Abstract

THE EVOLUTION OF*HOX* PARALOG GROUP 2 GENE EXPRESSION AND REGULATION IN THE JAPANESE MEDAKA

by

Adam Davis

July, 2011

Director of Dissertation: Edmund J. Stellwag, Ph.D.

Department: Biology

Hox paralog group 2 (PG2) genes are evolutionarily conserved developmental regulatory genes that function to specify rhombomere (r) and pharyngeal arch (PA) identities in animal embryos. Several rounds of whole genome duplications in animals, including one specific to ray-finned fishes, and post-genome duplication independent gene loss have resulted in divergent *Hox* PG2 gene complements across evolutionarily divergent teleost fishes. Divergence in gene complements may have, in part, been responsible for generating divergent *Hox* PG2 gene expression patterns and specification of hindbrain and PA-derived structures during the evolution of osteichthyan embryogenesis. In this dissertation, I describe the cDNA cloning and expression analysis of Japanese medaka (*Oryzias latipes*) *Hox* PG2 genes. I show that there are only two functional canonical genes, *hoxa2a* and *b2a*, and that a previously identified *hoxa2b* gene is a transcribed pseudogene, *ψhoxa2b*. The canonical genes, *hoxa2a* and *b2a*, were expressed in developing rhombomeres and PAs in a manner that was generally conserved throughout the osteichthyans. By contrast, *ψhoxa2b* was expressed at detectable levels only in noncanonical

Hox PG2 gene expression domains, including the ventral-most aspect of the neural tube, the pectoral fin buds and the caudal-most region of the embryonic trunk, indicative that regulatory control elements needed for spatiotemporal specification of expression have diverged from the canonical orthologs. In order to understand whether sequence divergence within *cis*-regulatory control elements are linked to the divergent expression patterns of the medaka *hoxa2* paralogs, conserved genomic sequences upstream of the medaka *hoxa2a* and *ψhoxa2b* genes were tested functionally using a transgenic GFP reporter system. The medaka *hoxa2a* r3/5 enhancer region (r3/5ER) was shown to direct reporter gene expression in r4, PA2 and the posterior PAs, while the r3/5ER of *ψhoxa2b* directed reporter gene expression in r3-7, PA2 and the posterior PAs, which is different from transgenic mapping studies of the orthologous regions tested in chick and mouse embryos. These analyses provide evidence for significant post-genome duplication divergence in *cis*-regulatory element function in the r3/5ER of osteichthyans. Further, they question the ancestral nature of the r3/5ER prior to the evolutionary split of sarcopterygians (lobe-finned fishes) from the actinopterygians (ray-finned fishes).

THE EVOLUTION OF*HOX* PARALOG GROUP 2 GENE EXPRESSION AND REGULATION IN THE JAPANESE MEDAKA

A Dissertation

Presented to the Faculty of the Department of Biology

East Carolina University

In Partial Fulfillment of the Requirements for the Degree

Doctor of Philosophy

by

Adam Davis

July, 2011

THE EVOLUTION OF*HOX* PARALOG GROUP 2 GENE EXPRESSION AND REGULATION IN THE JAPANESE MEDAKA

by

Adam Davis

APPROVED BY:

DIRECTOR OF DISSERTATION: _____
Edmund J. Stellwag, Ph.D.

COMMITTEE MEMBER: _____
Jean-Luc Scemama, Ph.D.

COMMITTEE MEMBER: _____
Anthony A.Capehart, Ph.D.

COMMITTEE MEMBER: _____
Brett D.Keiper, Ph.D.

CHAIR OF THE DEPARTMEN OF BIOLOGY: _____
Jeff McKinnon, Ph.D.

DEAN OF THE GRADUATE SCHOOL: _____
Paul J. Gemperline, Ph.D.

DEDICATION

I dedicate this work to my mother, Anne Eve Davis, for encouraging me to be incessantly curious, my wife, Rebecca Marie Davis, for encouraging me to go back to school after working in industry and being patient and supportive throughout this entire process and my son, Finnigan Jack Davis, for keeping me lighthearted and sane through the final stages of this process. I have fond memories of writing Chapter 4 of this dissertation when Finn was just two months old and was sleeping on my lap in the wee hours of the night.

ACKNOWLEDGEMENTS

First and foremost, I gratefully acknowledge my Doctoral Dissertation Advisor, Dr. Edmund J. Stellwag, for his excellent training in molecular genetic techniques and for teaching me to think systematically and pushing me to strive to be my best. I thank Dr. Stellwag for introducing me to evolutionary and developmental biology and forcing me to carefully dissect research from other projects in the attempt to relate them to my own research project.

For my other committee members, I thank Dr. Jean-Luc Scemama for his excellent guidance in molecular genetic and developmental biological techniques, including whole-mount *in situ* hybridization. I also thank Dr. Scemama for letting me utilize his laboratory space and equipment for my doctoral training. I thank Dr. Brett D. Keiper for his guidance on the microinjection process of medaka embryos. Without his expertise in microinjection, this research may have never taken off. I thank Dr. Anthony A. Capehart for his encouraging words and lending his expertise in developmental biology to me. I thank you all for your guidance and serving on my committee.

I thank Dr. Terry L. West for keeping me afloat financially from year to year during my doctoral training and allowing me to obtain excellent teaching experience.

To my previous lab mates, Pierre Le Pabic, Micheal Reubens and Micheal Odom, thank you all for your technical support. I wish you all well in your future endeavors.

On a more personal level, I would like to acknowledge the support of friends and family, most notably my wife, Rebecca Davis, my sisters, Ginger Davis and Fawne Weaver, Trushar ("Bobby") Patel, Micheal ("Mikey") Odom, the Ackermans (Dee, Rob, Earl, Rhonda and, I suppose, Lance), Rob Edler, Johnathan Weaver, Pierre Le Pabic, Micheal Reubens and, of course, my son, Finnigan ("The Finnster", "The Munch") Davis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: AN INTRODUCTION TO THE POST-GENOME DUPLICATION EVOLUTION OF *HOX* PARALOG GROUP 2 GENES IN THE OSTEICHTHYES

The field of evolutionary developmental biology seeks to provide a scientific explanation for the molecular mechanisms, processes and events that have generated the remarkable diversity of living forms that have emerged over the eons of life on earth. As a credit to the field, advancements stemming from the early 1980's showed that the incredible morphological diversity exhibited among the metazoans evolved upon an unexpected unitary framework of key developmental regulators, as exemplified by the *Hox*/*Hom* genes. Beginning with their discovery in *Drosophila* and subsequently throughout the bilaterians, animals that exhibit bilateral symmetry, the *Hox* complex served as the cornerstone for a new paradigm in thinking about the molecular systems responsible for shaping animal morphology. Their highly conserved sequences, stereotypical clustered organization and conserved expression patterns appeared contrary to them playing a role in the evolution of morphological diversity. However, recent evidence suggests that diversification of function, i.e. morphological specification coupled to evolutionary divergence, may be the products of two distinct but related processes, which are gene or genome duplication and divergence of duplicate gene function. While conservation of *Hox* gene coding sequences was discovered to be a hallmark of this clustered gene family, evidence has been forthcoming that the morphological specification characteristic of this gene family, commonly observed as colinearity between the physical order of the stereotypically clustered genes and their spatio-temporal order of expression during embryogenesis, may be altered by divergence in the *cis*-regulatory control sequences that regulate patterns of expression. The fundamental premise explored in this dissertation is that post-genome duplication divergence in *Hox* gene function is linked to evolutionary divergence in *cis*-regulatory element function.

An examination of this premise required the identification of a suitable model system and the characterization of *cis*-regulatory element function for an appropriate collection of genes. The following narrative provides the background for the selection of a suitable model system and the attendant issues relevant to the formulation of central hypotheses and working approaches.

Divergence of *Hox* Gene Clusters in the Osteichthyes

Multiple genome level duplications have expanded the total number of *Hox* gene clusters from a single cluster containing fourteen genes in the chordate or hemichordate ancestor of vertebrates to a collection of four or as many as eight clusters in the stem lineages leading to the tetrapods or teleosts, respectively. Current molecular evidence has shown that tetrapods contain thirty-nine *Hox* genes distributed among four clusters (Stellwag, 1999; Prince, 2002; Amores et al., 2004; Moghadam et al., 2005; Hoegg et al., 2007; Hurley et al., 2007) (Fig. 1-1A). At least one more whole-genome duplication event occurred specifically in the actinopterygian (ray-finned fish) lineage after their evolutionary split from sarcopterygians (lobe-finned fishes) and resulted in seven to eight *Hox* clusters (Stellwag, 1999; Prince, 2002; Amores et al., 2004; Moghadam et al., 2005; Hoegg et al., 2007; Hurley et al., 2007; Mungpakdee et al., 2008a) (Fig. 1-1A). Molecular phylogenetic studies have shown that the ray-finned fish specific genome duplication occurred prior to the evolutionary radiation of the Teleostei, the most diversified group of all vertebrates with over 23, 500 species (Nelson, 1994; Miya et al., 2003; Amores et al., 2004; Moghadam et al., 2005; Hoegg et al., 2007; Hurley et al., 2007).

Post-genome duplication independent gene loss across evolutionarily divergent teleost fish lineages has produced several divergent teleost *Hox* gene complements. For instance, the

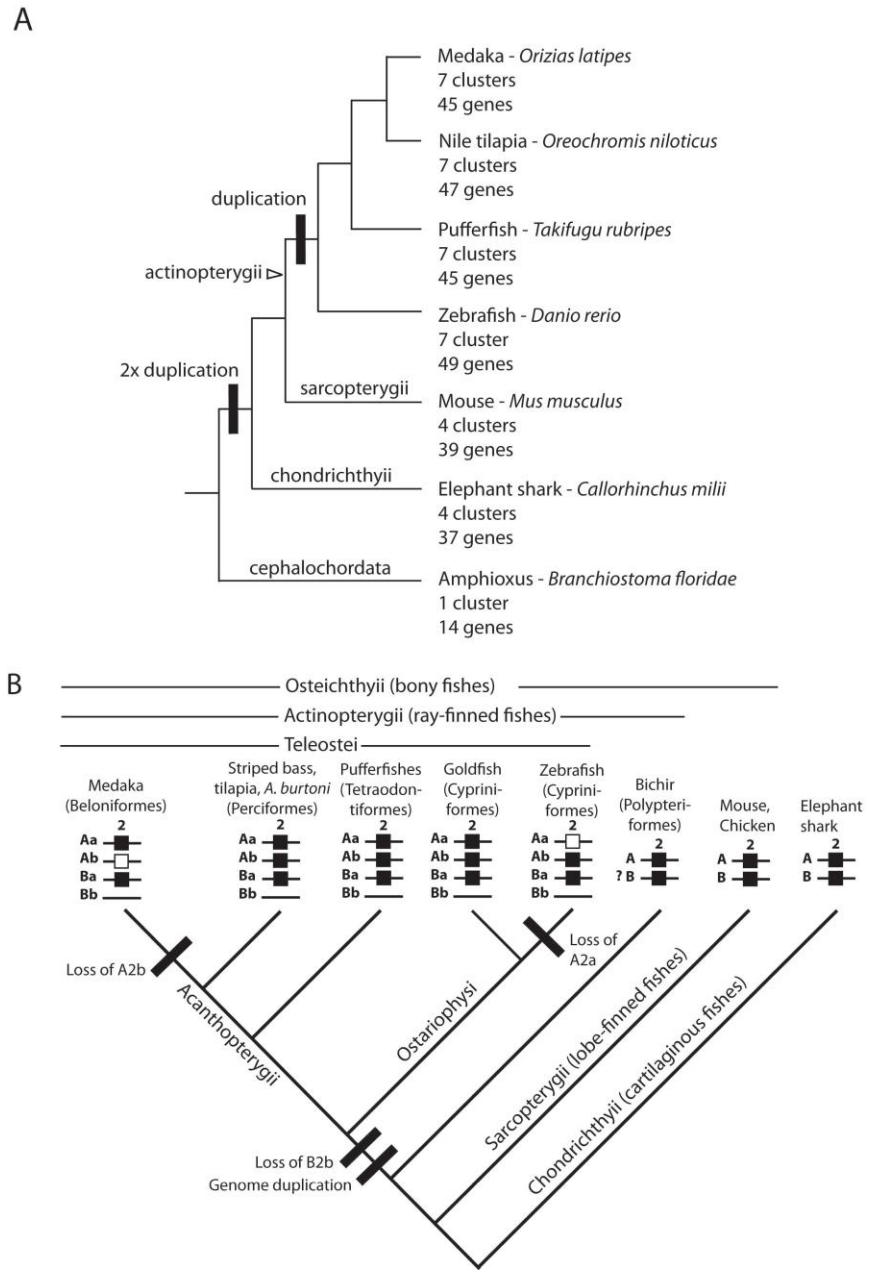


Fig. 1-1. *Hox* gene complement evolution in the Osteichthyes. (A) Phylogeny of osteichthyan *Hox* clusters. (B) Phylogeny of osteichthyan *Hox* paralog group 2 gene complements. Phylogenies based on Steinke et al. (2006).

Japanese medaka (*Oryzias latipes*) (superorder Acanthopterygii; order Beloniformes) and the Japanese pufferfish (*Takifugu rubripes*) (superorder Acanthopterygii; order Tetraodontiformes) each have seven *Hox* clusters comprised of 45 genes, the Nile tilapia (*Orceochromis niloticus*) (superorder Acanthopterygii; order Perciformes) has seven *Hox* clusters with 47 genes and the zebrafish (*Danio rerio*) (superorder Ostariophysii; order Cypriniformes) contains 7 *Hox* clusters with 49 genes (Fig. 1-1A). The divergence in *Hox* cluster gene complements provides an excellent platform to study how genome duplications and independent gene losses have affected the expression patterns and functions of *Hox* genes during embryogenesis across evolutionarily divergent osteichthyans.

Divergence of *Hox* Paralog Group 2 Gene Complements in the Osteichthyes

Perhaps the most extensively studied osteichthyan *Hox* genes in the field of evolutionary and developmental biology are from paralog group 2 (PG2). Based on current data from genomic and functional genetic studies, it has been hypothesized that the most recent common ancestor of sarcopterygians and actinopterygians had two *Hox* PG2 genes, *Hoxa2* and *b2* (Amores et al., 1998 and 2004; Stellwag, 1999) (Fig. 1-1B). These genes have been retained in tetrapods (Fig. 1-1B). Phylogenetic reconstructions that include a whole genome duplication event prior to the radiation of teleosts support a post-genome duplication ancestral *Hox* PG2 gene complement of four genes (*hoxa2a*, *a2b*, *b2a* and *b2b*) (Fig. 1-1B). The absence of a detectable *hoxb2b* gene in extant teleosts leads to the conclusion that this gene was lost prior to the emergence of the Teleostei (Fig. 1-1B). The further absence of a *hoxa2a* gene in zebrafish, but not in any members of the superorder Acanthopterygii, suggests the loss of this gene occurred after that of *hoxb2b* and that it was restricted to a clade including the zebrafish but not the acanthopterygians. The loss of *hoxa2b* in medaka, but not in any other documented teleost of the

Acanthopterygii, suggests that this gene loss occurred relatively late in the radiation of teleosts and was restricted to a clade including the medaka (Davis et al., 2008).

Hox PG2 Gene Expression and Function in the Osteichthyes

A comparison of the *Hox* PG2 embryonic hindbrain gene expression patterns among osteichthyans has revealed conserved expression patterns in the rostrally located hindbrain rhomobomeres (r2–5), irrespective of the evolutionary distance among the taxa being compared (Gendron-Maguire et al, 1993; Rijli et al., 1993; Prince et al., 1998; Grammatopoulus et al., 2000; Pasqualetti et al., 2000; Scemama et al., 2002, 2006; Baltzinger et al., 2005; Le Pabic et al., 2007; Davis et al., 2008). The degree of conservation in *Hox* PG2 gene expression patterns within the rostral rhombomeres was underscored by complete conservation in the osteoichthyan anterior boundaries of expression for *Hox* A and B cluster genes: expression always being localized to either the rhomobomere 1 and 2 (r1/2) boundary in the case of *Hox A* cluster PG2 genes or to the rhombomere 2 and 3 (r2/3) boundary for *Hox B* cluster PG2 genes. The conserved *Hox* PG2 gene expression patterns in the rostral hindbrain across evolutionarily divergent osteichthyans is suggestive of conserved hindbrain patterning functions. *Hox* PG2 gene function in the rostral hindbrain has been mapped in the mouse, where it has been shown that *Hoxa2* and *b2* control the segmentation of the anterior hindbrain, axonal guidance of the Vth and VIIth cranial motor nerve axons out of r2/r4 and specification of the somatic motor component of the VIIth cranial nerve exiting r4, respectively (Barrow and Capecchi, 1996; Gavalas et al., 1997, 2003; Davenne et al., 1999; Barrow et al., 2000). Moreover, analyses of homozygous *Hoxa2*/*b2* null mutant mice have shown the absence of rhombomeric boundaries between r1–r4, which has been interpreted as evidence of synergy between mouse *Hoxa2* and *b2* gene products in the specification of proper segmentation of the rostral rhombomeres (Davenne

et al., 1999). Given the conserved nature of the Vth and VIIth cranial motor nerve exit points from the hindbrain in comparisons between the mouse and zebrafish (Chandrasekhar, 2004), coupled with related expression patterns of *Hox* PG2 genes in the rostral hindbrain of tetrapods and teleosts, it is probable that the *Hox* PG2 genes in teleosts function similarly to those of their mouse orthologs.

Beyond their hindbrain expression and role in patterning the hindbrain, *Hox* PG2 genes have been shown to be expressed in the pharyngeal arches, which are tissue primordia that give rise to the craniofacial bones. Further, osteichthyan *Hox* PG2 genes function as homeotic selector genes in patterning the craniofacial derivatives of the second pharyngeal arch (PA2) (Gendron-Maguire et al., 1993; Rijli et al., 1993; Grammatopoulos et al., 2000; Pasqualetti et al., 2000; Hunter and Prince, 2002; Baltzinger et al., 2005; Le Pabic et al., 2010). At present, it appears that the two major osteichthyan lineages, those leading to teleosts or tetrapods, have partitioned selector gene activity in distinct ways. Tetrapods have a single *Hox* cluster-encoded PG2 gene with pharyngeal arch expression and PA2-directed selector gene activity (*Hoxa2*) (Gendron-Maguire et al., 1993; Rijli et al., 1993; Grammatopoulos et al., 2000; Pasqualetti et al., 2000; Baltzinger et al., 2005). Knockout and knockdown experiments of *Hoxa2* in tetrapods have resulted in the homeotic transformation of PA2-derived bony elements into first pharyngeal arch (PA1)-like bony elements (Gendron-Maguire et al., 1993; Rijli et al., 1993; Baltzinger et al., 2005). Further, misexpression of *Hoxa2* in PA1 resulted in the respecification of PA1-derived bony elements to resemble those of PA2 (Grammatopoulos et al., 2000; Pasqualetti et al., 2000). For teleosts, *hoxa2b* and *b2a* of zebrafish were shown to be expressed in the pharyngeal arches and function redundantly as selector genes of PA2 identity, such that the knockdown of both *Hox* PG2 genes was required to induce a homeotic PA2 to PA1 transformation (Hunter and Prince,

2002). In tilapia, in which *hoxa2a, a2b* and *b2a* are all expressed in the pharyngeal arches, individual knockdowns of each *Hox* PG2 gene resulted in the anteriorizing homeotic transformations in PA2, which suggested that each gene is able to function independently as a selector gene in patterning the PA2 identity (Le Pabic et al., 2010). The differences in expression and functional diversification in the pharyngeal arches among the aforementioned osteichthyans leaves uncertainty regarding the ancestral osteichthyan *Hox* PG2 gene pharyngeal arch expression pattern and function in determining PA2 identity. Interestingly, recent expression analyses of *Hox* genes in the dogfish shark (*Scyliorhinus canicula*) have shown that both *Hoxa2* and *b2* are expressed in PA2 and the posterior pharyngeal arches (Oulion et al., 2011), and these analyses suggest that tetrapods have lost the *cis*-regulatory activity that enables *Hoxb2* to be expressed and to ultimately function in PA2.

Hox Paralog Group 2 Enhancer Regions in Osteichthyans

Given that *Hox* genes are conserved in their coding sequences, the divergence of expression and function of *Hox* PG2 genes is most likely the result of mutational changes in their *cis*-regulatory sequences (Caroll, 2008). Results from functional studies of the regulatory loci controlling expression of osteicthyan *Hox* PG2 genes have identified three characteristic regulatory modules responsible for directing expression in rhombomeric compartments of the hindbrain (Fig. 1-2). A large and complex region upstream of *Hoxa2* has been shown to direct *Hoxa2* expression within r3 and r5 of the hindbrain and is comprised of several Krox20 binding sites that work in conjunction with several other rhombomeric elements (RE1-5; BoxA) (Frasch et al., 1995; Nonchev et al., 1996a and b; Maconochie et al., 2001: Tümpel et al., 2002 and

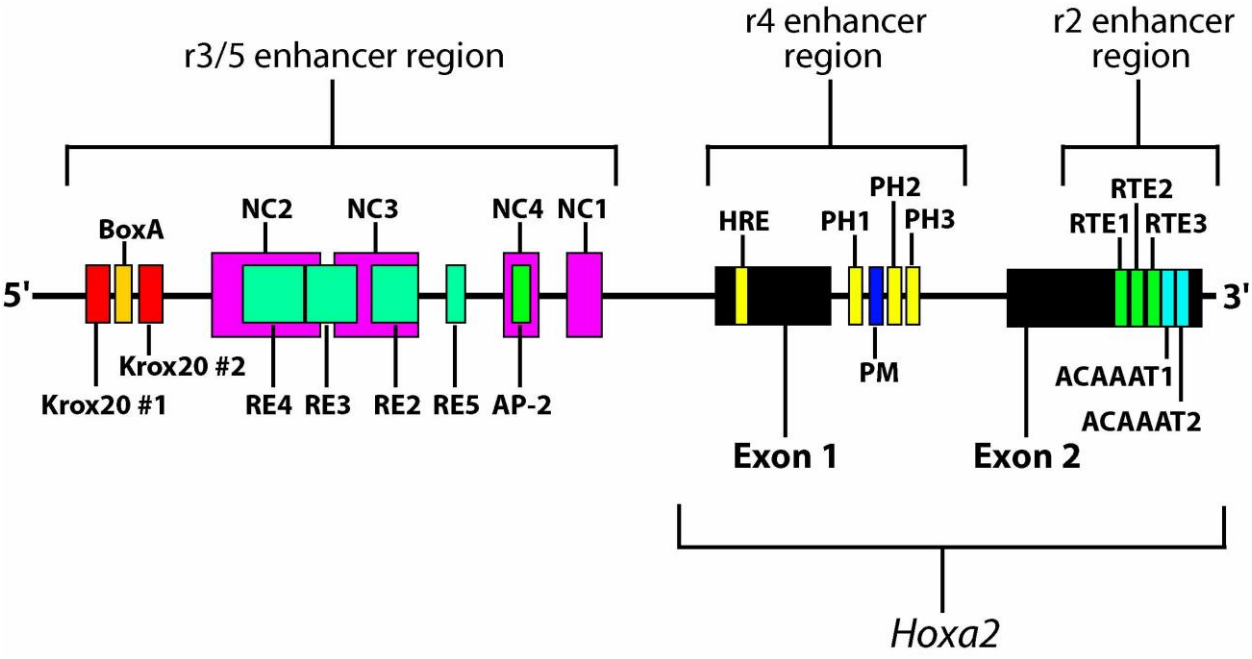


Fig. 1-2. Enhancer modules that control *Hoxa2* gene expression in the hindbrain and pharyngeal arches of tetrapods. AP-2, activator protein complex 2; HRE, Hox responsive element; NC, neural crest; PH, Hox/Pbx; PM, Prep/Meis; RE, rhombomeric element; RTE, rhombomere two element.

2006). Interspersed among the *cis*-regulatory elements that direct *Hoxa2* expression in r3 and r5 are several elements that have been shown to direct *Hoxa2* expression in the cranial neural crest cells (CNCCs) that delaminate from the hindbrain and populate PA2 and the posterior pharyngeal arches (Maconochie et al., 1999). A second module consisting of four Hox/Pbx binding elements, one located in exon 1 upstream of the hexapeptide (designated as HRE for Hox responsive element) and three within the intron (designated as PH1-3), and 1 Prep/Meis (PM) binding element also located in the intron are responsible for directing *Hoxa2* expression in r4 of the hindbrain (Tümpel et al., 2006 and 2007; Lampe et al., 2008). Finally, a third region is located downstream of the homeodomain within exon 2 and appears to be responsible for directing *Hoxa2* expression in r2. This region consists of three r2 elements (RTE1-3) that restrict *Hoxa2* expression to r2 and two ACAAT motifs (ACAAT1-2) that act as Sox2 binding elements (Tümpel et al., 2006 and 2008). Most of the regulatory elements directing *Hoxa2* expression in r3/5, r4, r2 and the CNCCs have been identified in mouse (Frasch et al., 1995; Nonchev et al., 1996b; Maconochie et al., 1999 and 2001; Tümpel et al., 2007 and 2008; Lampe et al., 2008), with large subsets of these elements shown to be present and functional in chick embryos (Nonchev et al., 1996b; Maconochie et al., 2001; Tümpel et al., 2007; Lampe et al., 2008). It should be noted that the regulatory elements that direct mouse and chicken *Hoxa2* expression were tested in their respective host species (homologous expression systems).

Comparative genomic analyses of osteichthyan *Hox* PG2 genes have shown conservation between tetrapod and teleost *Hoxa2* regulatory sequences corresponding to the r3/5, r4 and r2 enhancers (Tümpel et al., 2006; Tümpel et al., 2008; Raincrow, 2010). However, functional analyses of these teleost *Hox* PG2 enhancer regions have been limited to heterologous expression systems, such that the r3/5, r4 and r2 enhancer regions of *hoxa2a* and *a2b* of fugu and *hoxa2a*

and *ψhoxa2b* of medaka were tested in chicken embryos (Tümpel et al., 2006). Interestingly, heterologous reporter gene assays of the r3/5, r4 and r2 enhancer regions of fugu and medaka *hoxa2a* failed to demonstrate reporter gene expression in the hindbrain or pharyngeal arches of chicken embryos but the paralogous regions for fugu *hoxa2b* and medaka *ψhoxa2b* were shown to function in chicken embryos in a similar manner to the r3/5, r4 and r2 enhancer regions of mouse and chicken *Hoxa2* (Tümpel et al., 2006). While these functional genomic results appeared to be consistent with the expression results for fugu *hoxa2a* and *a2b*, which were shown to be expressed in r1 and r2 and r2-5, respectively, they were not for medaka *hoxa2a* and *ψhoxa2b*, which were shown to be expressed in r2-r8 and in noncanonical *Hox* PG2 expression domains, respectively (Amores et al., 2004; Tümpel et al., 2006; Davis et al., 2008).

Goal of Thesis

The goal of my thesis was to study the effects of the ray-finned fish-specific whole genome duplication and post-genome duplication independent gene loss on the evolution of *Hox* PG2 *cis*-regulatory activity during Japanese medaka (*Oryzias latipes*) embryonic development. Gene duplication, followed by sequence alteration has been hypothesized to to play a significant role in generating genetic and morphological novelty (Taylor and Raes, 2004). However, the most likely fate of a duplicated gene is loss of function, either through mutations in coding sequences leading to null alleles or in noncoding regulatory sequences leading to changes in gene expression (Haldane, 1933). Alternate fates of gene duplication include retention of both duplicates where either one gene acquires a novel and beneficial function while the other retains the function of the ancestral gene or in which both duplicates undergo degenerate and complementary mutations, such that their *cis*-regulatory elements operate to recapitulate the function of their single ancestral gene (Force et al., 1999). Interestingly, very few reporter gene

expression analyses have been conducted for studying *Hox* gene *cis*-regulatory element evolution in gene duplicates in teleosts, and of these, none were performed using teleost model systems to assay reporter gene expression (Tümpel et al., 2006). Advances in vector-based integration systems, such as the *Tol2* transposon system, allows for the use of fish model systems to study *cis*-regulatory element control of gene expression.

The use of medaka, which has a divergent *Hox* PG2 gene complement from those of other teleosts, for studying *cis*-regulatory activity on *Hox* PG2 gene expression provides an excellent basis for the comparison of *cis*-regulatory element evolution in teleost *Hox* PG2 gene expression. In chapter 2 of this thesis, I show that medaka *hoxa2a* and *b2a* possess similar expression patterns to orthologous *Hoxa2* and *b2* genes of other osteichthyans. Medaka *hoxa2b*, however, was shown to be expressed in noncanonical *Hox* PG2 domains, including the caudal-most region of the embryonic trunk, the ventral-most aspect of the neural tube and the distal mesenchyme of the pectoral fin buds (Described in Chapter 2). Sequence analysis of the cDNA sequence of medaka *hoxa2b* showed the presence of several termination codons in exon 2 in the sequences encoding the homeodomain. Based on the combined results of *in situ* hybridization, cDNA sequencing and comparative genomics, we were able to provide evidence that the *hoxa2b* sequence encoded an expressed pseudogene, *ψhoxa2b*.

The divergent expression of medaka *ψhoxa2b* from its paralog, *hoxa2a*, as well as its orthologs in other teleosts, helped to generate hypotheses regarding the structural and functional nature of the *cis*-regulatory elements directing expression of medaka *hoxa2a* and *ψhoxa2b* (described in Chapter 2). Towards this end, I developed a research project utilizing the *Tol2* transposon mechanism of transgene integration into the medaka genome for studying the r3/5 intergenic enhancer region upstream of medaka *hoxa2a* and *ψhoxa2b*, which is described in

Chapter 3 of this thesis. In mouse, the orthologous genomic sequences corresponding to the r3/5 enhancer region (r3/5ER) were shown to direct mouse *Hoxa2* in r3 and r5 of the hindbrain, PA2 and the posterior pharyngeal arches (Frasch et al., 1995; Nonchev et al., 1996a and b, Maconochie et al., 2001; Tümpel et al., 2002; Tümpel et al., 2006). Therefore, I hypothesized that the r3/5ER of medaka *hoxa2a* would function in a conserved manner to that of the r3/5ER of mouse *Hoxa2*, which directs gene expression in r3, r5 and the pharyngeal arches. I further hypothesized that, based on the unusual expression pattern of the medaka *ψhoxa2b* transcript, that the r3/5ER of medaka *ψhoxa2b* would direct transcript expression in noncanonical domains instead of the hindbrain or pharyngeal arches. Contrary to the hypothezied expression patterns, I observed that the medaka *hoxa2a* r3/5ER directed expression in r4 and the pharyngeal arches and the medaka *ψhoxa2b* r3/5ER directed expression in r3-r7 of the hindbrain and in the pharyngeal arches. These results were unexpected and showed that the regulatory sequences tested in transgenic reporter gene assays did not phenocopy the transcription patterns of either the *hoxa2a* or *ψhoxa2b* genes in medaka. Further, these results point to evolutionary divergence in function of the r3/5ER in the gene lineages leading to mouse *Hoxa2*, medaka *hoxa2a* and medaka *ψhoxa2b*.

The functional genomics study outlined in Chapter 3 provides an excellent basis to study how genome duplications and post-genome duplication independent gene loss affect the evolution of *cis*-regulatory element function. The functional mapping of *cis*-regulatory elements within a specific model system, such as the mouse or chicken, can provide an explanation of how kernels, or subcircuits within genetic regulatory networks, are involved in patterning specific morphological structures (Davidson, 2006). Further, comparative genomic sequence analyses of the mapped *cis*-regulatory enhancers with orthologous conserved noncoding sequences of

evolutionarily divergent species allows us to formulate hypotheses concerning the conservation of function of the subcircuits that pattern homologous morphological features between species. This is especially true when the developmental genes in question show similar expression patterns in comparisons among evolutionarily divergent animals. Results present in Chapter 3 show that conservation of expression patterns and genomic sequences do not necessarily ensure the conservation of the regulatory function of conserved noncoding sequences between evolutionarily divergent species and suggest that conserved noncoding sequences may take part in the generation of morphological novelties. More studies must be performed to understand how genome duplications affect the functional nature of regulatory subcircuits on their involvement in patterning morphological structures.

The results reported in Chapter 3 also underscore the importance of using homologous reporter gene systems to study the function of *cis*-regulatory elements. The reporter gene expression results that I observed from the medaka *hoxa2a* and *ψhoxa2b* r3/5ERs within medaka embryos were markedly different than those reported when the same enhancer regions were analyzed in chicken embryos (Tümpel et al., 2006). For instance, while Tümpel et al. (2006) showed that the medaka *hoxa2a* r3/5ER did not direct any observable gene expression within the chicken embryo, my study showed that the same enhancer region directed reporter gene expression in r4 of the hindbrain and the CNCCs in medaka embryos. Further, Tümpel et al. (2006) showed that the *ψhoxa2b* r3/5ER directed reporter gene expression in r3 and r5 but not in the CNCCs of chicken embryos, while my study showed that this same enhancer region directed expression in r3-7 and the CNCCs of medaka embryos. The lack of medaka *hoxa2a* and *ψhoxa2b* r3/5ER-driven reporter gene expression in r4 and the CNCCs and in r4, r6, r7 and the CNCCs, respectively, of chick embryos suggests that these enhancer regions were not efficient in

utilizing the *trans*-acting factors that were present in the heterologous chick model system. Further, these results show that caution must be used when interpreting results from heterologous reporter gene systems, especially when studying conserved noncoding sequences from species for which suitable homologous hosts are not available.

In Chapter 4 of this thesis, I investigated the nested expression patterns of medaka *Hox* PG3-6 genes in the developing posterior pharyngeal arches. This study is the first of its kind and provides a basis for comparative studies of posterior arch-expressing *Hox* genes in evolutionarily divergent teleosts. The bony structures of the pharyngeal jaw apparatus have been shown to be variable in morphology across evolutionarily divergent teleosts (e.g.: Langille and Hall, 1987; Parenti, 1987; Le Pabic et al., 2009). *Hox* genes occupy a significant role in patterning the bony architecture arising from the pharyngeal arches (Gendron-Maguire et al., 1993; Rijli et al., 1993; Grammatopoulus et al., 2000; Hunter and Prince, 2002; Baltzinger et al., 2005; Crump et al., 2006; Minoux et al., 2009; Le Pabic et al., 2010). The nested expression patterns of *Hox* genes along the anterior-posterior axis of developing embryos has given rise to the *Hox* code hypothesis, which states that a given combination of *Hox* gene products is necessary for specifying segmental identities along the A-P axis (Krumlauf, 1994). Given that medaka and other beloniform fishes possess autopomorphic bony characteristics arising from the posterior arches, it is possible that the nested *Hox* gene expression patterns presented in Chapter 4 represent a *Hox* code that is specific to beloniform fishes.

Choice of Experimental System

The Japanese medaka (*Oryzias latipes*) was chosen for my study for several reasons. Medaka is one of a few vertebrate model systems used for studying mechanisms of development (Metscher and Ahlberg, 1999; Wittbrodt et al., 2002). The genome of medaka has been fully

sequenced and the developmental stages are well documented for this species (Iwamatsu, 2004; Kasahara et al., 2007). Further, medaka possesses traits that bode well for its utility as a tractable laboratory system. It is a successful species for which aquarium maintenance and breeding conditions are well documented (Kirchen and West, 1999). Medaka spawns on a daily basis to yield up to 25-35 fertilized eggs per female, which is important for performing the microinjection of constructs containing *cis*-regulatory elements used for reporter gene assays. Of particular interest to my research project, the genomic sequences specific to medaka *hoxa2a* and *ψhoxa2b* provide an excellent source of genetic material for studying how genome duplications affect the evolution of *cis*-regulatory circuitry in directing spatio-temporal gene expression patterns during embryogenesis.

CHAPTER 2: JAPANESE MEDAKA *HOX* PARALOG GROUP 2: INSIGHTS INTO THE EVOLUTION OF *HOX* PG2 GENE COMPOSITION AND EXPRESSION IN THE OSTEICHTHYES

Introduction

Hox genes are a family of evolutionarily related developmental regulatory genes that serve as critical genetic determinates of regional tissue identity along the anterior–posterior (A–P) axis of animal species (McGinnis and Krumlauf, 1992). They are organized on the genome of chordates in clusters comprising as many as 14 genes that are expressed along the A–P axis of developing embryos collinear with their physical location within a cluster (Holland and Garcia-Fernandez, 1996; Ferrier et al., 2000; Powers and Amemiya, 2004). Multiple genome level duplications have expanded the total number of *Hox* clusters from one in chordates to four in tetrapods, 7 or 8 in most teleosts and even 13 in the salmoniform fishes (Stellwag, 1999; Amores et al., 2004; Moghadam et al., 2005; Woltering and Durston, 2006; Hoegg et al., 2007; Hurley et al., 2007; Mungpakdee et al., 2008). A consequence of these repeated genome duplications has been the generation of a stereotypical series of 13 clustered gene paralog groups that differ in gene number depending on the historical timing of gene losses relative to genome duplications. The most extreme case of gene content variation within a paralog group of a single species is the Japanese pufferfish, which has lost the entire complement of *Hox* paralog group 7 genes while retaining six of the possible eight genes in paralog group 9 (Amores et al., 2004). In addition to variation in the gene number within and among paralog groups, results from functional genetic studies have shown that individual genes within a paralog group can exhibit heterogeneous activities; sometimes demonstrating divergent, synergistic or even redundant functions relative to their paralogous counterparts (Condie and Capecchi, 1993; Hunter and Prince, 2002; Lynch and

Conery, 2003). The heterogeneous activities of *Hox* gene paralogs prompt questions about the lineage-specific effects of gene loss on the activity of surviving genes within a paralog group. Does *Hox* gene loss within a paralog group constrain the activity of the remaining genes, resulting in canalization of function and reduced evolutionary divergence; or, does it shift the spectrum of interactions between and among cross-regulatory *Hox* genetic networks resulting in homeotic transformations? Given the extensive evolutionary sculpting of the clustered *Hox* genes over the course of vertebrate history, answers to these fundamental questions will provide insight into the pathways by which evolution shapes development and affects organismal morphology.

Research in our laboratory has been directed toward understanding the effects of teleost *Hox* paralog group 2 (*Hox* PG2) post-genome duplication divergence on the evolution of developmental gene function. Genetic and developmental studies of osteichthyian *Hox* PG2 gene functions have shown that these genes appear to act primarily as homeotic selectors in the specification of 2nd pharyngeal arch (PA2)-derived structures (Gendron-Maguire et al., 1993; Rijli et al., 1993; Grammatopoulos et al., 2000; Pasqualetti et al., 2000; Hunter and Prince, 2002; Baltzinger et al., 2005). At present, it appears that the two major osteichthyan lineages, those leading to teleosts or tetrapods, have partitioned selector gene activity in distinct ways: tetrapods have a single *Hox* cluster-encoded PG2 gene with PA2-directed selector gene activity (*Hoxa2*), whereas teleosts, represented by zebrafish, have two selector genes, *hoxa2b* and *b2a* that function redundantly (Gendron-Maguire et al., 1993; Rijli et al., 1993; Grammatopoulos et al., 2000; Pasqualetti et al., 2000; Hunter and Prince, 2002; Baltzinger et al., 2005). This difference in functional diversification leaves uncertainty as to whether one (*Hox A*) or two (*Hox A* and *Hox B*) *Hox* PG2 genes functioned as selector genes of PA2 identity in the common osteichthyian

ancestor. Unfortunately, a scarcity of information concerning the *Hox* PG2 gene composition, expression and function among osteichthyans precludes discrimination between these alternative possibilities and limits our understanding of the evolutionary developmental history of osteichthyian *Hox* PG2 gene function.

Recent results from our laboratory, based on the cloning and expression analyses of striped bass (*Morone saxatilis*) and Nile tilapia (*Oreochromis niloticus*) *Hox* PG2 genes, have shown that these teleosts, unlike zebrafish (*Danio rerio*), have three expressed genes, including *hoxa2a, a2b* and *b2a* (Scemama et al., 2006; Le Pabic et al., 2007). Genomic sequencing of the *Hox* cluster genes among other teleosts, including the Japanese medaka (*Oryzias latipes*) and the pufferfishes (*Takifugu rubripes, Spheroides nephalus* and *Tetraodon nigroviridis*) have provided evidence for three Hox PG2 genes also, although neither their expression patterns nor functions are clear (Amores et al., 2004; Jaillon et al., 2004; Naruse et al., 2004).

Phylogenetic reconstructions that include a whole genome duplication event at the incipient stage of teleost evolution support a post-genome duplication ancestral *Hox* PG2 gene complement of four genes (*hoxa2a, a2b, b2a* and *b2b*), see Figure 2-1. The absence of a detectable *hoxb2b* gene in extant teleosts leads to the conclusion that this gene was lost relatively early and likely earlier than the divergence of zebrafish from the teleost stem lineage. The further absence of a *hoxa2a* gene in zebrafish, but not in any members of the superorder Acanthopterygii, suggests the loss of this gene occurred after that of *hoxb2b* and that it was restricted to a clade including the zebrafish but not the acanthopterygians. Accordingly, the established *Hox* PG2 gene composition of extant teleosts shows that the complement of genes represented by zebrafish is not pleisiomorphic for teleosts and indicates that the functional nature of the two zebrafish genes may have been affected by the loss of *hoxa2a*. Our results from

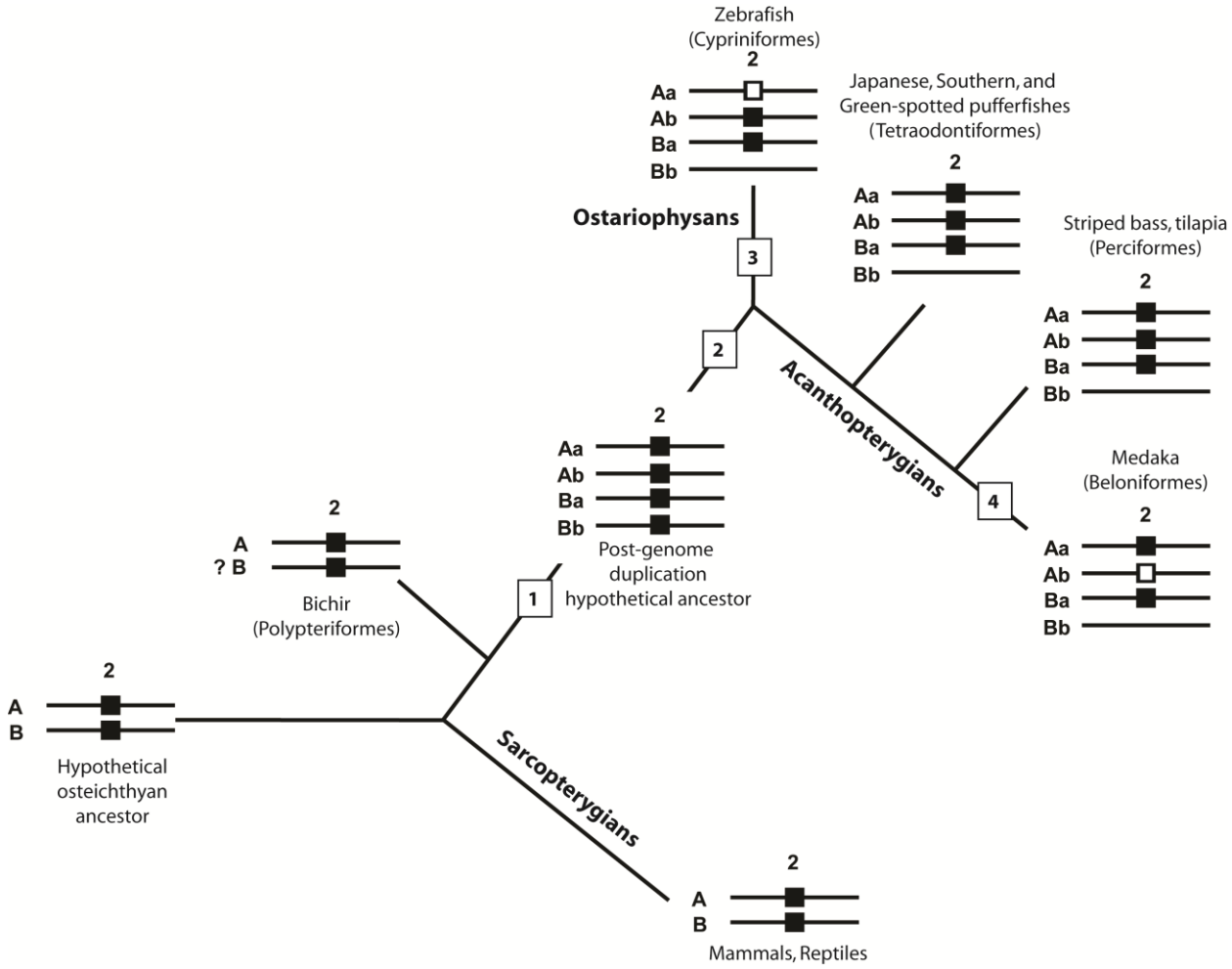


Fig. 2-1. *Hox* PG2 gene complement evolution in the Osteichthyes. Phylogeny based on Steinke et al. (2006). (1) Genome duplication; (2) *hoxb2b* gene loss; (3) *hoxa2a* gene loss; (4) *hoxa2b* gene loss.

expression patterns and functional studies in tilapia and striped bass suggest that *hoxa2a* may function as a nonredundant selector gene in PA2 and that *hoxa2b* and *b2a* may have functions distinct from their strict zebrafish orthologs (Le Pabic et al., 2007, 2008). If the activities of the *Hox* PG2 genes from tilapia and striped bass are characteristic of other teleosts with three *Hox* PG2 genes, then it seems possible that the loss of zebrafish *hoxa2a* may indeed have affected the functional nature of the surviving zebrafish *Hox* PG2 paralogs.

Beyond their role as selector genes of PA2 identity, osteichthyan *Hox* PG2 genes have been shown also to influence hindbrain development and functional specification (Barrow and Capecchi, 1996; Gavalas et al., 1997; 2003; Davenne et al., 1999; Barrow et al., 2000). A comparison of the *Hox* PG2 embryonic hindbrain gene expression patterns among osteichthyans revealed a mosaic distribution in which expression in the rostrally located hindbrain rhomobomeres (2–5) was relatively well conserved, irrespective of the evolutionary distance among the taxa being compared, whereas expression in the most caudal rhombomeres (6 and 7) was divergent even among taxa that were relatively closely related (Scemama et al., 2002, 2006; Le Pabic et al., 2007). The degree of conservation in *Hox* PG2 gene expression patterns within the rostral rhombomeres was underscored by complete conservation in the osteoichthyan anterior boundaries of expression for *Hox* A and B cluster genes: expression always being localized to either the rhomobomere 1 and 2 (r1/2) boundary in the case of *Hox A* cluster PG2 genes or to the rhombomere 2 and 3 (r2/3) boundary for *Hox B* cluster PG2 genes. This degree of conservation was also reflected in the close sequence relationships among *cis*-acting regulatory elements that directed *Hox* PG2 gene expression in both tetrapods and teleosts (Tümpel et al., 2006). A number of conserved regulatory sequences that include those that have been shown to direct *Hox* PG2 gene expression in r3/5 (Krox20) or r4 (Hoxb1/Pbx/Meinox) have been mapped

functionally in the mouse and identified bioinformatically for other osteichthyans (Shamet al., 1993; Frasch et al., 1995; Nonchev et al., 1996a,b; Ferretti et al., 2000; Maconochie et al., 2001; Scemama et al., 2002; Tümpel et al., 2002, 2006, 2007).

The conserved *Hox* PG2 gene expression patterns in the rostral hindbrain across evolutionarily divergent osteichthyans is suggestive of conserved hindbrain patterning functions. *Hox* PG2 gene function in the rostral hindbrain has been mapped in the mouse, where it has been shown that *Hoxa2* and *b2* control the segmentation of the anterior hindbrain, axonal guidance of the Vth and VIIth cranial motor nerve axons out of r2/r4 and specification of the somatic motor component of the VIIth cranial nerve exiting r4, respectively (Barrow and Capecchi, 1996; Gavalas et al., 1997, 2003; Davenne et al., 1999; Barrow et al., 2000). Moreover, analyses of homozygous *Hoxa2*/*b2* null mutant mice have shown the absence of rhombomeric boundaries between r1–r4, which has been interpreted as evidence of synergy between mouse *Hoxa2* and *b2* gene products in the specification of proper segmentation of the rostral rhombomeres (Davenne et al., 1999). Given the conserved nature of the Vth and VIIth cranial motor nerve exit points from the hindbrain in comparisons between the mouse and zebrafish (Chandrasekhar, 2004), coupled with related expression patterns of *Hox* PG2 genes in the rostral hindbrain of tetrapods and teleosts, it is probable that the *Hox* PG2 genes in teleosts function similarly to those of their mouse orthologs.

Despite the extensive similarity in the *Hox* PG2 gene expression patterns in the rostral hindbrain of osteichthyans, the caudal rhombomeres exhibit extensive divergence in expression patterns even among relatively closely related taxa. As is the case for PA2 specification, it is impossible to formulate meaningful hypotheses regarding the pattern of hindbrain expression in the common osteichthyan ancestor. Although comparisons of the expression patterns and

deduced function of teleost *Hox* PG2 genes from zebrafish, striped bass and tilapia suggest that differential post-genome duplication gene loss may have influenced hindbrain activity as well as selector gene function, further investigation of other species with the same or different *Hox* PG2 gene complement will be required to better understand the effects of post-duplication gene loss.

The Japanese medaka (medaka) is one of only a very few vertebrate developmental and genomic model organisms and, as such, represents an important resource for evolutionary developmental studies (Metscher and Ahlberg, 1999; Wittbrodt et al., 2002). Of particular interest to our laboratory, a comprehensive genomic characterization of the medaka *Hox* clusters revealed a *Hox* PG2 gene complement interpreted to be composed of three functional genes, *hoxa2a*, *a2b* and *b2a* (Kurosawa et al., 2006). Given our hypotheses concerning the correlation among *Hox* PG2 gene complements, their expression patterns and gene functions, we were interested to determine the expression patterns of the three genes from medaka to establish whether, as we expected, they were similar to other acanthopterygians with three *Hox* genes. We also wanted to understand whether the expression patterns of medaka *Hox* PG2 genes varied in a manner consistent with the evolutionary divergence of the bony elements derived from PA2 in medaka compared with other teleosts.

Medaka and other members of the order Beloniformes have been shown to be deficient in skeletal elements common among other teleosts (Rosen and Parenti, 1981; Parenti, 1987). We were particularly interested in the absence of the PA2-derived interhyal bone, which is a relatively small, but important, rounded cartilage typically located between the hyosymplectic and the distal most aspect of the ceratohyal. In most teleosts, the skeletal elements of the hyoid arch have been demonstrated to be composed of four elements, which include the ventrally located basihyal, ceratohyal, interhyal and the dorsal hyosymplectic (composed of the

hyomandibular and symplectic), as well as dermal elements, the branchiostegal rays and the opercular bone. The dermal bones are common across all teleosts and function in respiration. As for the other bony elements, the dorsal hyosymplectic functions to bridge the upper jaw, or 1st pharyngeal arch (PA1)-derived palatoquadrate to the base of the cranium and the interhyal functions as an articulation point between the hyosymplectic and the ceratohyal (Schaeffer and Rosen, 1961). The interhyal forms a critical component of the suspensorium and has been argued to be central to the feeding mechanism employed by teleosts in which it is found. It has been concluded based on functional morphological studies that the interhyal serves as a ‘‘mobile pivot’’ for the hyoid bar (Schaeffer and Rosen, 1961), which is the ossified bone derived from the ceratohyal and surrounding hypohyal and epihyal cartilages. In the absence of the interhyal, the beloniform hyoid bar is connected directly to the hyosymplectic via ligaments and the range of motion exhibited by the hyoid apparatus is more limited than in species in which the interhyal is present (Parenti, 1987). Thus medaka and other beloniforms display limited movement of the branchial apparatus, which restricts the mechanics of orobranchial chamber expansion and jaw protrusion within this group and precludes a highly efficient form of suction feeding that is common among other teleosts (Parenti, 1987). This derived beloniform-specific bony architecture poses questions about whether the *Hox* PG2 genes in this group have evolved unique expression patterns and functions that are divergent from teleosts that have retained these bony elements.

In this article, we report evidence from molecular cloning and characterization of medaka *Hox* PG2 genes that supports a unique gene composition with two functional genes, *hoxa2a* and *b2a*, as well as a pseudogene, *ψhoxa2b*, that was previously misassigned as a gene. Gene expression studies based on *in situ* hybridization of medaka whole-mount embryos revealed that

the two functional genes are similar to orthologs and co-orthologs from other teleosts in their general patterns of hindbrain and pharyngeal arch expression; however, they share especially similar expression patterns with zebrafish in comparisons to the more closely related species, striped bass and tilapia. We examine the relationships among the *Hox* PG2 gene expression patterns of these divergent teleosts and consider their evolutionary implications in light of post-genome duplication effects on lineage-specific gene loss events.

Materials and Methods

Japanese medaka *Hox* PG2 and *egr2* gene cDNA cloning

Medaka *hoxa2a*, *hoxa2b*, *hoxb2a* and *egr2* partial cDNAs were generated by RT–PCR using total RNA isolated from stage 19 (2 somite stage) medaka embryos according to the manufacturer's procedure (Totally RNAs, Applied Biosystems, Foster City, CA). The primers used for the amplification of *hoxa2a*, *a2b* and *b2a* partial cDNAs were designed based on medaka genomic sequences (Accession numbers: AB207976, AB207985 and AB20799) (Kurosawa et al., 2006) to amplify a 556, 1018 and 617 bp fragment of each transcript, respectively (meda2aF: 5'- GATGCGGGAGAAGAAAGC-3' and meda2aR: 5'-CAGTGGGTGATGGATTGG-3'; meda2bF: 5'-GTGGTTTTATCAACAGCC-3' and meda2bR: 5'-CAGTAAATCAGGTTTTGC-3'; medb2aF: 5'-CGCACCGCCTACACCAAC-3' and medb2aR: 5'-GAAGGTGAGGTCAGGGAG-3'). *Egr2* primers were designed based on conserved sequences shared among mouse, zebrafish and striped bass cDNA sequences (Egr2F: 5'-GGCTACCCTCTGCTTACAGTC-3' and Egr2R: 5'-GGAGGTGGATTTTGGTGTGTC-3') (Scemama et al., 2006). The PCR products generated from RT–PCR-mediated amplification of medaka embryonic total RNA were cloned into pCR II vectors (Invitrogen, Carlsbad, CA), according to the manufacturer's instructions. Confirmation and orientation of PCR products

corresponding to inserts from plasmid cDNA clones were determined by restriction endonuclease digestion and DNA sequencing using dideoxyterminator sequencing chemistry (Big Dye v. 3.0, Applied Biosystems, Foster City, CA). The partial cDNA sequences are available in the Genbank database under the accession numbers EU483626, EU483627, EU483628 and EU493253, for *hoxa2a*, *a2b*, *b2a* and *egr2*, respectively.

Whole-mount *in situ* hybridization

Medaka (golden strain) were obtained from Carolina Biological Supply Company and were cultivated in 1X embryo rearing medium (ERM) (17mM NaCl, 0.4mM KCl, 0.66mM MgSO4.7H2O, 0.27mM CaCl2.2H2O, pH 7.2) at 26–28°C under a 14/10 hr light/dark cycle (Oxendine et al., 2006). Adults were maintained in 3:2 female to male ratios. Embryos used for *in situ* hybridization were collected after spawning from females by placing adult females in a sterile 150 mm diameter Petri dish and gently scraping eggs from their abdomens with a 3x2 inch index card. Embryos were raised until hatching in 150 mm Petri dishes at 26°C in 1X ERM and staged developmentally according to Iwamatsu (2004). Embryos were anesthetized with MS-222 (0.04% w/v) prior to fixation in 4% paraformaldehyde (PFA) overnight at 4°C, de-chorionated with fine forceps and scalpels, dehydrated in a graded series of methanol in phosphate buffered saline with 0.1% Tween 20 (PBT) and stored in 100% methanol at -20°C until use (Inohaya et al., 1995).

Whole-mount *in situ* hybridization assays were performed in 1.7 mL sterile centrifuge tubes containing ten embryos each according to a modification of Inohaya's and Takamatsu's methods (Inohaya et al., 1995, 1999; Takamatsu et al., 2007). All experiments used digoxigenin (DIG)-labeled sense and antisense riboprobes that were produced and purified according to Scemama et al. (2006). Sense riboprobes were used in control experiments to assess nonspecific

binding. Proteinase K digestions were performed at 28.5°C for 5 min for stages 18–20, 7 min for stages 21–24, 10–15 min for stages 25–30, and 20 min for stages 30 and beyond. Embryos were pre-hybridized for at least 2 hr at 65°C in 1 mL hybridization buffer (50% formamide, 5X sodium saline citrate (SSC), 50 mg/mL heparin, 0.1% Tween 20, 500 mg/mL torula RNA, pH 6.0), and hybridized in 250 mL hybridization buffer with 0.33–0.66 ng/mL DIG-labeled RNA probes for at least 16 hr. Four 1 mL post-hybridization washes were performed at 65°C for 1 hr each in 50% formamide, 2X SSC-0.1% Tween 20 (SSCT) followed by one wash in 1mL of 2X SSCT for 30 min at 65°C, two washes in 1mL of 0.2X SSCT for 1 hr each at 65°C, and two washes in 1 mL of PBT for 5 min each at room temperature (RT). Embryos were treated in a blocking solution of 1 mL of 5% sheep serum (Sigma, St. Louis, MO) in PBT for at least 90 min at RT. Embryos were then incubated overnight at 4°C in a solution containing 1 mL of 1:5000 anti-DIG antibody-AP (Roche, Indianapolis, IN) in PBT overnight at 4°C, washed 8X for 15 min each at RT in 1 mL of PBT to remove excess unbound antibody and then washed twice for 5 min at RT using 1 mL of 0.1M Tris, pH 9.5, 0.05M MgSO4, 0.1M NaCl, 0.1% Tween 20 (Alkaline phosphatase buffer). This wash buffer was removed and replaced with 0.5 mL of alkaline phosphatase buffer containing 0.4 mg/mL of nitro-blue tetrazolium chloride (Roche, Indianapolis, IN) and 0.2 mg/mL of 5-bromo-4-chloro-30-indolylphosphate p-toluidine salt (Roche, Indianapolis, IN) to visualize the location of probe binding and activity. Development of DIG- labeled probe signal, examination of embryos and digital photography of embryos were performed as described in Scemama et al. (2006).

In comparing medaka *Hox* PG2 gene expression patterns to those of other teleosts, morphological features, including the intermediate brain vesicle (IBV), midbrain/hindbrain boundary, rhombomeres (r), otic vesicles (OV), pectoral fins (PF), caudal region (CR) and

somites (s) within developing embryos were used as morphological landmarks. To define putative boundaries within the hindbrain times prior to the development of morphological landmarks, the zinc finger *egr2* gene (*krox20* ortholog), a marker for presumptive rhombomeres 3 and 5 (Voiculescu et al., 2001) was used for comparative in situ hybridization experiments. The expression pattern of medaka *egr2* has been described in detail by Kage et al. (2004).

Genomic sequence comparisons

The *hoxa2b* genes, introns and upstream *hoxa9b–hoxa2b* intergenic regions from *D. rerio* (accession number AL645795), *O. latipes* (accession number AB232919), *Astatotilapia burtoni* (accession number EF 594311) and *T. rubripes* accession number DQ481664) were compared using Vista (http://genome.lbl.gov; Frazer et al., 2004). The DNA regions included in the analysis encompassed the *hoxa2b* gene and 4000 bp of its upstream 5'-intergenic sequence. The Shuffle-Lagan option in Vista, which detects rearrangements and inversions, was used for the alignment of the sequences. The following parameters were selected in the presentation of the results: window of 100 bp, minimum conservation width of 100 bp and conservation identity of 70 %.

Results

Cloning and assignment of medaka *Hox* PG2 and *egr2* genes

Cloning of partial cDNA's obtained from RT–PCR amplification of total RNA prepared from 1 day post-fertilization (dpf) medaka embryos generated clones with inserts of three distinct lengths. Single clones representing each of the three insert lengths were designated either pOlaa2a (556 bp insert length), pOlaa2b (1008 bp insert length), or pOlab2a (617 bp insert length) based on their sequence relationships to previously characterized genomic sequences, respectively. Sequences corresponding to the inserts from pOlaa2a, pOlaa2b and pOlab2a were

99% similar to medaka genomic sequences corresponding to *hoxa2a* (GenBank accession number: AB207976) and *hoxb2a* (GenBank accession number: AB207991) and identical to *hoxa2b* (GenBank accession number: AB207985) (Kurosawa et al., 2006), respectively. Based on these results, we concluded that the cDNA inserts from pOlaa2a, pOlaa2b, and pOlab2a corresponded to complementary regions from medaka *hoxa2a*, *a2b* and *b2a* genes, respectively. Sequence alignment of the cDNA insert from pOlaa2b with the annotated genomic sequence of medaka *hoxa2b* (GenBank accession number: AB207985) revealed that the exon splice junction was misassigned in the annotated genomic sequence, which resulted in 10 bp of the bioinformatically deduced intron being included in exon 1 (Fig. 2-2). *In silico* translation of the re-annotated sequence revealed that the amino terminal 138 amino acid sequence was uninterrupted by translational terminators but included a well-conserved hexapeptide motif (EYPWMN) characteristic of those common to *Hox* PG2 genes. However, translation terminators were identified beginning in the region corresponding to exon 2 upstream of the putative homeodomain (Fig. 2-2). We interpreted these results as indicative that medaka hoxa2b was a transcribed pseudogene potentially capable of producing a translational product with a conserved hexapeptide motif but not a homeodomain. In addition to the cloning of the *Hox* PG2 genes and the *hoxa2b* pseudogene, we amplified and cloned a 353 bp medaka *egr2* partial cDNA (pOlaegr2). The cDNA insert sequence of pOlaegr2 was 99% similar to the striped bass *egr2* gene (Genbank accession number: DQ383280) and 83% similar to the zebrafish *egr2b* gene (Genbank accession number: BC081622), which provided evidence that the cloned cDNA in pOlaegr2 encoded a part of the medaka *egr2* gene.

Medaka *hoxa2a* and *b2a* gene expression analysis

We conducted whole-mount *in situ* hybridization analyses of medaka embryos with

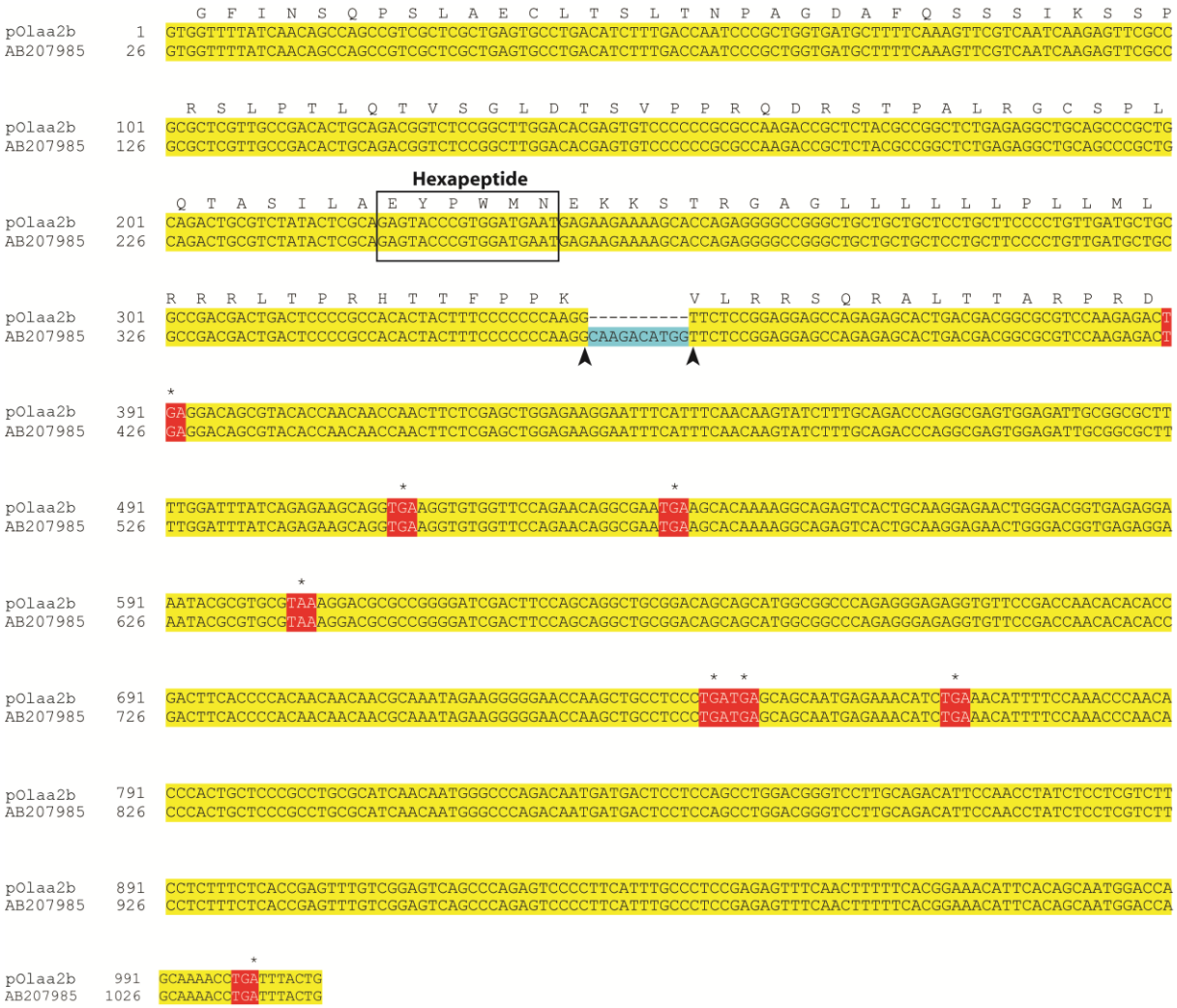


Fig. 2-2. Sequence comparisons between the cDNA sequence from clone pOlaa2b and the genomic DNA sequence reported in GenBank. Identical sequences are highlighted in yellow. Boxed region indicates the sequence encoding and corresponding to the hexapeptide. The arrow heads show the positions of the splice junctions. The sequence highlighted in blue shows the 10 bp of intronic DNA that was included in exon 1 as reported in GenBank accession number AB207985. The sequences highlighted in red and overlying asterisks (*) indicate terminator codons.

antisense DIG-labeled riboprobes to determine the expression patterns of *hoxa2a* and *b2a*. We further compared the medaka *hoxa2a* and *b2a* gene spatio-temporal expression patterns to those of zebrafish, tilapia, striped bass and tetrapods.

During early osteichthyan embryonic development, the *Hox* PG2 gene expression patterns have been shown to be generally established as faint stripes within the rostral aspect of the hindbrain, after which they quickly expand to include more caudal regions of the hindbrain. *In situ* hybridization of medaka whole-mount embryos with antisense riboprobes directed toward *hoxa2a* and *b2a* revealed that the patterns of *hoxa2a* and *b2a* gene expression early in development were largely consistent with these general osteichthyan patterns. This was reflected by their expression as single stripes in presumptive r2/r3 (*hoxa2a*) or r3/r4 (*hoxb2a*) at the one somite stage (stage 18/19-data not shown) followed by the caudal expansion of *hoxa2a* expression into r4 and r5 at stage 19/20 (3 somites) of development (Fig. 2-3A). By comparison, *hoxb2a* expression remained restricted to r3 and r4 (Fig. 2-3C). The rostral boundaries of *hoxa2a* and *b2a* hindbrain expression observed during the earliest stages of medaka development were representative of the rostral-most boundaries at any time during hindbrain development and were consistent with identical rostral hindbrain expression boundaries observed among *Hox* PG2 orthologs of other vertebrates (Sham et al., 1993; Vesque et al., 1996; Vieille-Grosjean et al., 1997; Prince et al., 1998; Pasqualetti et al., 2000; Scemama et al., 2002, 2006; Baltzinger et al., 2005; Le Pabic et al., 2007), which suggests that the genetic mechanisms specifying the rostral limits of *hoxa2* and *b2* hindbrain expression have been conserved throughout the Osteichthyes.

The foregoing pattern observed during the establishment phase of osteichthyan *Hox* PG2 gene expression was generally followed by post-establishment phase caudal expansion of expression to include r6 and r7 of the hindbrain with concomitant expression in cranial neural

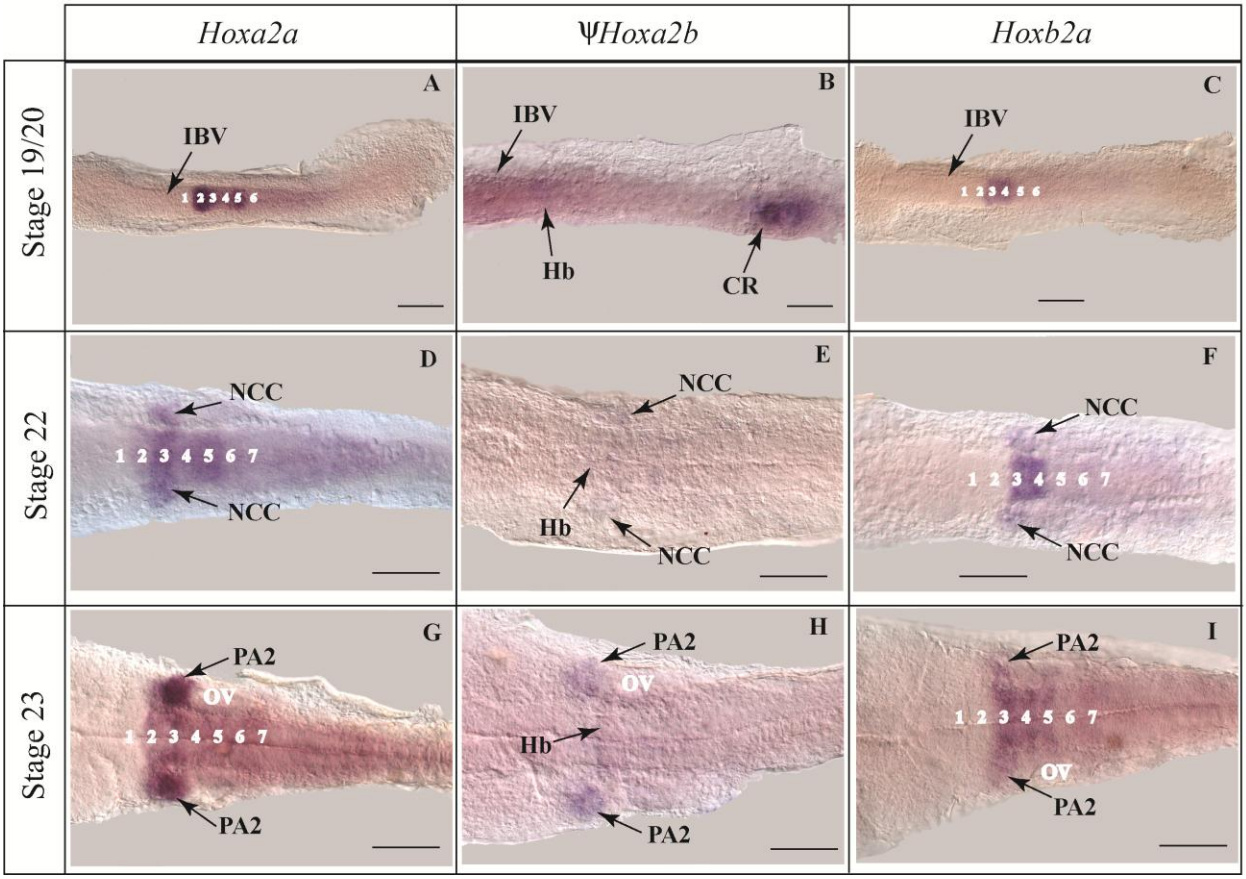


Fig. 2-3. Whole-mount *in situ* hybridization analysis of medaka *hoxa2a* (A, D, G), *ψhoxa2b* (B, E, H) and *hoxb2a* (C, F, I) gene expression at stages 19/20 (3s) (A–C), 22 (9s) (D–F), and 23 (12s) (G–I). All embryos were mounted with their anterior sides to the left and their dorsal sides toward the reader. Rhombomere numbers are indicated by either white or black numbers on the embryos. CR, caudal region; Hb, hindbrain; IBV, intermediate brain vesicle; NCC, neural crest cells; OV, otic vesicle; PA2, 2nd pharyngeal arch. Scale bars equal 0.1 mm.

crest cells (CNCCs) that migrated ventro-laterally from the hindbrain into the developing pharyngeal arches. In keeping with these characteristic osteichthyan patterns, post-establishment phase medaka *hoxa2a* gene expression expanded caudally to include r6 and r7 by stage 22 (nine somites, Fig. 2-3D) and simultaneously extended into the region corresponding to the pathway of CNCC migration emanating from r4 and into PA2. Further medaka *hoxa2a* expression in r2–r7 was continuous up to stage 39 (9 dpf), after which expression of this gene was no longer detectable in the hindbrain (Figs. 2-3G, 2-4A and D, 2-5D and data not shown). At stage 22, *hoxb2a* expression was also detected in the CNCCs migrating out of the hindbrain and into PA2, whereas its expression in the hindbrain remained restricted to the r3/r4 region (Fig. 2-3F). Later at stage 23 (12 somites) *hoxb2a* expression was observed in all the rhombomeres, with the exception of r6 (Fig. 2-3I). This pattern changed by stage 25/26 (50–54 hpf), such that expression was observed in all rhombomeres, including r6 (Fig. 2-4C), where it remained continuously expressed until around stage 39, after which expression became undetectable (Figs. 2-4F, 2-5F and data not shown). A comparison of *hoxa2av*and *b2a* expression among medaka and othervteleosts revealed that medaka *hoxa2a* expression patterns were most similar to tilapia and striped bass whereas *hoxb2a* patterns, particularly the absence of expression in r6 at stage 23, were most similar to zebrafish at a comparable period of development.

Teleost *Hox* PG2 gene expression patterns in migratory CNCCs are always observed initially in the dorsal-most aspect of PA2, from where expression expands to encompass more ventral aspects of this arch before the onset of a cascading dorso-ventral pattern in the more posterior arches. A consequence of this differential rostro-caudal migration pattern results in rostral CNCCs migrating to their specific arch destinations earlier then more caudal CNCCs in an apparent rostro-caudal, dorso-ventral wave of *Hox* expression. We observed that the overall

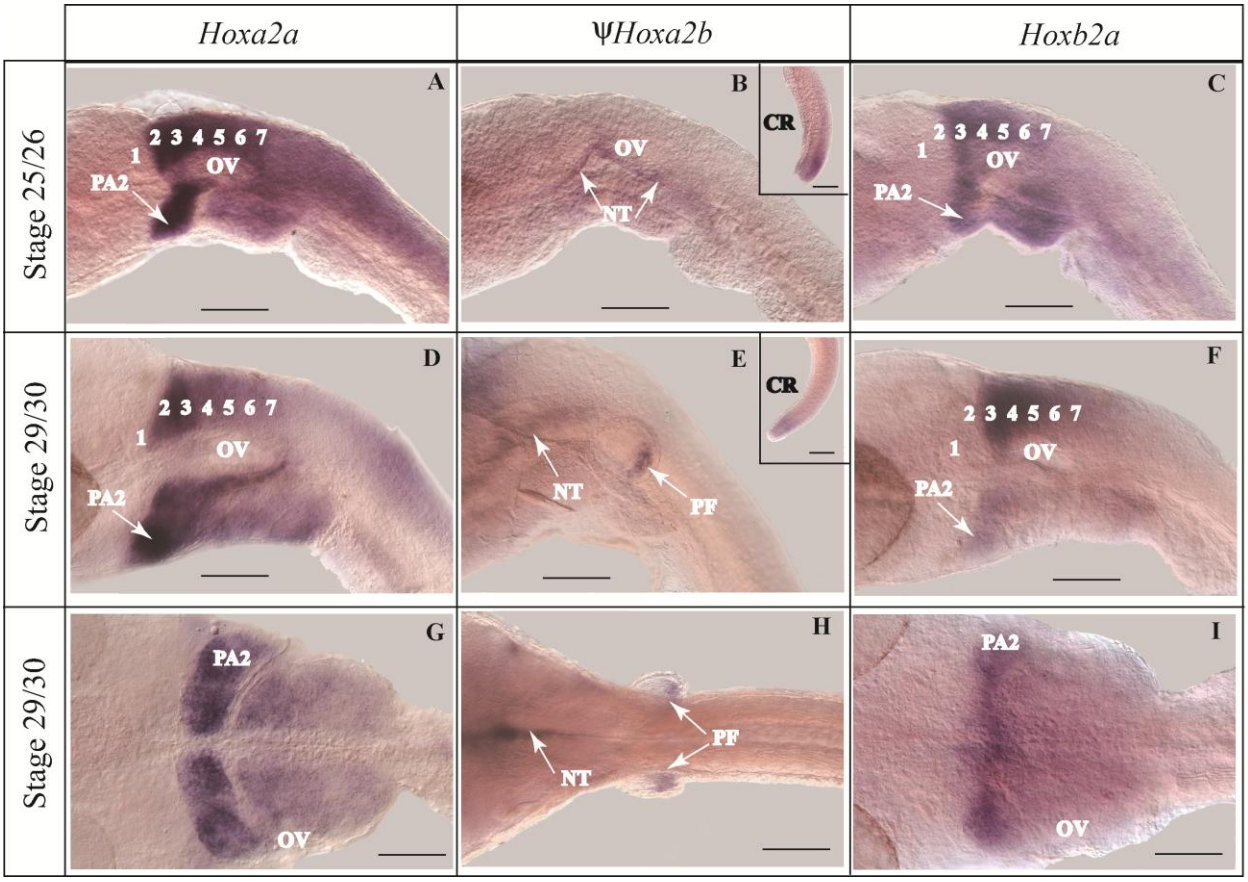


Fig. 2-4. Whole-mount *in situ* hybridization of medaka *hoxa2a* (A, D, G), *ψhoxa2b* (B, E, H) and *hoxb2a* (C, F, I) gene expression at stages 25/26 (50–54 hpf) (A–C) and 29/30 (74–82 hpf) (D–I). (A–F) embryos were mounted with their anterior sides to the left and their lateral sides toward the reader. (G–I) embryos were mounted with their anterior sides to the left and their ventral sides toward the reader. Rhombomere numbers are indicated by white numbers on the embryos. Note that arrowheads in panels A and D point to the finger-like *hoxa2a* expression pattern in the dorsal domain of PA2. Similarly, arrowheads in panel G point to a pair of thumb-like *hoxa2a* expression domains in the ventral region of PA2. CR, caudal region; NT, neural tube; OV, otic vesicle; PA2, 2nd pharyngeal arch; PF, pectoral fin buds. Scale bars equal 0.1mm.

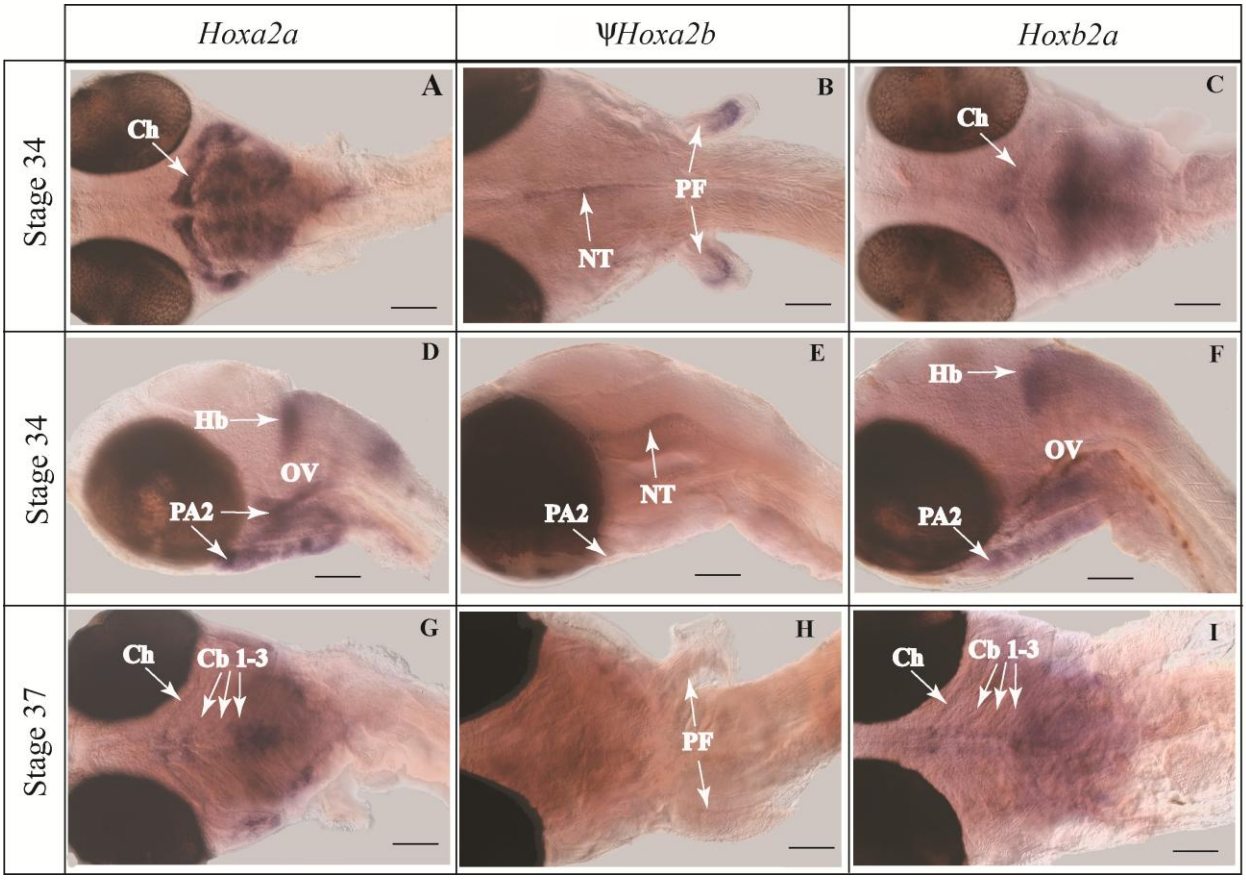


Fig. 2-5. Whole-mount *in situ* hybridization of medaka *hoxa2a* (A, D, G), *ψhoxa2b* (B, E, H) and *hoxb2a* (C, F, I) gene expression at stages 34 (121 hpf) (A–F) and 37 (7 dpf) (G–I). (A–C, G–I) embryos were mounted with anterior sides to the left and their ventral sides toward the reader. (D–F) embryos were mounted with their anterior sides to the left and their lateral sides toward the reader. Cb, ceratobranchial; Ch, ceratohyal; Hb, hindbrain; NT, neural tube; PA2, 2nd pharyngeal arch; PF, pectoral fin buds. Scale bars equal 0.1mm.

patterns of medaka *hoxa2a* and *b2a* expression in migrating CNCCs follow a similar pattern of rostro-caudal, dorso-ventral expression changes in which both medaka *hoxa2a* and *b2a* expression were observed in the CNCCs populating PA2 earlier then expression in the more posterior arches at stage 23 (Fig. 2-3G and I). By stage 25/26 both genes were expressed throughout PA2 and the posterior arches (Fig. 2-4A and C) and remained expressed in these compartments until their eventual decline to undetectable expression levels after stage 37 (7 dpf) (Figs. 2-4D, F, G and I, 2-5A, C, D, F, G and I and data not shown). In comparison to CNCC-specific *Hox* PG2 gene expression patterns in other teleosts, the caudal expansion of medaka *hoxa2a* and *b2a* expression into the posterior arches resembled the caudal expansion of *hoxa2a*, *a2b* and *b2a* expression in the posterior arches of striped bass/tilapia or zebrafish *hoxa2b* at similar developmental stages (Scemama et al., 2006; Le Pabic et al., 2007). The lone exception to this pattern of caudal expansion was observed for the zebrafish *hoxb2a* gene, which remained restricted to PA2 (Hunter and Prince, 2002).

In addition to the pharyngeal arch patterns described above, a distinctive *hoxa2a* expressing caudal projection that extended from the dorsal-most aspect of PA2 to a posterior location immediately ventral to the developing otic vesicle was observed beginning at stage 25/26 (Fig. 2-4A). This projection appeared to expand caudally and assumed the form of a thin slightly curved finger-like projection that terminated under the posterior-most margin of the otic vesicle by stage 29/30 (74–82 hpf) (Fig. 2-4D). This pattern was observed to persist in a slightly altered shape up until stage 37, after which it diminished to an undetectable level (Fig. 2-5A, D and G). It is interesting to note that this *hoxa2a* expression pattern foreshadows, both spatially and temporally, the pattern of the cartilaginous elements that generate the hyomandibular. Similarly, in the ventral domain of PA2 in medaka we observed a distinct pattern of *hoxa2a*

expressing CNCCs at stage 29/30 that projected in a thumb-like shape from the lateral-most region of PA2 in a rostro-medial direction toward the ventral midline of the medaka embryo (Fig. 2-4G). Beginning at stage 32/33 (53–58 hpf) this expression pattern changed, such that the region of *hoxa2a* more closely resembled a shape comparable to the ceratohyal (Fig. 2-5A and data not shown). A similar pattern in which *hoxa2a* expression mirrored the shape of the ceratohyal was also observed for striped bass (Scemama et al., 2006). Moreover, the pattern observed for medaka *hoxa2a* expression in the dorsal aspect of PA2 that mirrored the structure of the hyomandibular was also observed for *hoxa2a* genes from tilapia and striped bass and the *hoxa2b* gene from zebrafish (Hunter and Prince, 2002; Scemama et al., 2006; Le Pabic et al., 2007), suggesting that the *hoxa2* genes in all four species function similarly in patterning the hyomandibular development.

During teleost and tetrapod embryogenesis, it has been observed that *Hox* PG2 gene expression diminishes in the pharyngeal arch compartments during chondrogenesis but that persistence of *Hox* PG2 gene expression into the chondrogenic phase of PA2 development correlates with PA2-specific selector gene activity. We observed that both medaka *hoxa2a* and *b2a* were expressed into the chondrogenic phase in PA2, after which their expression decreased as arch-derived cartilaginous structures developed (Fig. 2-5A, C, D, F, G and I). A similar pattern of persistent expression into the chondrogenic stage of PA2 development was observed for zebrafish *hoxa2b* and *b2a* (Hunter and Prince, 2002). By comparison, in striped bass and tilapia only *hoxa2a* was expressed into the chondrogenic phase of PA2 development (Scemama et al., 2006; Le Pabic et al., 2007), which suggests that teleosts with three *Hox* PG2 genes may have partitioned the function of these genes differently than teleosts with only two *Hox* PG2 genes.

Medaka *Hox* PG2 gene hybridization competition experiments

As medaka is more closely related to striped bass and tilapia than to zebrafish, and because neither striped bass nor tilapia showed detectable *hoxb2a* expression during the chondrogenic phase of PA2 development, we were concerned that the medaka *hoxb2a* expression observed during this time was the result of cross-hybridization between the medaka *hoxb2a* riboprobes and *hoxa2a* mRNA. To examine the potential cross-hybridization between medaka *hoxa2a* and *b2a* riboprobes and their respective mRNA targets, we conducted whole embryo *in situ* hybridization competition experiments using labeled antisense riboprobes and 2-, 5- and 10-fold molar excess of *in vitro* transcribed homologous or paralogous medaka unlabeled competitor antisense transcripts as described in ‘‘Materials and Methods.’’ The results showed that only whole-mount embryos hybridized with labeled *hoxa2a* or *b2a* antisense riboprobes in the presence of their respective homologous *in vitro* transcribed unlabeled antisense RNAs showed reduced expression signals. Further, the extent of reduction in the hybridization signal was directly proportional to the amount of unlabeled competitor RNA added to the hybridization reaction, indicative that neither of these two *Hox* PG2 gene-derived probes cross-hybridize (data not shown). Based on these results, we concluded that both *hoxa2a* and *b2a* from medaka were expressed during the chondrogenic phase of PA2, which is a pattern more similar to zebrafish than for the more closely related acanthopterygians striped bass and tilapia, both of which possess three functional *Hox* PG2 genes.

Medaka *ψhoxa2b* pseudogene expression analysis

Based on the observation that medaka *ψhoxa2b* is a transcribed pseudogene and that all functional teleost *hoxa2b* genes characterized to date are expressed in common embryonic compartments, e.g. the rhombomeres and pharyngeal arches (Prince et al., 1998; Hunter and

Prince, 2002; Amores et al., 2004; Scemama et al., 2006; Le Pabic et al., 2007), we were interested to determine whether the medaka *ψhoxa2b* pseudogene was expressed in the same locations of the developing embryo as functional *Hox* PG2 genes in other species. Unlike the expression of its functional medaka paralogs, *hoxa2a* and *b2a*, or the expression patterns exhibited by its functional orthologs from other species, medaka *ψhoxa2b* showed barely detectable expression in the hindbrain and the pharyngeal arches at any time during embryogenesis. As an indication of the weakness of *ψhoxa2b* expression in the hindbrain and other compartments typical of canonical *Hox* PG2 genes, we always observed expression of medaka *hoxa2a* or *b2a* within an hour after the addition of the chromogenic substrate during development of *in situ* hybridization reactions. However, even the very weak signals observed for medaka *ψhoxa2b* expression in the hindbrain and pharyngeal arches always required several days of incubation after substrate addition, despite comparable specific activities among the riboprobes used for all *Hox* PG2 in situ hybridization studies (Fig. 2-3B, E and H).

In comparison to the weak signal generated from the hindbrain and pharyngeal arches, noncanonical robust expression of medaka *ψhoxa2b* was observed in three regions, including the caudal-most area of the embryonic trunk, the distal mesenchyme of the developing fin buds and the ventral aspect of the neural tube extending from the level of the second pharyngeal arch caudally to the level of the caudal-most pharyngeal arches. Expression in the caudal trunk region was detected between stages 19/20 and 25/26 (Figs. 2-3B; 2-4B, insert) before becoming fainter by stage 29/30 (Fig. 2-4E, insert) and undetectable by stage 34 (data not shown), whereas expression in the ventral region of the neural tube and developing fin buds were first observed at stage 25/26 or 29/30, respectively, and ended in both regions at stage 37 (Figs. 2-4B, E, H; 2-5B, E, H).

Discussion

Medaka *Hox* PG2 gene complement

The results from cDNA cloning, DNA sequencing and comparative genomics reported in this study showed that the medaka *Hox* PG2 gene complement was composed of two functional genes, *hoxa2a* and *b2a*, each of which were expressed in canonical *Hox* PG2 gene expression domains, including the rhombomeres of the hindbrain and the developing pharyngeal arches. We also demonstrated that the genomic sequence complementary to the medaka *hoxa2b*-encoding cDNA possessed a series of translational termination codons in regions that corresponded to the second exon, which likely rendered the translation product of the *hoxa2b* mRNA devoid of a functional homeodomain and inactive as a classical *Hox* gene. In further support of its pseudogene status, whole-mount *in situ* hybridization using *ψhoxa2b* DIG-labeled riboprobes revealed that the *ψhoxa2b* mRNA was expressed at levels comparable to other *Hox* PG2 genes only in noncanonical expression domains, which included the caudal-most region of the embryonic trunk, ventral-most aspect of the neural tube and pectoral fin buds (Figs. 2-3B, 2-4B, E and H, 2-5B and E). The expression of *ψhoxa2b* mRNA in these noncanonical expression domains suggests that the *cis*-regulatory sequences controlling expression of the transcribed pseudogene may have diverged from their nonpseudogene counterparts in other species, possibly as a result of relaxed selection after mutational inactivation of the functional *hoxa2b* gene in the lineage leading to the medaka. As expected, comparative genomic analyses conducted in our laboratory have shown putative medaka regulatory regions corresponding to phylogenetically conserved consensus osteichthyan regulatory elements from the *hoxa2b- a9b* intergenic region are more divergent from other teleost species than would be expected based on the taxonomic relationships to these species. This sequence divergence is particularly evident in the region

corresponding to the proximal promoter but also seems to be characteristic of other putative regulatory regions as well (see Fig. 2-6). We interpret this result as evidence of relaxed selection after mutational inactivation of a functional *hoxa2b* gene in the medaka lineage.

Although medaka *ψhoxa2b* does not function as a classical *Hox* PG2 gene, the exon 1-specificvhexapeptide sequence of the medaka *ψhoxa2b* transcript may be functional, because its amino acid sequence encodes a hexapeptide motif, EYPWMN that is highly similar to motifs from other *Hox* PG2 genes (data not shown). The only difference in the amino acid sequence of the medaka *ψhoxa2b* hexapeptide domain compared with other teleost *hoxa2b* hexapeptides is a substitution of asparagine (N) for lysine (K) at the carboxy-terminal end of the hexapeptide (data not shown). Four of the five conserved amino acids (YPWM) have been shown to be integral for binding with the hydrophobic pocket of the TALE protein, Pbx (Piper et al., 1999). LaRonde-LeBlanc and Wolberger (2003) recently demonstrated that the hexapeptides of numerous homeodomain proteins were able to bind to their cofactors despite variability in the hexapeptide sequence, including variation in the C-terminal amino acid. These results suggest that the hexapeptide sequence encoded by medaka *ψhoxa2b* may be functional and capable of binding *Hox*-specific hexapeptide cofactors. It is interesting to speculate that binding of the medaka *choxa2b*-derived hexapeptide to *Hox*-binding co-activators could influence *Hox* activity and affect morphological specification in the embryonic compartments in which the putative *ψhoxa2b* -encoded protein is expressed. We are currently conducting oligonucleotide morpholino-mediated knockdowns of medaka *ψhoxa2b* expression to determine whether changes in expression of the putative hexapeptide-encoding protein influences embryonic morphological specification.

Osteichthyan hindbrain *Hox* PG2 gene expression patterns

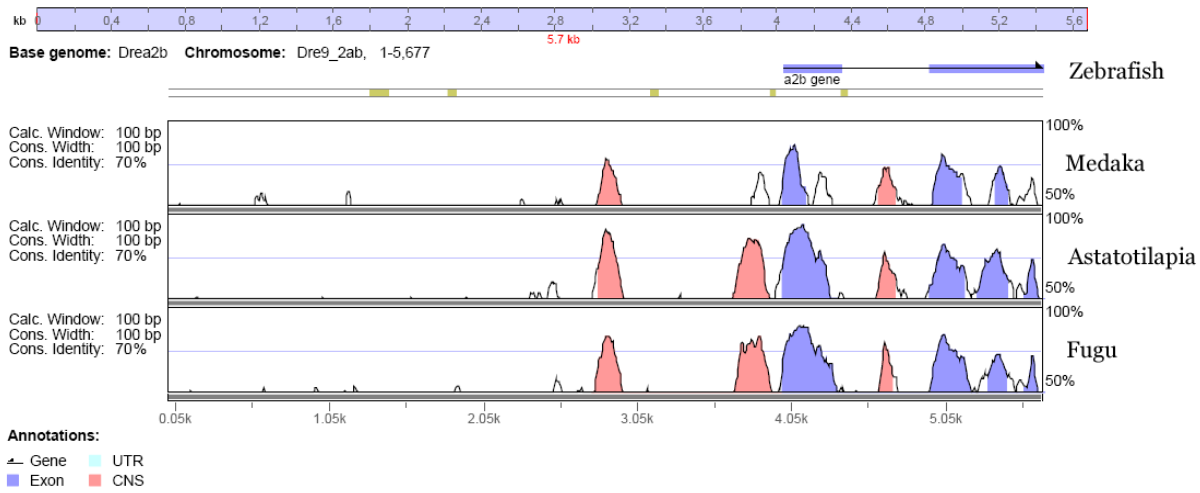


Fig.2- 6. mVista sequence alignment plot for *Danio rerio*, *Oryzias latipes*, *Astatotilapia burtoni*, and *Takifugu rubripes hoxa2b* genes, intron, and intergenic DNA. The regions encompassing the complete *hoxa2b* gene including the intron and 4,000 bp of *hoxa9b-hoxa2b* intergenic regions were used in the analysis. Peaks shown within each frame represent the levels of sequence similarity in a 100 bp window. Blue-shaded peaks correspond to exons whereas red-shaded peaks correspond to conserved noncoding sequences. The red-shaded peaks approximately 1 kb upstream of the exons of the *hoxa2b* genes correspond closely to functionally tested regulatory regions that direct teleost and tetrapod *Hoxa2* expression in r3/5, PA2 and in the posterior arches (Maconochie et al., 1999, 2001).

In addition to the characterization of the medaka *Hox* PG2 gene complement and the *ψhoxa2b* gene expression pattern, we examined the spatio-temporal patterns of *hoxa2a* and *b2a* expression throughout medaka embryonic development. Both the medaka *hoxa2a* and *b2a* genes were expressed earliest at sites that were generally restricted to a region encompassing r2–r5 and included the rostral limits of expression at the r2/3 or r3/4 boundaries, respectively. This early establishment phase pattern of expression appears to be relatively common among osteichthyans, all of which have very similar spatio-temporal expression patterns. These strong similarities in expression patterns may reflect widespread evolutionary conservation among the *cis*-acting regulatory elements responsible for controlling early rostral rhombomere-specific *Hox* PG2 hindbrain expression. Support for this view is provided by genomic sequence comparisons that have shown DNA sequence elements identified in the intron and upstream of the medaka *hoxa2a* coding sequence that are similar to evolutionarily conserved *cis*-acting regulatory elements corresponding to the r3/5, r2 and r4 enhancer modules that were initially characterized in tetrapods and subsequently identified in diverse teleost species (Tümpel et al., 2006; Lampe et al., 2008).

The functional nature of the teleost *Hox A* cluster PG2 rhombomeric elements have been confirmed through transgenic assays in chick embryos, wherein reporter gene constructs containing elements located upstream of teleost *Hox A* cluster PG2 genes directed expression in r3/5, and regulatory elements located within exon 2 and the intron directed expression in r2 and r4, respectively (Tümpel et al., 2006). Therefore, the similarities in spatio-temporal expression patterns and sequence suggest that the conserved medaka *hoxa2a*-specific sequences function similarly to their counterparts in tetrapods and other teleosts by directing *hoxa2a* expression in r2–r5 during early hindbrain development. Similarly, the intergenic region upstream of medaka

hoxb2a shared sequence elements shown to be required for directing *hoxb2a* expression in r3/5 and r4 of other osteichthyans (Scemama et al., 2002). These sequence elements included sites specific for Krox20, a transcription factor that, in part, has been shown to direct tetrapod *Hoxa2* and *b2* gene expression in r3/5 (Sham et al., 1993; Nonchev et al., 1996a,b; Vesque et al., 1996; Maconochie et al., 2001), and Hox/Pbx and Meis, transcription factors integral for directing tetrapod *Hoxa2* and *b2* gene expression in r4 (Ferretti et al., 2000; Tümpel et al., 2006, 2007).

The conserved regulatory sequences and expression patterns in the rostral hindbrain of teleost *Hox* PG2 genes are very likely to have direct functional consequences related to rhombomeric identity and cranial motor neuron specification. In mouse, knockout experiments have shown that *Hoxa2* and *b2* control the segmentation of the anterior hindbrain and the axonal guidance of the Vth and VIIth cranial motor nerve axons out of r2/r4 and specification of the somatic motor component of the VIIth cranial nerve exiting r4, respectively (Barrow and Capecchi, 1996; Gavalas et al., 1997, 2003; Davenne et al., 1999; Barrow et al., 2000). Analyses of homozygous *Hoxa2*/*b2* null mutant mice have shown the absence of rhombomeric boundaries between r1–r4 but not between r4 and more caudal rhombomeres, which has been interpreted as evidence of synergy between mouse *Hoxa2* and *b2* gene products in the specification of proper rostral rhombomere segmentation (Davenne et al., 1999). As the exit points of the Vth and VIIth cranial motor nerves are conserved between mouse and zebrafish (Chandrasekhar, 2004) and the rostral hindbrain expression of teleost *Hox A* and *B* cluster PG2 genes are similar to those of mouse *Hoxa2* and *b2*, it is possible that teleost *Hox* PG2 genes function similarly to those of their mouse orthologs in patterning the rostral hindbrain. We hypothesize this would also be the case for medaka because it shows conserved expression of *hoxa2a* and *b2a* genes in r2–r5. We are testing this hypothesis by conducting antisense morpholino knockdown experiments of *hoxa2a*

and *b2a* in medaka and monitoring rhombomere boundaries through *in situ* hybridization with *mariposa* antisense RNA probes, a marker that has been shown to be expressed specifically at inter-rhombomere boundaries (Moens et al., 1996).

A comparison of the *hoxb2a* expression patterns in the caudal-most rhombomeres, r6 and r7, of osteichthyans revealed considerable evolutionary divergence, indicative of an uncoupling between the regulation of rostral and caudal *hoxb2a*-specific rhombomere expression. We observed that medaka *hoxb2a* was first expressed exclusively in r7 during early hindbrain development (stage 23) but expanded to include both r6 and r7 beginning at stage 25/26 and continuing until the cessation of expression in the hindbrain. By contrast, zebrafish *hoxb2a* expression was observed in r7 but not in r6 during early hindbrain development (Prince et al., 1998); however, *hoxb2a* expression in striped bass was absent from both r6 and r7 throughout development (Scemama et al., 2002). Although the functional consequences of divergent *hoxb2a* expression in r6 and r7 are unknown, they are most likely owing to genomic sequence divergence in the enhancer modules responsible for directing *hoxb2a* expression in these rhombomeres. To date, the regulatory elements that direct *Hox* PG2 gene expression in r6 and r7 have not been identified. However, the failure to identify such sequences using conventional comparative genomic analyses is not unexpected. *In silico*-based methods for identification of putative *cis*-regulatory elements commonly rely on conservation of sequences among evolutionarily divergent species that share similar expression patterns. The extensive divergence of *hoxb2a* expression patterns in r6 and r7 suggests that identification of putative *cis*-regulatory elements based on conservation of sequences among divergent taxa using comparative genomic methods may prove fruitless because in this case we have shown that taxonomically related species differ significantly in their expression patterns. To address these problems, we are

currently conducting comparative genomic sequence analyses of evolutionarily divergent teleosts that share similar expression patterns to determine if these taxa share common regulatory elements (Stellwag and Scemama, manuscript in preparation).

Osteichthyan neural crest and pharyngeal arch *Hox* PG2 gene expression

Our results from whole-mount *in situ* hybridization of medaka *Hox* PG2 gene expression within the cranial neural crest revealed that expression of both *hoxa2a* and *b2a* commenced simultaneously in the CNCCs migrating ventro-laterally out of r4 and into the developing second pharyngeal arch primordium. A virtually identical pattern of expression was observed for the two *Hox* PG2 genes from zebrafish (Prince et al., 1998) but not from acanthopterygians with three functional *Hox* PG2 genes (Le Pabic et al., 2007). In these acanthopterygians, either *hoxa2a* (tilapia) or *hoxb2a* (striped bass) was initially expressed alone, followed later by expression of the other *Hox* PG2 genes (Le Pabic et al., 2007). At even later times, we observed that expression persisted until the chondrogenic phase of PA2 development for both *Hox* PG2 genes from medaka and zebrafish, but only for *hoxa2a* from the striped bass and tilapia (Hunter and Prince, 2002; Scemama et al., 2006; Le Pabic et al., 2007). Given the closer taxonomic affiliation among medaka, tilapia and striped bass in relation to zebrafish, it was difficult to reconcile the overall similarity of medaka and zebrafish *Hox* PG2 gene expression patterns in r4-derived migratory CNCCs and their longer persistence during PA2 development. A possible explanation for these seemingly contradictory results is that the loss of *Hox A* cluster-derived PG2 genes in the medaka and zebrafish lineages may have affected *Hox* PG2-related CNCC expression. The potential functional consequences of these altered expression patterns are significant. Hunter and Prince (2002) and Baltzinger et al. (2005) observed that persistence of *Hox* PG2 gene expression in PA2 of zebrafish and African clawed frog correlated with their

function as homeotic selector genes in the second arch. Based on these observations, they hypothesized that persistent expression of *Hox* PG2 genes was required for PA2-directed selector gene activity. Assuming their hypothesis was correct, and based on the persistent expression of medaka *hoxa2a* and *b2a* in PA2 until second arch chondrogenesis, it would be consistent to hypothesize that the two medaka *Hox* PG2 genes, like those from zebrafish, function redundantly as selector genes of PA2 identity, whereas only a single *Hox A* cluster PG2 gene (*hoxa2a*) functions in this manner in tilapia and striped bass. We are presently conducting a functional genetic characterization of the medaka *Hox* PG2 genes using antisense morpholino-mediated genetic knockdowns and misexpression studies to test this hypothesis.

In addition to expression of *Hox* PG2 genes in CNCCs localized in PA2, we observed a spatial pattern of medaka *hoxa2a*-expressing CNCCs in the dorsal and ventral domains of PA2 that was similar to the orthologous and co-orthologous *hoxa2*-expressing cells in PA2 of tilapia/striped bass and zebrafish, respectively. Specifically, the spatial distribution of the *hoxa2*-expressing cells within the dorsal and ventral domains of PA2 of these species closely mirrored the morphology of cartilaginous elements that subsequently form the hyomandibular and ceratohyal, respectively. These observations suggest that the expression of medaka *hoxa2a* in the CNCCs, like the *Hox A* cluster PG2 genes of other teleosts, functions to pattern the geometry of PA2-derived cartilaginous elements. An extensive body of evidence has shown that spatially restricted signals emanating from endodermal and ectodermal sources surrounding PA2-localized CNCCs are required to pattern the pharyngeal skeleton through their action on *Hox* PG2 expressing CNCCs (Couly et al., 2002; Creuzet et al., 2005; Crump et al., 2006). Recent evidence reported by Crump et al. (2006) showed that in *moz* null mutants of zebrafish, which are deficient in both *hoxa2b* and *b2a* gene activity, the CNCCs that normally responded to first

pharyngeal pouch-derived endodermal signals to form the rostral region of the hyomandibular and the symplectic became refractory to these signals and failed to form their normal PA2 skeletal structures. Instead, the absence of *Hox* PG2 gene activity resulted in both the failure of the dorsal CNCCs in PA2 to chondrify, and lead to the re-specification of the intermediate CNCCs in PA2, which resulted in a duplicate palatoquadrate, adjacent to, and in the same dorso-ventral position as the wild-type palatoquadrate of PA1 (Crump et al., 2006). These results showed that persistent *Hox* PG2 gene expression in PA2-localized CNCCs was required for the response of these cells to patterning signals arising from specific locations surrounding PA2. It further showed that absence of *Hox* PG2 gene expression reprogrammed cells to be both refractory to these signals but receptive to extrinsic epithelial signals from PA1 (Crump et al., 2006). Based on Crump et al.'s (2006) results, and our observations concerning the close relationship between the spatial distribution of *hoxa2a*-expressing CNCCs and the morphology of the hyomandibular and ceratohyal, it seems reasonable to conclude that the shapes of these bones are determined by the combined action of persistent *Hox* PG2 gene expression within PA2-localized CNCCs that respond to extracellular signals emanating from tissues surrounding the pharyngeal arch mesenchyme. Given the similarities of *Hox* PG2 gene expression patterns in PA2 among osteichthyans, it appears that a combination of diffusible extracellular signals adjacent to the arch and persistent *Hox* expression in CNCCs are necessary for morphological patterning of cartilaginous elements in the hyoid arch and have been evolutionarily conserved among the Osteichthyes.

Despite sharing similar *Hox* PG2 gene expression domains in PA2 with other osteichthyans, medaka and other members of the Beloniformes have evolved apomorphic differences in the skeletal elements commonly specified by the 2nd and posterior pharyngeal

arch compartments. Of particular interest in relation to this study was the absence of the interhyal bone in medaka. This cartilaginous element serves as an articulation point between the 2nd pharyngeal arch-derived ceratohyal and the hyosymplectic and allows for increased buccopharyngeal expansion in teleosts during feeding (Schaeffer and Rosen, '61). Evidence from Schilling and Kimmel (1997) showed that in the zebrafish, the ceratohyal, symplectic and hyomandibular begin to chondrify as three separate chondrogenic foci. This occurs within a single hyoid precartilage condensation along the ventro-lateral region of PA2 within a 4 hr window (53–57 hpf) and prior to the commencement of interhyal chondrification (Schilling and Kimmel, 1997). The ceratohyal of zebrafish subsequently elongates medially toward the ventral midline of the zebrafish head (Schilling and Kimmel, 1997). By 68 hpf, as the symplectic and hyomandibular chondrifications fuse to one another to form the hyosymplectic, the interhyal begins to chondrify between the lateral end of the ceratohyal and the fusion point of the symplectic and hyomandibular (Schilling and Kimmel, 1997; Kimmel et al., 1998). By contrast, Langille and Hall (1987) showed that in medaka at developmental stage 30 (stage 34 in Iwamatsu (2004)), the ceratohyal begins to chondrify at the ventral midline behind the eyes and chondrogenesis extends latero-caudally where the chondrifying ceratohyal eventually contacts and articulates directly with the developing hyosymplectic. This pattern of PA2 cartilage development provides evidence of a marked contrast between the mechanisms of PA2 chondrogenesis in medaka and zebrafish. As we did not observe any obvious differences for the PA2-directed medaka *Hox* PG2 gene expression patterns in comparison to zebrafish, this contrast between the dynamics of medaka and zebrafish 2nd arch chondrogenesis suggests these structures and their shapes are influenced by one of several possible mechanisms: either they are controlled by differential sensitivities to diffusible signals of *Hox* expressing CNCCs in these

two species or by species-specific qualitative/quantitative differences in diffusible signals, or a combination of these factors.

In addition to their expression in PA2, we have shown that medaka *Hox* PG2 genes, like their strict orthologs from tilapia and striped bass are expressed in the posterior arches, PA3-7 (Scemama et al., 2006; Le Pabic et al., 2007). Our results showed that medaka *Hox* PG2 gene expression persisted in PA3-7 until the onset of chondrogenesis, which is similar to zebrafish *hoxa2b* but strikingly different from zebrafish *hoxb2a*, the expression of which was undetectable in the posterior arches (Hunter and Prince, 2002). Dual knockdowns of zebrafish *Hox* PG2 genes failed to affect posterior arch development (Hunter and Prince, 2002), which suggests that neither gene plays a role in the development of this region. By comparison, single gene knockdowns of tilapia *Hox* PG2 genes, each appear to affect the architecture of the bony elements derived from the posterior pharyngeal arches (Le Pabic et al., 2008). The general similarity in *hoxb2a* expression patterns in the posterior arches of medaka and tilapia suggests that their *hoxb2a* genes may function in a related fashion in posterior arch specification and that *Hox* PG2 gene activity may not be restricted exclusively to PA2 in the Osteichthyes. If this is the case, it points to a possible interaction of *Hox* PG2 genes with other *Hox* genes that are expressed in the posterior arches, such as *hoxb3a*, *b4a* and *b5a* (Miller et al., 2004).

Evolution of osteichthyan *Hox* PG2 genes

Phylogenetic reconstructions that incorporate a whole genome duplication event in the actinopterygian stem lineage at the incipient stage of teleost evolution support an ancestral *Hox* PG2 gene complement that includes four genes; *hoxa2a*, *a2b*, *b2a* and *b2b* (Fig. 2-1). A comparison of the *Hox* PG2 gene composition among extant teleosts, including medaka, shows that only a single gene, *hoxb2b*, appears to have been lost in every lineage, indicative that this

gene was lost relatively early after the actinopterygian-specific total genome duplication. Aside from the wholesale loss of *hoxb2b*, two teleost *Hox A* cluster genes, *hoxa2a* or *a2b*, appear to have been lost later during teleost evolution and independently in two separate lineages. The deterioration of *hoxa2a* and *hoxa2b* to pseudogenes in the lineages leading to zebrafish or medaka, respectively, appear characteristic of ostariophysans, in the case of *hoxa2a*, or to a restricted subset of species within the acanthopterygians, in the case of *hoxa2b*. The independent post-genome duplication loss of three of the four *Hox* PG2 genes in the teleosts suggests the constituent genes belonging to this paralog group are prone to independent gene loss events. Numerous studies have shown that genetic redundancy leads to the rapid accumulation of deleterious alleles and ultimately to inactivation (Aparicio et al., 1997; Amores et al., 1998; Amores et al., 2004). If functional redundancy disposes *Hox* PG2 genes to inactivation, then we hypothesize that the activities of *Hox* PG2 genes from teleosts that have retained three functional genes should be distinct from one another, under stronger selection, and therefore less subject to mutational inactivation. Preliminary examination of tilapia *Hox* PG2 gene function suggest that the three tilapia genes, *hoxa2a*, *a2b* and *b2a* each have distinct but overlapping functions (Le Pabic et al., 2008), which we view as indicative that divergence of function favors gene retention within this paralog group. It is interesting in this context that the tetrapod *Hox* PG2 genes, *hoxa2* and *b2*, also appear to be functionally different from one another, such that *hoxa2* alone serves as a homeotic selector gene whereas *hoxb2* appears to play a role in neuronal specification in the hindbrain (Gendron-Maguire et al., 1993; Rijli et al., 1993; Barrow and Capecchi, 1996; Gavalas et al., 1997, 2003; Davenne et al., 1999; Barrow et al., 2000; Grammatopoulos et al., 2000; Pasqualetti et al., 2000; Baltzinger et al., 2005). Comparisons of *Hox* PG2 gene function across the Osteichthyes, point to a distribution in which functional redundancy appears to have been

restricted to post-genome duplication teleost clades that have retained only two functional genes. However, this may not have been the case historically. Despite the differential partitioning of *Hox* PG2 gene functional redundancy within specific teleosts clades, *Hox A* cluster genes from throughout the Osteichthyes characteristically function as selector genes to specify the identity of PA2. This allows us to infer that *Hox A* cluster genes functioned as selector genes in the common osteichthyan ancestor and that they may have functioned in this capacity since the inception of the vertebrates. Unfortunately, we are unable to infer the ancestral condition of the *Hox B* cluster PG2 genes because the gene from zebrafish, and presumably medaka, has selector gene activity, whereas the co-orthologous genes from the mouse and frog lack this function. This divergence in *hoxb2* function between the two major osteichthyan stem lineages makes it equally likely that either the common ancestral *hoxb2* gene possessed an activity that was subsequently lost in tetrapods, or that it did not have selector gene activity in the common ancestor but gained it in the lineage leading to zebrafish and possibly medaka. A resolution to this problem will require *Hox* PG2 gene functional studies in basal gnathostomes and in pre-genome duplication actinopterygians. Establishment of the ancestral function of *Hox B* PG2 genes will be critical in addressing which of several possible genetic mechanisms, i.e. sub- or neofunctionalization, are operative in maintenance of post-duplication *Hox* PG2 genes. It will also provide the basis for understanding whether the divergent functions of the two tetrapod *Hox* PG2 genes represent a condition characteristic of the common ancestor of the Osteichthyes or whether these two genes were functionally redundant.

CHAPTER 3: EVOLUTIONARY DIVERGENCE OF GENE REGULATION BETWEEN *HOXA2A* AND *ψHOXA2B* IN THE JAPANESE MEDAKA

Introduction

Clustered *Hox* genes are a family of evolutionarily related developmental regulatory genes that function to pattern regional tissue identities along the anterior-posterior (A-P) axis of animal species (McGinnis and Krumlauf, 1992). *Hox* clusters comprise up to 14 genes and the genes are expressed along the A-P axis during embryonic development collinear with their physical location within a cluster (Holland and Garcia-Fernandez, 1996; Ferrier et al., 2000; Powers and Amemiya, 2004). Multiple genome level duplications have expanded the total number of *Hox* clusters from one in chordates to four in tetrapods and at least seven or eight in most teleost fishes (Stellwag, 1999; Amores et al., 2004; Moghadam et al., 2005; Hoegg et al., 2007; Mungpakdee et al., 2008a). Post-genome duplication independent gene loss has generated clustered paralog groups that differ in gene number depending on the historical timing of gene losses relative to genome duplications (Amores et al., 2004; Le Pabic et al., 2007; Davis et al., 2008; Davis and Stellwag, 2010).

In addition to variation in gene number within and among paralog groups, results from expression and functional genetic studies have shown that duplicate genes can exhibit either similar or heterogenous expression patterns. In the cases in which the expression patterns among paralog group members differ they often appear to be the result of sequence divergence within *cis*-regulatory elements (Amores et al., 2004; Scemama et al., 2006; Tümpel et al., 2006; Hurley et al., 2007; Le Pabic et al., 2007, 2009; Davis et al., 2008; Mungpakdee et al., 2008b; Davis and Stellwag, 2010). A particular case of heterogeneous expression patterns exhibited by duplicate genes is that of *Hoxa2* and *b2* of tetrapods. Mouse *Hoxa2* is expressed in rhombomeres (r) 2-8

of the developing hindbrain and in the cranial neural crest cells (CNCCs) that delaminate from r4 and r6/7 and populate the second pharyngeal arch (PA2) and the posterior arches. *Cis*-regulatory elements (CREs) that direct mouse *Hoxa2* expression in r2 are located in exon 2 of *Hoxa2*, the r4 CREs are located in the intron and exon 1 and the CREs that direct *Hoxa2* in r3, r5 and the CNCCs are located in the *Hoxa3-a2* intergenic region (Frasch et al., 1995; Nonchev et al., 1996a and b; Maconochie et al., 2001; Tümpel et al., 2002, 2006, 2007 and 2008; Lampe et al., 2008). By contrast, mouse *Hoxb2* is expressed in r3-8 of the hindbrain but not in the CNCCs. The CREs that direct mouse *Hoxb2* in r3, r4 and r5 are located in the mouse *Hoxb3-b2* intergenic region (Sham et al., 1993; Vesque et al., 1996; Ferretti et al., 2000). The variation in mouse *Hoxa2* and *b2* expression patterns are mirrored by their functional divergence. Knockout experiments have shown that *Hoxa2* controls the segmentation of the anterior hindbrain and the axonal guidance of the Vth and VIIth cranial motor nerve axons out of r2/r4 while *Hoxb2* controls the specification of the somatic motor component of the VIIth cranial nerve exiting r4 (Barrow and Capecchi, 1996; Gavalas et al., 1997 and 2003; Davenne et al., 1999; Barrow et al., 2000). Further, *Hoxa2* is involved in patterning the craniofacial elements arising from PA2 and the posterior arches in tetrapods (Gendron-Maguire et al., 1993; Rijli et al., 1993; Grammatopoulus et al., 2000; Pasqualetti et al., 2000; Baltzinger et al., 2005; Minoux et al., 2009).

Research in our laboratory has been directed toward understanding the effects of teleost *Hox* paralog group 2 (PG2) post-genome duplication divergence on the evolution of developmental gene function and regulation. Phylogenetic reconstructions that include a whole genome duplication event at the incipient stage of teleost evolution support a post-genome duplication ancestral *Hox* PG2 gene complement consisting of two *Hoxa2* genes, *hoxa2a* and

a2b (Fig. 3-1A). The absence of a *hoxa2a* gene in zebrafish, but not in any members of the superorder Acanthopterygii, suggests that the loss of *hoxa2a* was restricted to a clade including zebrafish but not the acanthopterygians (Fig. 3-1A). Recent results from our laboratory, based on cloning and expression analyses of several acanthopterygian teleosts, including striped bass (*Morone saxatilis*), Nile tilapia (*Oreochromis niloticus*) and Japanese medaka (*Oryzias latipes*), have shown that while striped bass and Nile tilapia have two functional *Hoxa2* genes, *hoxa2a* and *a2b*, medaka has just one, *hoxa2a* (Scemama et al., 2006; Le Pabic et al., 2007; Davis et al., 2008) (Fig. 3-1A). Further, while *hoxa2a* and *a2b* genes of most teleosts are expressed in a conserved manner in much of the hindbrain (r2-r5), PA2 and the posterior arches similar to orthologous genes in tetrapods (Prince and Lumsden, 1994; Nonchev et al., 1996a, 1996b; Pasqualetti et al., 2000; Hunter and Prince, 2002; Amores et al., 2004; Baltzinger et al., 2005; Scemama et al., 2006; Le Pabic et al., 2007), the medaka expressed pseudogene (*ψhoxa2b*) was expressed in noncanonical *Hox* PG2 domains, including the ventral-most aspect of the neural tube, the distal mesenchyme of the pectoral fin buds and the caudal-most region of the embryonic trunk (Fig. 3-1B) (Davis et al., 2008). We interpreted this loss of canonical expression to be a reflection of sequence changes resulting from relaxed selection within the medaka *ψhoxa2b cis*-regulatory elements that direct hindbrain expression among canonical *hoxa2b* genes following the mutational inactivation of the *hoxa2b* coding sequence (Davis et al., 2008).

Results from comparative and functional genomic analyses of the regulatory loci controlling expression of vertebrate *Hox* PG2 genes have identified a conserved intergenic

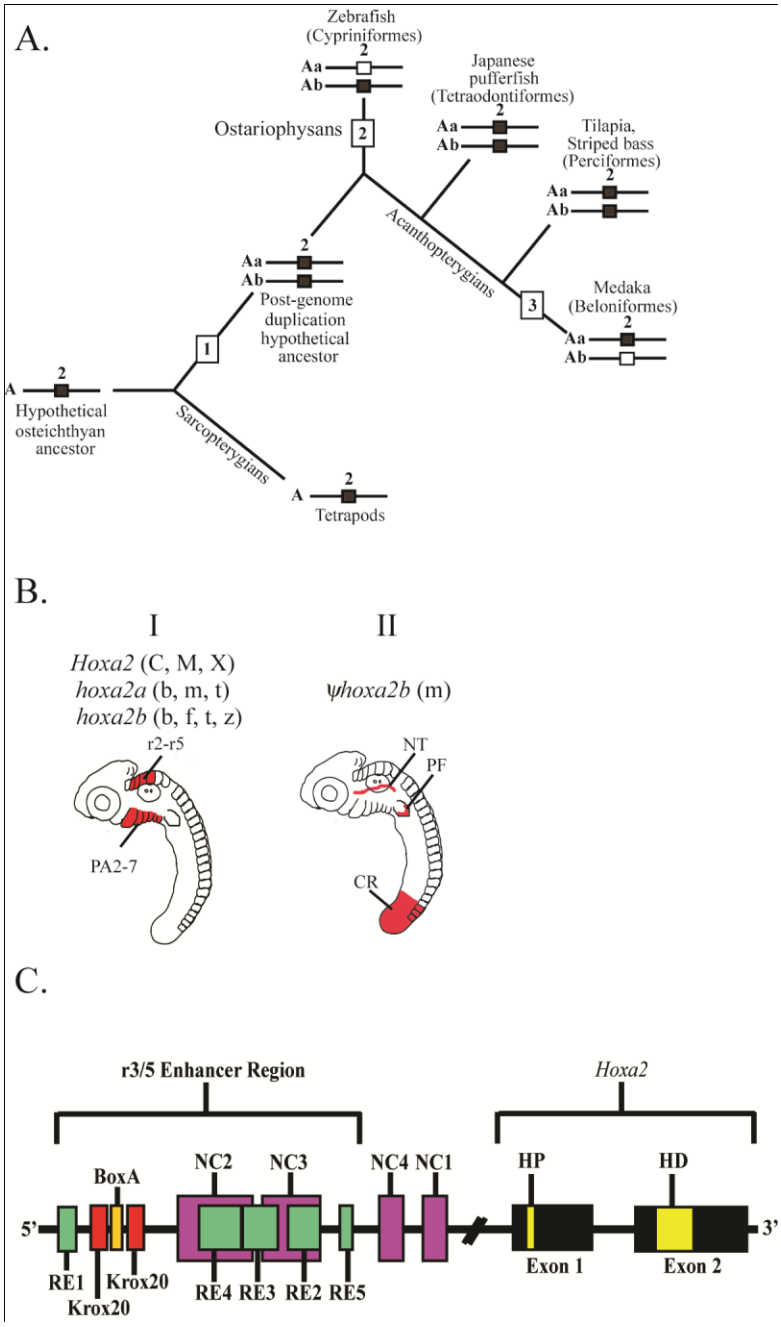


Fig. 3-1. **Evolution of *Hoxa2* gene complement and expression in the Osteichthyes.** (A) Phylogeny based on Steinke et al. (2006). (1) Genome duplication; (2) *hoxa2a* gene loss; (3) *hoxa2b* gene loss. (B) Diagrams of (I) conservative hindbrain (r2-r5) and pharyngeal arch (PA2-7) expression of tetrapod *Hoxa2* and teleost *hoxa2a* and *a2b* genes and (II) noncanonical *Hox* PG2 gene expression of medaka *ψhoxa2b*. (C) Genomic map of the *Hoxa2* r3/5ER characterized

in tetrapods and teleosts. The genomic map was not drawn to scale. RE, rhombomeric element.b, striped bass; C, chick; CR, caudal region; f; fugu; HD, homeodomain; HP, hexapeptide; m, medaka; M, mouse; NC, Neural Crest; NT, neural tube; PF, pectoral fin; r, rhombomere; t, tilapia; X, *Xenopus*; z, zebrafish.

region upstream of *Hoxa2* in vertebrates (Maconochie et al., 1999 and 2001; Tümpel et al., 2002, 2006 and 2007). This region has been termed the r3/5 enhancer region (r3/5ER) and, accordingly, it was shown using reporter gene expression constructs in the mouse and chicken to be responsible for directing *Hoxa2* gene expression in r3 and r5 of the hindbrain and in the CNCCs that migrate to and populate PA2 and the posterior pharyngeal arches (Fig. 3-1C) (Maconochie et al., 2001; Tümpel et al., 2002, 2006 and 2007). Many of the elements comprising the r3/5ER have been identified in the mouse developmental model system and include binding sites for Krox20, a transcription factor that has been shown to play a critical role in the specification of hindbrain segmentation (Nonchev et al., 1996a and b; Maconochie et al., 2001). Other elements, including a BoxA site and several other rhombomeric elements (RE1-5), have been shown to function in conjunction with Krox20 in potentiating *Hoxa2* expression in r3 and r5 (Frasch et al., 1995; Nonchev et al., 1996a and b; Maconochie et al., 2001; Tümpel et al., 2002; Tümpel et al., 2006). The r3/5ER has been shown to function similarly for directing r3 and r5 expression in the hindbrain for *Hoxa2* of mouse and chicken and *hoxa2b* of fugu and *ψhoxa2b* of medaka (Tümpel et al., 2002 and 2006). However, it must be noted that, unlike the r3/5ERs of mouse and chicken *Hoxa2,* which were tested in their respective host species (homologous expression systems), functional analyses of the r3/5ERs for the fugu and medaka *Hoxa2* duplicates were tested in chicken embryos (heterologous expression system) (Tümpel et al., 2006). Several sequence elements located within and downstream of the mouse r3/5ER have been shown to direct mouse *Hoxa2* expression in the CNCCs of the hyoid and post-otic CNCC streams originating from the hindbrain at the level of r4 and r6/r7 and populating PA2 and the posterior pharyngeal arches, respectively (Maconochie et al., 1999). These elements include neural crest elements 1-4 (NC1-4). NC2 and NC3 are located within the r3/5ER of mouse

Hoxa2 and NC1 and NC4 are located downstream of the r3/5ER (Maconochie et al., 1999) (Fig. 3-1C). All four elements were shown to be necessary for directing *Hoxa2* expression in the CNCCs migrating to and populating PA2 and the posterior pharyngeal arches (Maconochie et al., 1999). Interestingly, the r3/5ERs of fugu *hoxa2a* and *a2b* and medaka *hoxa2a* and *ψhoxa2b* did not drive reporter gene expression in the CNCCs when they were tested in chicken embryos (Tümpel et al., 2006), which suggested that these sequences did not possess elements that could direct reporter gene expression in the CNCCs.

Heterologous reporter gene assays of teleost genomic sequences performed in chicken embryos showed that the r3/5ERs of fugu *hoxa2b* and medaka *ψhoxa2b* were able to direct reporter gene expression in r3 and r5 of the chick hindbrain but the r3/5ERs of fugu and medaka *hoxa2a* were not (Tümpel et al., 2006). Although the heterologous reporter gene expression results documented by Tümpel et al. (2006) were in general agreement with the hindbrain expression patterns for fugu *hoxa2a* and *a2b*, which were shown to be expressed in r1/r2 and r2-r5, respectively, they were not in agreement with native medaka *hoxa2a* and *ψhoxa2b* expression patterns. Medaka *hoxa2a,* like *Hoxa2* of tetrapods and *hoxa2a* and *a2b* genes of most teleosts, was shown to be expressed in the hindbrain (r2-r7) and pharyngeal arches, whereas medaka *ψhoxa2b* was expressed predominately in noncanonical *Hox* PG2 domains (Davis et al., 2008).

Despite the sequence conservation of *cis*-regulatory elements between the r3/5ERs of teleost *hoxa2a* and *a2b* genes and tetrapod *Hoxa2* genes shown by Tümpel et al. (2006), our reporter gene expression analyses showed that the medaka *hoxa2a* r3/5ER directed reporter gene expression in r4 of the medaka hindbrain, the migratory CNCCs of the hyoid and post-otic CNCC streams and the post-migratory CNCCs in PA2 and the posterior pharyngeal arches. These results were different from reporter gene assays of the r3/5ER of mouse *Hoxa2*, which

directed reporter gene expression in r3 and r5, and the medaka *hoxa2a* r3/5ER when it was tested in chicken embryos, which did not direct any reporter gene expression in the hindbrain or pharyngeal arches. Further, reporter gene mediated *cis*-element mapping studies conducted in medaka using the medaka r3/5ER showed that only genomic sequences corresponding to the conserved sequence elements RE3 and RE2 were required for generating reporter gene expression in r4 and the CNCCs. Comparative genomic sequence analyses of the orthologous region from several *Hoxa2* genes of evolutionarily divergent vertebrates, including shark, tetrapods and teleosts, revealed the presence of conserved Hox/Pbx and Prep/Meis binding sites. In each case examined, including mouse and chicken using homologous reporter gene assays, Hox/Pbx and Prep/Meis binding sites play an integral role in directing *Hox* gene expression in r4 (Ferretti et al., 2000; Tümpel et al., 2007; Lampe et al., 2008). By comparison, the *ψhoxa2b* r3/5ER of medaka was shown to direct reporter gene expression in r3-7 of the hindbrain, the hyoid and post-otic migratory CNCCs and the post-migratory CNCCs in PA2 and the posterior arches. These results were unexpected for several reasons; when tested in the transgenic chicken model the *ψhoxa2b* r3/5ER of medaka directed in r3 and r5 but not in r4, r6, r7 or in the CNCCs (Tümpel et al., 2006); further, the natural expression of *ψhoxa2b* in medaka occurs in noncanonical *Hox* PG2 domains, including the caudal-most region of the embryonic trunk, the ventral-most aspect of the neural tube and the distal mesenchyme of the pectoral fin buds (Davis et al., 2008). Further, reporter gene assays of the medaka *ψhoxa2b* r3/5ER in their homologous host showed that the sequence orthologous to the medaka *hoxa2a* r3/5ER sequence, which directs reporter gene expression in r4, was able to drive expression in r3-r7 and the migratory and post-migratory CNCCs. Overall, these results show that the functional nature of the r3/5ER has diverged among osteichthyans. They also point to a deeply rooted r4/CNCC-specifying

element that is present in the regulatory sequences common to vertebrate *Hox* A clusters but has diverged functionally over the course of vertebrate evolution.

Methods and Materials

Tol2 plasmid construction

Transient and stable-line transgenic analyses employed the pT2AL200L200G plasmid vector for transmitting transposon insertions into medaka embryos (generous gift from Kawakami) (Urasaki et al., 2006). This plasmid vector contains the *Xenopus laevis* eFI-αS promoter upstream of *enhanced green fluorescent protein* (*eGFP*) and was used for positive controls to determine if the *Tol2* transposon system functions similarly in medaka relative to other osteichthyans. Positive controls were performed by co-microinjection of constructs containing the *Xenopus laevis* eFI-αS promoter and transposon mRNA. Negative controls were performed by co-microinjection of constructs containing the *Xenopus laevis* eFI-αS promoter without transposon mRNA. Medaka contains roughly 20-30 copies of the *Tol2* element in its genome, and these controls were performed to determine whether constructs could be integrated into the medaka genome independently of exogenously translated *Tol2* mRNA (Kawakami, 2007). The microinjection procedure used for medaka zygotes is described below.

In order to analyze the function of the CREs of the medaka *hoxa2a* and *ψhoxa2b* r3/5ERs, the eFI-αS promoter of the pT2AL200L200G was truncated to include only the region encompassing the TATA box so as to diminish the amount of transcription occurring from the vector and maximize transcription resulting from cloned sequences. A multiple cloning site was inserted upstream of the TATA box to allow for insertion of genomic DNA in the plasmid vector system (pTolTATA-βMCS). Design of the pTolTATA-βMCS vector system for transgenic analyses in developmental fish models is described by Reubens (M.S. Thesis, 2009, East

Carolina University). The pTolTATA-βMCS vector was also co-microinjected with and without transposon mRNA to test the efficacy of this vector system in the medaka model.

Medaka genomic DNA extraction

Adult medaka were anesthetized with MS-222 (0.04 w/v) prior to genomic DNA extraction. Tissues from the trunk of the fish behind the anus were homogenized in 2 mL of DNA extraction buffer (0.5% SDS, 50 mM Tris, 100 mM NaCl, 20 mM disodium EDTA, pH 8.0) and the resulting homogenate was poured into sterile 15 mL Falcon tubes (Becton Dickinson Labware). The tissue homogenizer was washed with an additional 2.0 mL of DNA extraction solution, which was combined with the previous homogenate in the sterile 15 ml polypropylene tube. Five microliters of RNase A (20 mg/mL) (Sigma) was added to the cell homogenate and the tube was incubated for 30 min at 37 °C. The homogenate was then treated with 100 μL of protease K (10 mg/mL) (Invitrogen) and incubated at 55 °C for 6 hours. The RNase A and protease K-treated homogenate was transferred to 1.5 mL centrifuge tubes (Fisher Scientific) and centrifuged for 30 seconds at 14,000 rpm using an Eppendorf 5514 C centrifuge. The resulting supernatant was then divided into two 500 μL aliquots in sterile 1.5 mL centrifuge tubes (Fisher Scientific), and DNA was purified by two Phenol Chloroform Isoamyl alcohol (PCI) extractions as follows: 500 μL of PCI (Ambion) was added to each 1.5 mL tube and mixed until a homogeneous emulsion formed. Microcentrifuge tubes containing the emulsion were centrifuged for 30 seconds at 14,000 rpm, and the aqueous layer was removed from the organic layer and then placed in a sterile 1.5 mL microcentrifuge tube. After a second PCI extraction, the aqueous phase was treated with 5 M NaCl, which was added to the aqueous extract, such that the final concentration of NaCl in the tube would be 0.3 M after the addition of 2 volumes of 100% ethanol (EtOH, Pharmco-Aaper). After addition of the ethanol, the tubes were mixed

gently to allow the DNA to precipitate, after which the precipitated DNA was collected by centrifugation at 10,000 x g for 1 minute. The supernatant was removed by pipetting and the sedimented DNA was washed twice with 70% EtOH and incubated at 37 °C until dry. Once dry, the DNA was suspended in 100 μL of Ambion Nuclease Free Water. The quality of the genomic DNA was assayed by agarose gel electrophoresis on a 0.5 % gel. The migration of the genomic DNA was compared to that of the DNA fragments in the High Molecular weight standard (GIBCO BRL).

Amplification of medaka *hoxa2a* and *ψhoxa2b* regulatory regions and PCR-mediated deletion mutagenesis

Amplification of genomic DNA corresponding to *cis*-regulatory elements in the upstream DNA sequences, exons 1 and 2 and the intronic DNA of medaka *hoxa2a* and *ψhoxa2b* was performed using long range PCR. Primers and their hybridization coordinates to medaka genomic DNA with respect to ATG start site of medaka *hoxa2a* and *ψhoxa2b* are listed in Table 3-1. The PCR products generated from long-range PCR were cloned into pCR II vectors (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. Confirmation and orientation of PCR products corresponding to inserts from plasmid genomic DNA clones were determined by restriction endonuclease digestion. These clones were then used to amplify the r3/5 enhancer regions of medaka *hoxa2a* and *ψhoxa2b*, which were tested for enhancer activity in the pTolTATA-βMCS plasmid vector.

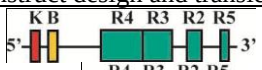
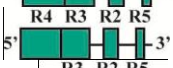
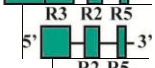
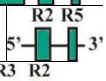
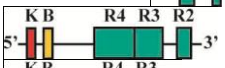
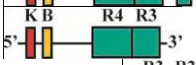
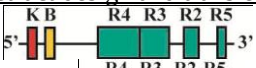
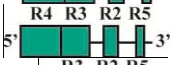
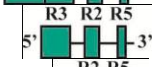
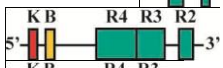
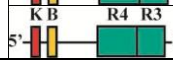
DNA sequences that contained the r3/5ERs of medaka *hoxa2a* and *ψhoxa2b* were amplified using PCR primers listed in Table 3-1. These primers were used to generate a series of

Table 3-1. Primers used for the amplification of the medaka *hoxa2a* and *ψhoxa2b* r3/5ERs and the exclusion of sequence elements for the functional testing of these regulatory regions.

Primer	Sequence 5' to 3'	5' start site (with respect to ATG start site of downstream gene)
	Medaka *hoxa2a* Genomic Primers	
A2a For	TTATTCCCACAACCCTTTCATTTCG	-2691
A2a Rev	CACACTCAGCCACAATCTCTTCTTC	1846
	Medaka *ψhoxa2b* Genomic Primers	
A2b For	ACACAGCAGGGGTCAACAATAGGTC	-3093
A2b Rev	ATAGGCAGAGCACGAAAACAAAATG	3193
	Medaka *hoxa2a* r3/5 Forward Primers	
AF1	GATCGATATCGAACAGGCTGAAATCCACTGAATGC	-1778
AF2	GATCGATATCGCTTCTAATCTGAGAAGCCAGTGTTTC	-1468
AF3	GATCGATATCATGTGTTGCGAGGGCACCGAGCTGTC	-1392
AF4	GATCGATATCGAGTAAGATTGATCGCGCACAGGCTTC	-1354
	Medaka *hoxa2a* r3/5 Reverse Primers	
AR1	GATCGAATTCGTTTGCTGTGGAACAGAGGAAAGAAG	-1247
AR2	GATCGAATTCTTATATACCAAACAAAGAGTCCTGG	-1303
AR3	GATCGAATTCTTACTCGCCAAAAGGTCTGACAGCTC	-1348
	Medaka *ψhoxa2b* r3/5 Forward Primers	
BF1	GATCGATATCATGTGCCAACACCCACTCACCCCAG	-1068
BF2	GATCGATATCCTTCGCTCCGCACCGAGGGCATCCTC	-868
BF3	GATCGATATCATGTTCTCTAAGGGCAAAGAGCTGTC	-803
BF4	GATCGATATCTGGAAAGATTGATCACACAGAATACC	-765
	Medaka *ψhoxa2b* r3/5 Reverse Primers	
BR1	GATCGAATTCAAAAAGCTGCAGGAAAAGGAGGGGATC	-671
BR2	GATCGAATTCCCGGGCTCTGAACAAAAGATTCCTG	-715
BR3	GATCGAATTCTTTCCAGCCAAGAGCTCTGACAGCTC	-759

nested deletion constructs encompassing the region extending from genomic sequence positions -1778 to -1247 upstream of medaka *hoxa2a* and -1068 to -671 upstream of medaka *ψhoxa2b* to include specific *cis*-regulatory sequences of the medaka *hoxa2a* and *ψhoxa2b* r3/5ERs that were functionally tested by Tümpel et al. (2006) in chicken embryos. Specific primer pairs and their amplified genomic DNA products are listed in Table 3-2. All 5'-located primers started with the sequence 5'-GATCGATATC-3' and 3'-located primers started with the sequence 5'-GATCGAATTC-3 to ensure that the PCR products could be digested with EcoRI and EcoRV restriction digestion enzymes and oriented 5' to 3' with respect to the TATA box and GFP sequences located downstream in the pTolTATA-βMCS vector. PCR products were purified using the QIAquick PCR purification kit (Qiagen, Valencia, CA) according to the manufacturer's instructions, restriction endonuclease digested with EcoRI and EcoRV and then ligated into the pTolTATA-βMCS plasmid vector using T4-DNA ligase (New England Biolabs, Ipswich, MA). The pTolTATA-βMCS plasmid vector was pre-digested with EcoRI and EcoRV, purified using the QIAquick PCR sequence kit (Qiagen, Valencia, CA) and treated with Shrimp alkaline phosphatase (New England Biolabs, Ipswich, MA) to remove phosphate groups at the digested sites of the vector to minimize vector religation. Plasmid vectors containing PCR-amplified genomic DNAs were transformed into *E. coli* JM109 cells. *E. coli* cells were screened for recombinants by digestion of plasmid isolated from transformants with EcoRI and EcoRV after purification of plasmids using the GenElute HP Plasmid Miniprep kit (Sigma-Aldrich) according to the manufacturer's instructions. Further confirmation of cloned PCR products in the pTolTATA-βMCS plasmid vector was performed using DNA sequencing of the inserts by

Table 3-2. Primer pairs used for the amplification of sequences of the medaka *hoxa2a* and *ψhoxa2b* r3/5ERs used in functional genomic analyses. Schematics of constructs and frequencies of transient transgenic embryos showing reporter gene expression in the hindbrain and CNCCs are shown.

Construct	Primer Pairs	Amplicon Length	Construct Schematic	Hindbrain Expression	CNCC Expression
			Medaka *hoxa2a* r3/5ER construct design and transient transgenic analysis		
1	AF1/AR1	531 bp		42/48 (87.5%)	42/48 (87.5%)
2	AF2/AR1	221 bp		64/84 (76%)	56/84 (67%)
3	AF3/AR1	145 bp		39/49 (80%)	41/49 (84%)
4	AF4/AR1	107 bp		7/47 (15%)*	7/47 (15%)*
5	AF1/AR2	475 bp		42/50 (84%)	42/50 (84%)
6	AF1/AR3	430 bp		0/52 (0%)	0/52 (0%)
7	AF3/AR2	89 bp		52/62 (84%)	52/62 (84%)
			Medaka *ψhoxa2b* r3/5 construct design and transient transgenic analysis		
8	BF1/BR1	397 bp		46/51 (90%)	46/51 (90%)
9	BF2/BR1	197 bp		47/52 (90%)	47/52 (90%)
10	BF3/BR1	132 bp		33/38 (87%)	33/38 (87%)
11	BF4/BR1	94 bp		27/52 (52%)*	27/52 (52%)*
12	BF1/BR2	353 bp		33/41 (80%)	33/41 (80%)
13	BF1/BR3	309 bp		9/64 (14%)*	15/64 (23%)*

dideoxyterminator sequencing chemistry (Big Dye v. 3.0, Applied Biosystems, Foster City, CA) on a 3130xl Genetic Analyzer (Applied biosystems, Foster City, CA).

Microinjection of medaka embryos

Cultivation of medaka was performed as described in Davis et al. (2008). Zygotes collected from females (stage 1, Iwamatsu (2004)) were transferred to ice-cold 1X medaka embryo rearing medium (ERM) (17 mM NaCl, 0.4 mM KCl, 0.66 mM $MgSO_4.7H_2O$, 0.27 mM $CaCl_2.2H_2O$) (Oxendine et al., 2006) between 30 min to 2 hr to arrest development. Zygotes were then physically transferred to specially fabricated agarose 'corrals' that were pre-chilled at 4 °C and that hold the zygotes in an orientation appropriate for microinjection. The agarose corrals were 1 mm wide x 1 mm deep x 4 mm long and were made using 60 ml of 1.5% agarose. To circumvent high internal egg pressures when microinjecting, the agarose corrals were overlaid with 10 ml of 15% Ficoll 400 pre-chilled at 4 °C. Zygotes were placed in their corrals and were allowed to equilibrate for at least 40 min in 15% Ficoll 400 on ice prior to microinjection. Zygotes were microinjected using needles derived from filamented borosilicate glass capillaries that had an outer diameter of 1 mm and an inner diameter of 0.58 mm (World Precision Instruments, Sarasota, FL). Glass needles were pulled on a Sutter P-97 needle puller (Novato, CA) housing a 1.5 mm trough filament using the following parameters: heat of 250, pull of 200, velocity of 100, time of 200 and pressure of 400. To avoid clogging from chorions, the tips of the needles were beveled at a 45° angle using a Narishige Micro-grinder (Tokyo, Japan). After beveling the tips, glass needles were stored by embedding their mid-sections in rounded stripes of Play-Doh (Hasbro, Pawtucket, RI) housed in 100 mm petri dishes with dampened Kimwipes (Kimberly Clark®) at 4 °C. The Play-Doh prevented glass needles from movement and breakage in the Petri plates. The dampened Kimwipe maintained a moist

environment inside the Petri dishes and kept the stored needles from clogging at the beveled ends. Microinjection of solutions into medaka zygotes with beveled needles was performed using a PV 820 pneumatic picopump (World Precision Instruments, Sarasota, FL). Twenty pounds per square inch (psi) of eject pressure and 3 psi of hold pressure were used for the medaka zygote microinjection process. Zygotes were microinjected with the following solutions for analyzing transient transgenic embryos: 25 ng/μl plasmid DNA, 25 ng/μl *Tol2* transposase RNA, 1X Yamamoto Buffer (128 mM NaCl, 2.7 mM KCl, 1.8 mM $CaCl_2$, 0.24 mM $NaHCO_3$, pH 7.3). After microinjection, embryos were carefully removed from their individual agarose corrals and transferred in approximately 1 ml of 15% Ficoll to a Petri dish containing 20 mL 1X ERM. After transfer they were allowed to equilibrate to less than 1% Ficoll for roughly 2 hours at room temperature (RT) without shaking. The medium housing the embryos was then replaced with fresh 1X ERM and the embryos were incubated at 28.5 °C to continue development.

Generation and visualization of transient and stable-line trasnsgenic medaka embryos

Microinjected embryos were raised in 1X ERM at 28.5 °C. After 24 hr, embryos were visualized under a Leica dissecting microscope. Dead eggs and embryos that showed extremely defective morphologies (i.e.: gastrulation defects) were discarded. All other embryos were transferred in 1X ERM containing 0.1 mM phenylthiourea (Sigma-Aldrich, Sweden AB) to reduce pigmentation (Karlsson et al., 2001). Embryos were visualized for *eGFP* expression using a Zeiss inverted compound microscope (Thornwood, NY) at selected times during development, focusing primarily on developmental stage 29/30 (74-82 hpf; Iwamatsu, 2004). At stage 29/30, the hindbrain and pharyngeal arches are easily distinguished morphologically in medaka embryos when they are in the chorion. Since embryos at stage 29/30 have guanophores that auto-fluoresce under GFP filters, illumination of specimens using both GFP and rhodamine

filters was performed to differentiate positive *eGFP* signal from auto-fluorescing pigmentation. Embryos were visualized using Axiovision AC 4.4 software. Depending on the strength of the *eGFP* signal, camera exposure times ranged from 1 to 2 sec for transient transgenic embryos. Brightfield images were also taken of medaka embryos to determine the relative origin of the *eGFP* signal within the hindbrain and pharyngeal arches in transient transgenic embryos. The developing otic vesicle was used to determine whether the *eGFP* signal was occurring anteriorly or posteriorly in the hindbrain. Rhombomeres 3 and 4 develop above the anterior region of the otic vesicle while r5, r6 and r7 develop above the mid- to posterior region of the otic vesicle.

Transient transgenic medaka embryos that were observed to be positive for *eGFP* expression in the hindbrain and pharyngeal arches showed mosaic expression in comparisons among embryos. For this reason, transient transgenic embryos that showed strong *eGFP* signal in the hindbrain and/or pharyngeal arches were raised to adulthood and mated with wild-type fish. We observed that more robust *eGFP* expression patterns were detected in the hindbrain and pharyngeal arches of stable-line transgenic medaka embryos compared to transient transgenics. Transient transgenic embryos were raised until hatching in 1X ERM at 28.5 °C. Hatched embryos were raised in 1X ERM in breeding nets in 5 gallon tanks until they were large and strong enough to swim against filter-generated currents. Adult female fish that developed from embryos that were originally scored as positive for *eGFP* expression in the hindbrain and pharyngeal arches were mated with wild-type males. F2 medaka embryos resulting from such genetic crosses were assayed for *eGFP* expression according to the procedure for transient transgenic embryos.

Whole-mount *in situ* hybridization

To corroborate the results of *eGFP* expression using a system independent of fluorescence, the expression of *eGFP* in embryos was visualized using *in situ* hybridization analysis with an anti-*eGFP* riboprobe. This method allowed for the visualization of *eGFP* expression in the absence of fluorescence originating from auto-fluorescing pigment cells, which was often the case when embryos were visualized under a GFP filter. We used whole-mount *in situ* hybridization to visualize *eGFP* transcripts at developomental stages 22 (nine somites), 29/30 and 34 (121 hpf). We assayed embryos at stage 22 because we wanted to determine if the CREs of the medaka *hoxa2a* and *ψhoxa2b* r3/5ERs were directing reporter gene expression in the migratory CNCCs of the hyoid and post-otic streams. We assayed embryos at stage 29/30 to visualize the exact rhombomeres and pharyngeal arches that were expressing *eGFP*. We assayed embryos at stage 34 to determine if the CREs of the medaka *hoxa2a* and *ψhoxa2b* r3/5ERs were directing expression in post-migratory CNCCs at the chondrogenic stages of pharyngeal arch development.

Embryos from stable-line transgenic medaka fish were collected, raised, anesthetized, fixed in 4% paraformaldehyde (PFA) and dehydrated as described in Davis et al. (2008). Medaka embryos were developmentally staged according to Iwamatsu (2004). Whole-mount *in situ* hybridization was performed according to Davis et al. (2008). All experiments used digoxigenin (DIG)-labeled sense and antisense riboprobes were produced and purified according to Scemama et al. (2006). Sense riboprobes were used in control experiments to assess nonspecific binding. Development of DIG-labeled probe signal, examination of embryos and digital photography of embryos was performed as described in Scemama et al. (2006). Morphological landmarks, including the midbrain/hindbrain boundary, rhombomeres (r), otic vesicles (OV), pectoral fins (PF), pharyngeal arches (PA) and somites (s) within developing

embryos were used to define the location of *eGFP* expression. *eGFP* signal was also determined using double whole-mount *in situ* hybridization assays with DIG-labeled antisense-*eGFP* riboprobes and DIG-labeled antisense-*hoxd3a* riboprobes or fluorescein-labeled *hoxb1a* riboprobes. Medaka *hoxd3a* is expressed in r6-8 of the hindbrain and *hoxb1a* is expressed in r4 (Hurley et al., 2007; Davis and Stellwag, 2010). Production of fluorescein riboprobes and double whole-mount *in situ* hybridization using DIG and fluorescein-labeled riboprobes was performed as documented in Scemama et al. (2006).

Comparative genomic sequence analysis

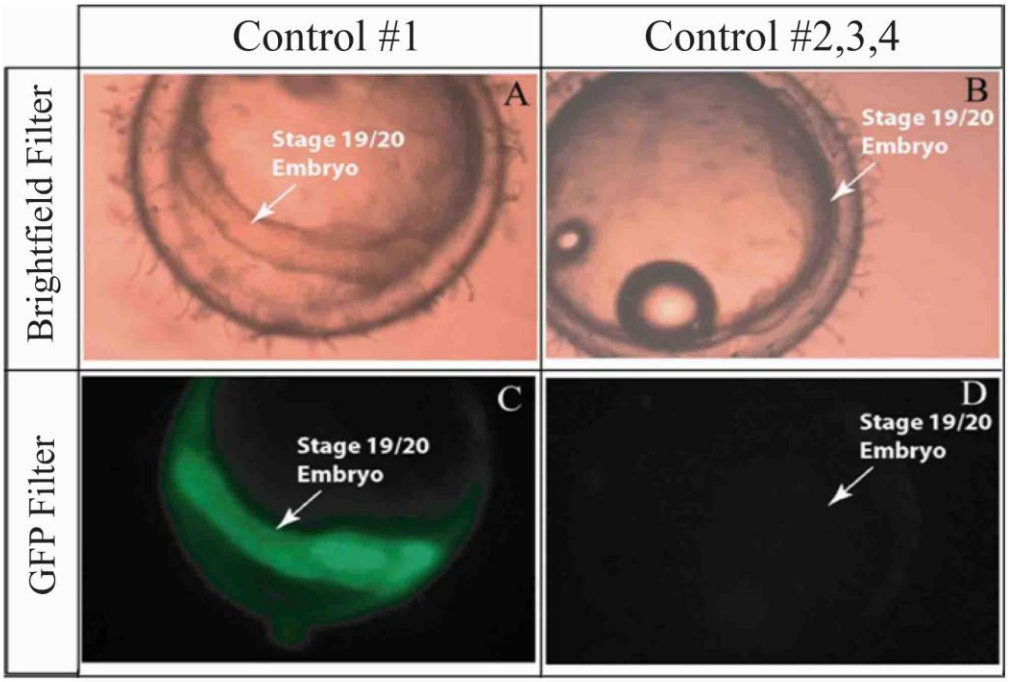
Genomic DNA sequences from the r3/5ERs of medaka *hoxa2a* and *ψhoxa2b* that were shown to be required for directing reporter gene expression in the rhombomeric and neural crest embryonic domains using functional genomic analyses outlined above were examined for putative transcription factor binding sites using the Transcription Element Search Software (TESS, Schug and Overton, 1997 at http://cbil.upenn.edu/tess) and JASPAR (Sandelin et al., 2004 at http://jaspar.genereg.net). We also performed a comparative genomic sequence analysis of the r3/5ERs between medaka *hoxa2a* and *ψhoxa2b* and the r3/5ERs of other vertebrate *Hoxa2* genes using the software Dialign-TX (http://dialign-tx.gobics.de/) (Subramanian et al., 2008) in order to determine if the putative transcription factor binding sites were conserved in other vertebrates. Amplification of the tilapia and striped bass *hox9b-a2b* genomic sequences were performed using long range PCR (striped bass, 5'-CCCACCAGAAAAAAGCGTTGTCC-3' and 5'-TGTAGTGTGAAGCAGAGGAAGG-3'; tilapia, 5'-CCTGATAACCCGTCATCCAACTG-3' and 5'-GGGATCTGGTGGCTGTTTATTGC'3'). Generation of the striped bass *hoxa3a-a2a* genomic sequence was performed by subcloning and primer walking. The *Hoxa2* r3/5ERs from genomic sequences of medaka (accession numbers AB232918 and AB232919) (Kurosawa et al.,

2006), fugu (accession numbers DQ481663 and DQ481664) (Lee et al., 2006), zebrafish (accession number AL645795, direct submission), Nile tilapia (AF533976 and this study) (Malaga-Trillo and Meyer, 2001), striped bass (this study), bichir (Chiu et al., 2004), mouse (accession number: NC000072) (Church et al., 2009), human (accession number: NG012078, direct submission), chicken (accession number: AC163712, direct submission), *Latimeria* (accession number: FJ497005) (Amemiya et al., 2010), horn shark (accession number: AF224262) (Kim et al., 2000) and dogfish (accession number: FQ032658) (Oulion et al., 2010) were compared using Dialign-TX (http://dialign-tx.gobics.de/) (Subramanian et al., 2008). All default and recommended parameters were used in the Dialign-TX web interface for conducting DNA sequence comparisons.

Results

Validation of the *Tol2* transposon system for medaka embryos

In order to determine if the *Tol2* transposon system functions similarly in medaka compared to zebrafish, we performed experiments using the pT2AL200L200G vector. This plasmid vector contains the *Xenopus laevis* eFI-αS promoter upstream of *eGFP*. We expected that positive control embryos, which were co-microinjected with pT2AL200L200G vector and *Tol2* transposon mRNA, would show green fluorescence during development. However, we were unsure if negative control embryos, which were microinjected solely with pT2AL200L200G vector, would not exhibit green fluorescence, since the *Tol2* element is present within the medaka genome (Kawakami, 2007). All positive control embryos showed strong reporter gene expression throughout the body of the medaka embryos (68/68; 100%) (Fig. 3-2A and C) while negative control embryos failed to show any reporter gene expression (0/52; 0%)



Control Experiment #	Constructs	% Positives
1	pT2AL200L200G + Transposon mRNA	68/68 (100%)
2	pT2AL200L200G - Transposon mRNA	0/52 (0%)
3	pTolTATA-βMCS + Transposon mRNA	0/64 (0%)
4	pTolTATA-βMCS - Transposon mRNA	0/49 (0%)

Fig. 3-2. Control experiments used for the validation of the *Tol2* transposon system for medaka embryos.

(Fig. 3-2B and D), which suggested that *eGFP* expression was not generated from microinjected vectors in the absence of exogenous *Tol2* transposon mRNA. Further, medaka zygotes microinjected with the pTolTATA-βMCS vector alone (0/49, 0%) or co-microinjected with vector and transposon mRNA (0/64, 0%) lacked detectable fluorescence (0/64, 0%) (Fig. 3-2B and D). These results showed that the *Tol2* transposon system and the pTolTATA-βMCS vector generated by Reubens (M.S. Thesis, 2009, East Carolina University) can be used in medaka embryos for studying *cis*-regulatory element control of gene expression.

Functional Genomic Analysis of the Medaka *hoxa2a* r3/5 enhancer region

Medaka *hoxa2a* is expressed in a conserved manner in the hindbrain and pharyngeal arches relative to *hoxa2a* and *a2b* genes from teleosts, including zebrafish, tilapia, striped bass and fugu and *Hoxa2* genes of tetrapods, including the mouse, chicken and frog (Prince et al., 1994; Nonchev et al., 1996; Prince et al., 1998; Pasqualetti et al., 2000; Hunter and Prince, 2002; Amores et al., 2004; Baltzinger et al., 2005; Scemama et al., 2006; Le Pabic et al. 2007; Davis et al., 2008). Based on this extensively conserved pattern of expression, we hypothesized that the r3/5 enhancer region upstream of medaka *hoxa2a* would function similarly to the r3/5 enhancer region in tetrapods, wherein it drives reporter gene expression in r3 and r5 of the hindbrain and in the CNCCs entering PA2 and the posterior arches.

Contrary to our hypothesis, transient and stable-line transgenic medaka embryos showed that the 531 bp construct containing the entire medaka *hoxa2a* r3/5ER and spanning from genomic bp positions -1778 to -1247 (with respect to the ATG start site of medaka *hoxa2a*) (Construct #1, Table 3-2) directed reporter gene expression in r4 of the hindbrain and in PA2 and the posterior pharyngeal arches (Fig. 3-3A-G). We observed *eGFP* expression in the hindbrain (87.5%) and pharyngeal arches (87.5%) in a high percentage of transgenic embryos (Table 3-2).

Whole-mount *in situ* hybridization of stable-line transgenic embryos showed that the *eGFP* expression in the hindbrain was restricted to r4 (3-3G), which was unexpected given that the orthologous r3/5ER of mouse directed reporter gene expression in r3 and r5 (Frasch et al., 1995; Nonchev et al., 1996a and b; Maconochie et al., 1999 and 2001; Tümpel et al., 2002 and 2006). This restricted and unusual pattern of expression prompted questions concerning the validity of rhombomere assignments. To authenticate these rhombomere assignments, additional whole-mount *in situ* hybridization experiments were conducted using medaka *hoxb1a*, which is expressed in r4 (Hurley et al., 2007) and *hoxd3a*, which is expressed in r6-r8 of the hindbrain but not in the pharyngeal arches (Davis and Stellwag., 2010), as molecular markers along with anti-*eGFP* riboprobes. Medaka *hoxb1a* antisense riboprobes were labeled with fluorescein. Both fluorescein-labeled *hoxb1a* and DIG-labeled *eGFP* riboprobes were observed to hybridize to their mRNA targets in r4 (Fig. 3-3H). When DIG-labeled *eGFP* and *hoxd3a* antisense riboprobes were used in tandem for *in situ* hybridization experiments, a one-rhombomere gap without any DIG-stained cells between the *eGFP* and *hoxd3a* expressing rhombomeres was observed (Fig. 3-3I). These results are consistent with expression of *eGFP* in r4 with a gap in expression corresponding to r5 and *hoxd3a* expression in r6-r8 (Fig. 3-3I). Further, *eGFP* expression was observed in the migratory CNCCs of the hyoid and post-otic streams at developmental stage 22, the post-migratory CNCCs in PA2 and the posterior pharyngeal arches at stage 29/30 and the chondrogenic CNCCs in PA2 and the posterior arches at stage 34 (Fig. 3-4 A-C). However, we observed much higher levels of *eGFP* expression in post-migratory CNCCs in the posterior pharyngeal arches than in PA2. These results are different from those observed

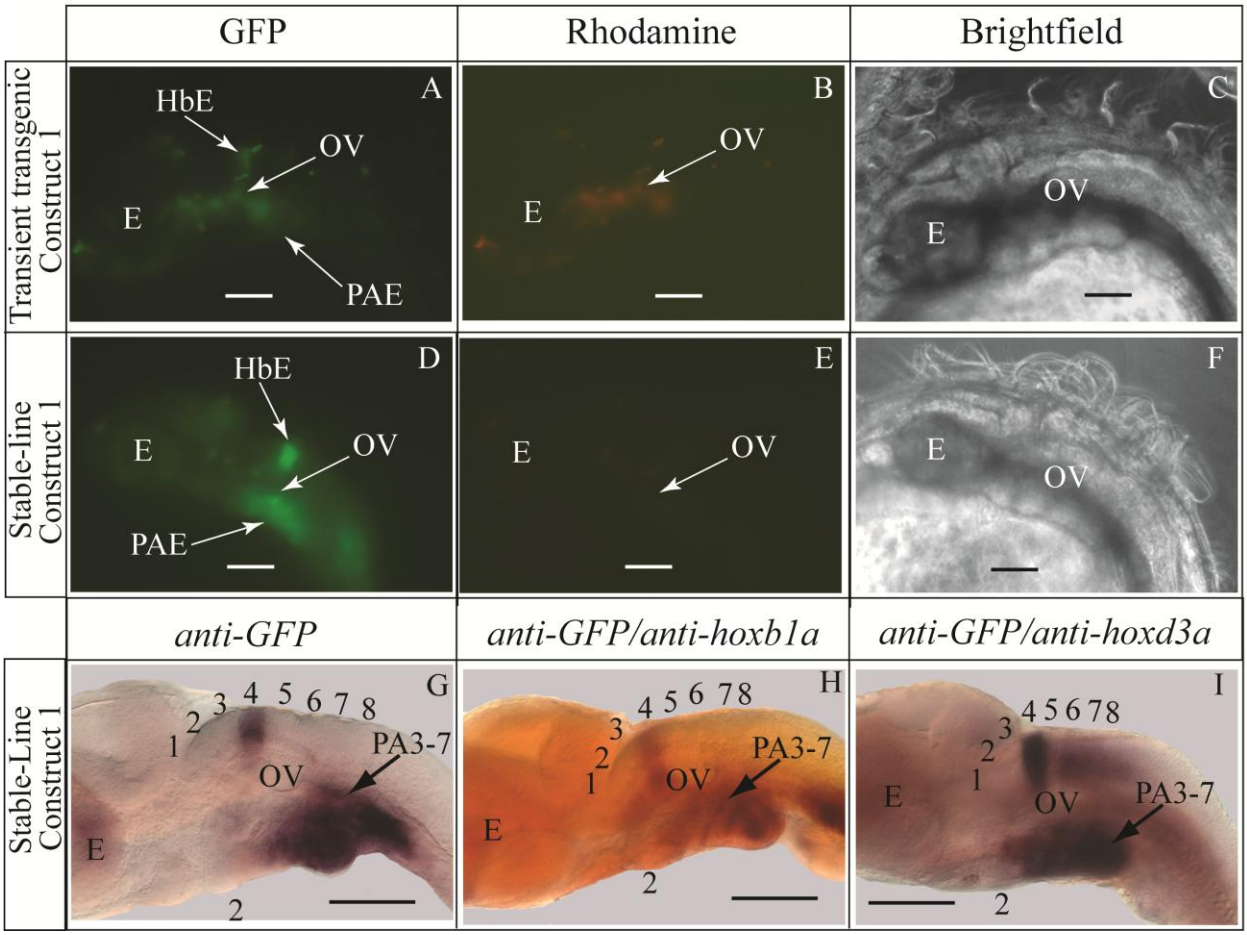


Fig. 3-3. Transient (A-C) and stable-line (D-I) transgenic data from the medaka *hoxa2a* r3/5ER (Construct #1). (A-F) Pictures of transient (A-C) and stable-line (D-F) transgenic embryos at stage 29/30 (72-84 hpf) were taken using GFP (A, D), rhodamine (B, E) and brightfield (C, F) filters. All embryos are still in their chorions and are positioned with their anterior sides to the left and their lateral sides to the reader. (G-I) Whole-mount *in situ* hybridization of medaka embryos using DIG-labeled anti-*eGFP* riboprobe (G), DIG-labeled anti-*eGFP* riboprobe with fluorescein-labeled anti-*hoxb1a* riboprobe (H) and DIG-labeled anti-*eGFP* and anti-*hoxd3a* riboprobes (I). Embryos were mounted with anterior sides facing left and lateral sides facing the reader. Rhombomere numbers are indicated by black numbers above the dorsal sides of the embryos. Pharyngeal arch 2 is indicated by the number 2 below the ventral sides of

the embryos. E, eye; HbE, hindbrain expression; OV, otic vesicle; PA, pharyngeal arch; PAE, pharyngeal arch expression. Scale bars equal 0.1 mm.

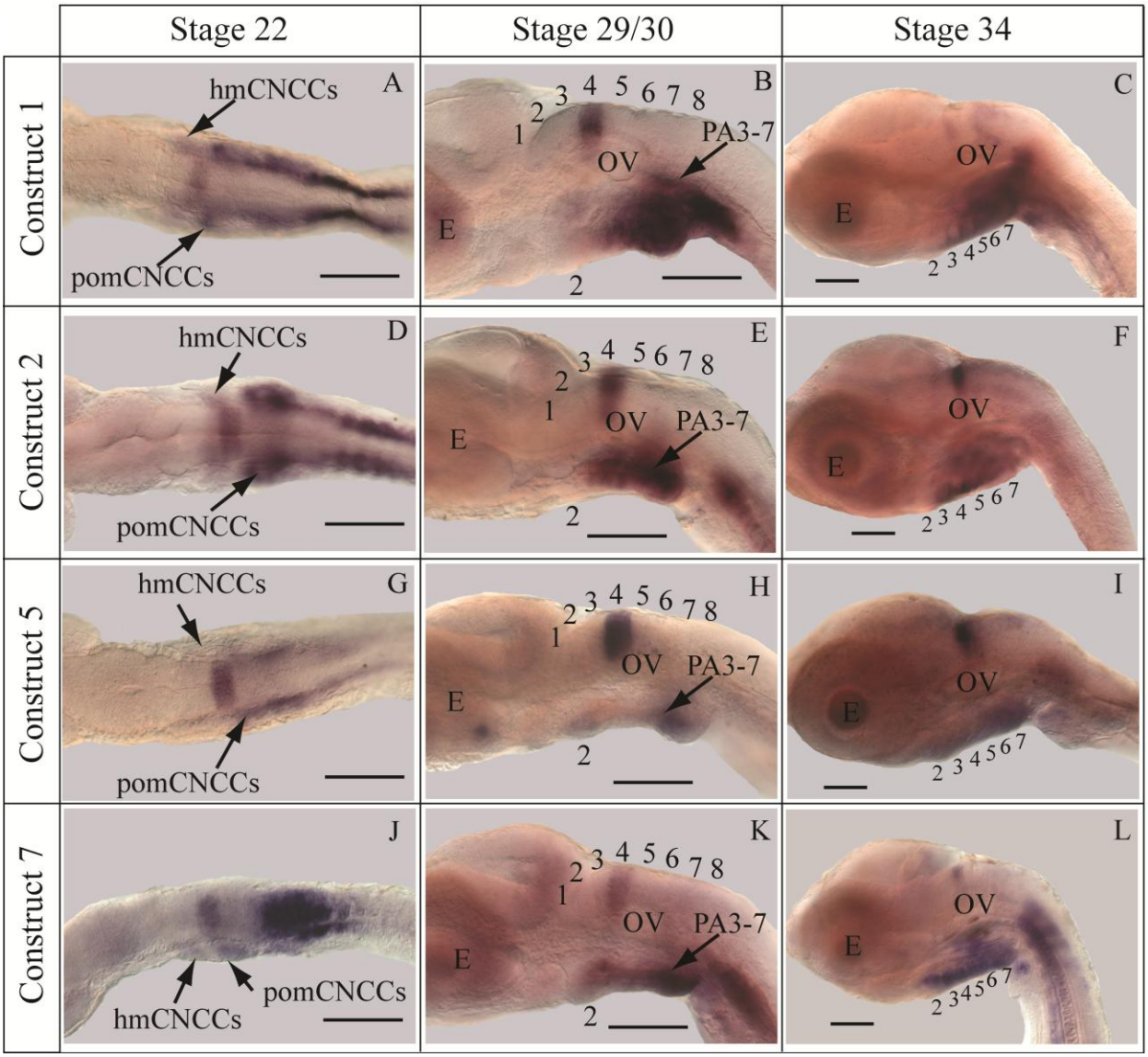


Fig.3-4. Whole-mount *in situ* hybridization of *eGFP* in stable-line *hoxa2a* r3/5ER transgenic medaka embryos generated with Construct #1 (A, B and C), Construct #2 (D, E and F), Construct #5 (G, H and I) and Construct #7 (J, K and L) at stages 22 (9 s) (A, D, G and J), 29/30 (72-84 hpf) (B, E, H and K) and 34 (121 hpf) (C, F, I and L). (A, D, G and J) Embryos were mounted with their anterior sides facing left and their dorsal sides facing the reader. (B, C, E, F, H, I, K and L) Embryos were mounted with their anterior sides facing left and lateral sides facing the reader. (B, E, H and K) Images are magnified to show rhombomere

placement. Rhombomere numbers are indicated by black numbers above the dorsal sides of the embryos. Pharyngeal arches are indicated by black numbers below the ventral sides of the embryos. E, eye; hmCNCCs, hyoid migratory cranial neural crest cells; OV, otic vesicle; PA, pharyngeal arch; pomCNCCs, post-otic migratory cranial neural crest cells. Scale bars equal 0.1 mm.

for the mouse *Hoxa2* r3/5ER, which was shown to direct reporter gene expression in r3, r5 and the CNCCs (Frasch et al., 1995; Nonchev et al., 1996a and b, Maconochie et al., 2001; Tümpel et al., 2002). Interestingly, these results are different from the functional genomic analysis of the medaka *hoxa2a* r3/5ER when it was tested in chicken embryos using electroporation of constructs containing the medaka *hoxa2a* r3/5ER. In the chicken embryo experiments, reporter gene expression was undetectable in the hindbrain, pharyngeal arches or any other sites (Tümpel et al., 2006).

A comparison of expression in transient and stable-line transgenic embryos generated with a series of nested deletion constructs extending from the 5'- and 3'-ends of the medaka *hoxa2a* r3/5ER was performed (Constructs #2-7). These constructs represented a set of nested deletions beginning from the 5'- and 3'-ends of a sequence that appeared to be orthologous to one that was shown to be responsible for directing reporter gene expression in r3 and r5 of the hindbrain in the mouse (*Hoxa2* r3/5ER) (Table 3-2). These deletions eliminated regions corresponding to previously mapped enhancer elements (Tümpel et al., 2006). A high percentages of transient transgenic embryos generated with construct #2, a 221 bp construct that spanned from genomic bp positions -1468 to -1247 and was devoid of Krox20 and BoxA but retained the RE4, RE3, RE2 and RE5 sequences orthologous to those previously mapped in tetrapods (Frasch et al., 1995; Nonchev et al., 1996a and b; Maconochie et al., 2001), showed *eGFP* expression in the hindbrain (76%) and pharyngeal arches (67%) (Table 3-2). Both transient and stable-line transgenic embryos generated with Construct #2 showed similar *eGFP* expression patterns to transgenic embryos generated with constructs containing the entire medaka *hoxa2a* r3/5ER (Construct #1). Specifically, we observed *eGFP* expression in r4 of the hindbrain, the migratory CNCCs of the hyoid and post-otic streams, the post-migratory CNCCs

of the posterior pharyngeal arches and the chondrogenic CNCCs of the posterior arches in construct #2-generated embryos (Fig. 3-4D-F). A high percentage of transient transgenic embryos generated with Construct #3, a 145 bp construct that spanned from genomic bp positions -1392 to -1247 and was devoid of Krox20, BoxA and RE4 sequences but retained RE3, RE2 and RE5 sequences (Table 3-2), showed *eGFP* expression in the hindbrain (80%) and pharyngeal arches (84%). While this expression was similar to the *eGFP* expression in transient transgenic embryos generated with Construct #1 and #2, we did not have success in obtaining stable-line transgenic embryos (see Appendix C). A much lower percentage of transient transgenic embryos generated with Construct #4, a 107 bp construct that spanned from genomic bp positions -1354 to -1247 and was devoid of Krox20, BoxA, RE4 and RE3 but retained RE2 and RE5 sequences (Table 3-2), showed *eGFP* expression in the hindbrain (15%) and in the pharyngeal arches (15%) when compared to embryos generated with Constructs #1, 2 and 3. Further, these embryos showed *eGFP* levels of expression that were barely detectable in the hindbrain and pharyngeal arches (data not shown). Unfortunately, we did not obtain any stable-line transgenic embryos generated with Construct #4 (Appendix C). A high percentage of transient transgenic medaka embryos generated with Construct #5, a 475 bp construct that spanned from genomic bp positions -1778 to -1303 and included the sequences orthologous to the Krox20, BoxA, RE4, RE3 and RE2 sequences of the mouse *Hoxa2* r3/5ER but was devoid of the RE5 sequence showed *eGFP* expression in the hindbrain (84%) and pharyngeal arches (84%) (Table 3-2). Stable-line transgenic embryos generated with Construct #5 showed a similar *eGFP* expression pattern to stable-line embryos generated with Construct #1 and #2, wherein expression was observed in r4 of the hindbrain, the migratory CNCCs of the hyoid and post-otic streams, the post-migratory CNCCs in PA2 and the posterior pharyngeal arches and the

chondrogenic CNCCs of PA2 and the posterior arches (Fig. 3-4G-I). Transient transgenic embryos generated with Construct #6, a 430 bp construct that spanned from genomic bp positions -1778 to -1348 and included the sequences for Krox20, BoxA, RE4 and RE3 but was devoid of the RE5 and RE2 sequences, did not show any *eGFP* expression in the hindbrain (0%) or in the pharyngeal arches (0%) (Table 3-2). Overall, these results showed that the sequence elements corresponding to Krox20, BoxA, RE4 and RE5 of the medaka *hoxa2a* r3/5ER were not necessary for directing reporter gene expression in the hindbrain or pharyngeal arches. However, the genomic region encompassing the RE3 and RE2 sequences were shown to be necessary for r4- and pharyngeal arch-directed reporter gene expression. Further, it appears that sequence elements embedded within regions corresponding uniquely to RE3 or RE2 interact cooperatively to modulate qualitative expression levels in r4.

The aforementioned *eGFP* expression patterns obtained from the nested deletion constructs of the medaka *hoxa2a* r3/5ER prompted us to develop a construct that only encompassed sequences that included the RE3 and RE2 elements (Construct #7; Table 3-2). This construct was 89 bp in length and spanned from genomic bp positions -1392 to -1303 (Table 3-2). As expected, a high percentage of transient transgenic medaka embryos generated with Construct #7 showed *eGFP* expression in the hindbrain (84%) and in the pharyngeal arches (84%) (Table 3-2). Stable-line transgenic medaka embryos showed *eGFP* expression in r4 of the hindbrain, the migratory CNCCs of the hyoid and post-otic streams and the post-migratory and chondrogenic CNCCs of PA2 and the posterior pharyngeal arches (Fig. 3-4J-L). Interestingly, stable-line transgenic embryos generated with Construct #7 showed much higher levels of *eGFP* expression in the post-migratory and chondrogenic CNCCs of PA2 than stable-line transgenic embryos generated with Constructs #1, 2 or 5 (see Fig. 3-4). Further, the *eGFP* expression in

PA2 was observed to be restricted to the ventral domain (Fig. 3-4K and L). These results show that there are transcription elements in the sequence spanning from position -1392 to -1303 upstream of medaka *hoxa2a* that are responsible for directing reporter gene expression in r4 and the migratory and post-migratory CNCCs. This sequence will be termed the r4/CNCC-specifying element hereafter. Further, these results also indicate that sequences flanking the r4/CNCC-specifying element repress this element from driving strong reporter gene expression in the ventral domain of PA2.

To determine putative transcription factor binding sites in the 89 bp r4/CNCC-specifying element, we performed a comparative geomic sequence analysis of this DNA sequence fragment with orthologous sequences from r3/5ERs upstream of *Hoxa2* genes of several vertebrates, including horn shark, dogfish, *latimeria*, chicken, mouse, human, bichir, zebrafish, tilapia, striped bass, fugu and medaka. Sequence alignment of the entire *Hoxa2* r3/5ER is shown in Appendix D. Sequence alignments revealed three regions of sequence that were highly conserved among vertebrate *Hoxa2* genomic sequences (Fig. 3-5). Analysis of these sequences showed that they corresponded to a Prep/Meis binding site (5'-CTGTCA-3') beginning at bp position -1371 of the medaka *hoxa2a* r4/CNCC-specifying element, a Hox/Pbx binding site (5'-AGATTGATCG-3') beginning at bp postion -1349 and a lymphoid enhancer binding factor 1 (LEF-1) binding site (5'-CTTTGTT-3') beginning at bp position -1320 (Fig. 3-5). Interestingly, the Prep/Meis and Hox/Pbx sites of medaka *hoxa2a* were observed to be identical in sequence to the functionally mapped Prep/Meis and Hox/Pbx binding sites in the mouse *Hoxb3-b2* intergenic region (Ferretti et al., 2000). The mouse Prep/Meis and Hox/Pbx binding sites were shown by

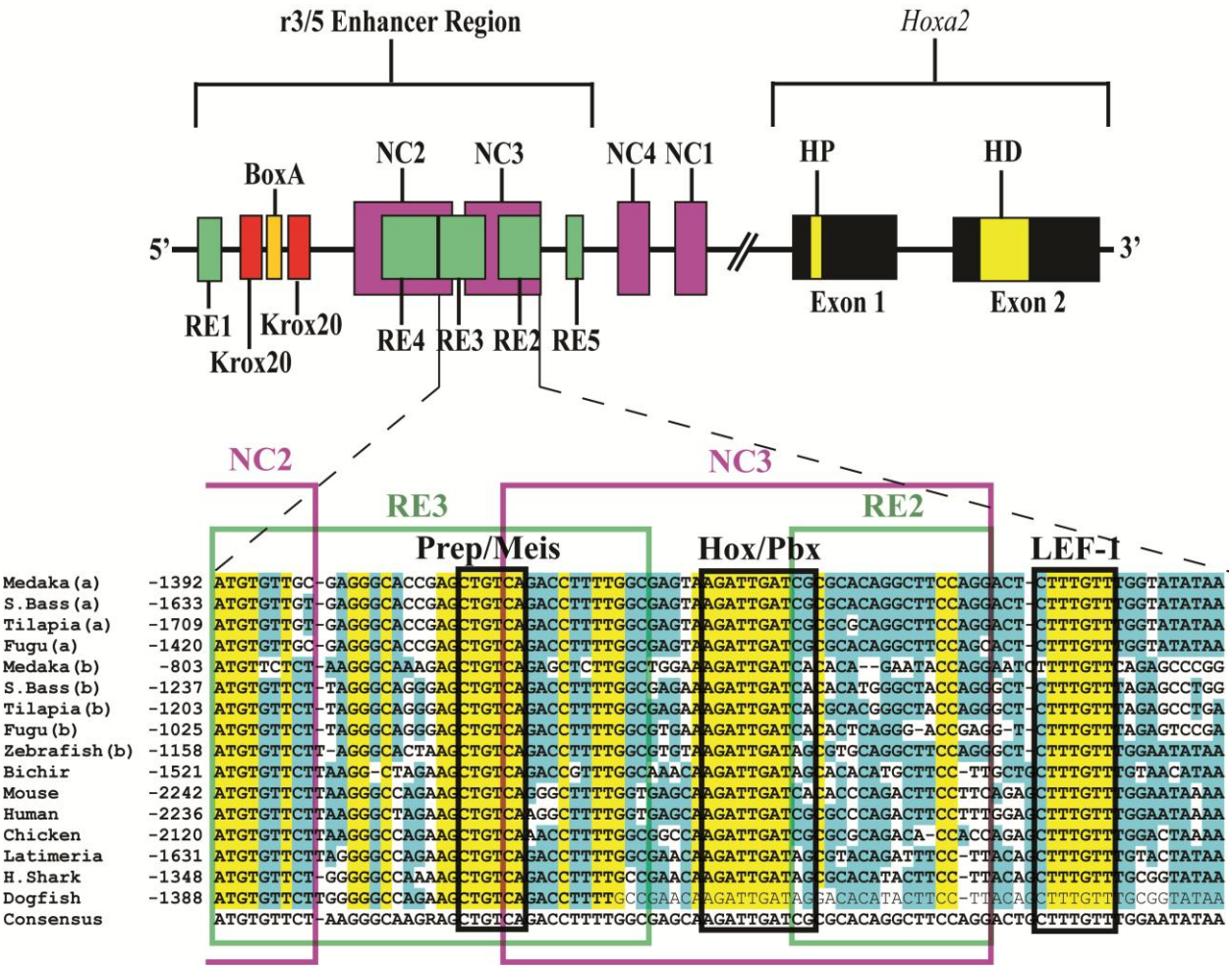


Fig. 3-5. Comparative genomic sequence analysis of the 89 bp DNA fragment of the medaka *hoxa2a* r3/5ER (top sequence) that directs expression in r4 and the CNCCs. This sequence was compared to orthologous DNA sequences of *hoxa2a* (denoted by a in parentheses after species name), *a2b* (denoted by b in parentheses after species name) of teleosts, *Hoxa2* of tetrapods and *Hoxa2* of *Latimeria*, horn shark and dogfish. Numbers correspond to genomic base pair positions relative to the ATG start site of the *Hoxa2* genes. The schematic diagram above the sequences corresponds to the *Hoxa2* r3/5ER and the relative location of the 89-bp DNA fragment of the medaka *hoxa2a* r4/CNCC-specifying element. The schematic is not drawn to scale. Base pairs colored in yellow correspond to complete conservation at particular sites

across all sequences examined. Base pairs colored in blue represent the majority of the sequences containing specific base pairs at specific sites. Purple boxed regions and green boxed elements correspond to neural crest and rhombomeric elements defined in the mouse r3/5ER (Maconochie et al., 1999 and 2001). Black boxed regions correspond to transcription factor binding sites identified in this study. A consensus sequence was derived from the aligned sequences. HD, homeodomain; HP, hexapeptide; LEF, lymphoid enhancer binding factor; NC, neural crest; RE, rhombomeric element.

Ferretti et al. (2000) to direct mouse *Hoxb2* expression in r4 and the migratory CNCCs. Further, the Prep/Meis site was shown to be 100% identical among all vertebrate *Hoxa2* genomic sequences analyzed. The Hox/Pbx site was shown to be nearly identical in sequence with exception to the last two bp of this sequence among all vertebrate *Hoxa2* genomic sequences. Several Hox/Pbx sites that were shown to be involved in directing the expression of several *Hox* genes, including *labial* of *Drosophila* and *Hoxb1, a2, b2* and *a3* of tetrapods were identical at positions 2, 3, 4 and 8, which were G, A, T and T respectively, as is the case in our analysis, but differed in the remaining two bp (Tümpel et al. 2007), which suggests that this variation does not alter Hox/Pbx binding sites from directing *Hox* expression. It is interesting that the orthologous sequences of mouse and chicken contain these highly conserved elements but do not direct expression in r4 of the hindbrain. Rather, Hox/Pbx and Prep/Meis binding sites located in Exon 1 and the intron of *Hoxa2* of chicken and mouse were shown to direct expression in r4 (Tümpel et al., 2007; Lampe et al., 2008). Therefore, it is possible that there are sequence elements located in the *Hoxa2* r3/5ER of tetrapods that restrict the genomic sequences located between RE3 and RE2 from directing expression in r4. Interestingly, the LEF-1 transcription factors contains a DNA binding domain of the High Mobility Group (HMG) box type and have been shown to be expressed in both migratory CNCCs as well in as the pharyngeal arches (Oosterwegel et al., 1993). However, mouse strains that were deficient of LEF-1 did not show any developmental abnormalities in structures derived from PA2 or the posterior pharyngeal arches (van Genderen et al., 1994).

Overall, our transient and stable-line transgenic results of the medaka *hoxa2a* r3/5ER show that this region is divergent in function from the orthologous r3/5ER upstream of mouse *Hoxa2*. While both regions were shown to direct expression in the CNCCs originating from the

hindbrain, the mouse r3/5ER directed reporter gene expression in r3 and r5 whereas the medaka *hoxa2a* r3/5ER contained an r4/CNCC-specifying element that potentiated reporter gene expression in r4 (Frasch et al., 1995; Nonchev et al., 1996a and b; Maconochie et al., 1999 and 2001). Further, our results differ from those reported by Tümpel et al. (2006), which showed that the medaka *hoxa2a* r3/5ER did not direct any reporter gene expression within the hindbrain or CNCCs of the chicken embryonic head.

Functional Genomic Analysis of the Medaka *ψhHoxa2b* r3/5 Enhancer Region

Unlike the conserved hindbrain and pharyngeal arch expression patterns common to many teleost *hoxa2a* and *a2b* and tetrapod *Hoxa2* genes, medaka *ψhoxa2b* is expressed in noncanonical *Hox* PG2 domains, which include the caudal-most region of the embryonic trunk, the ventral-most aspect of the neural tube and the distal regions of the pectoral fin buds (Davis et al., 2008). Based on the divergence in expression observed in comparisons of medaka with other teleost and tetrapod *hoxa2* genes, we hypothesized that the r3/5ER of medaka *ψhoxa2b* would not direct reporter gene expression in the hindbrain or CNCCs.

Contrary to our hypothesis, transient and stable-line transgenic medaka embryos showed that the 397 bp construct containing the entire medaka *ψhoxa2b* r3/5ER and spanning from genomic bp positions -1068 to -671 (with respect to the ATG start site of medaka *ψhoxa2b*) (Construct #8, Table 3-2) directed reporter gene expression in the hindbrain and in the pharyngeal arches (Fig. 3-6A-I). We observed a high percentage of our transient transgenic embryos showing *eGFP* expression in the hindbrain (90%) and pharyngeal arches (90%) (Table 3-2). Further, whole-mount *in situ* hybridization analyses on stable-line transgenic embryos

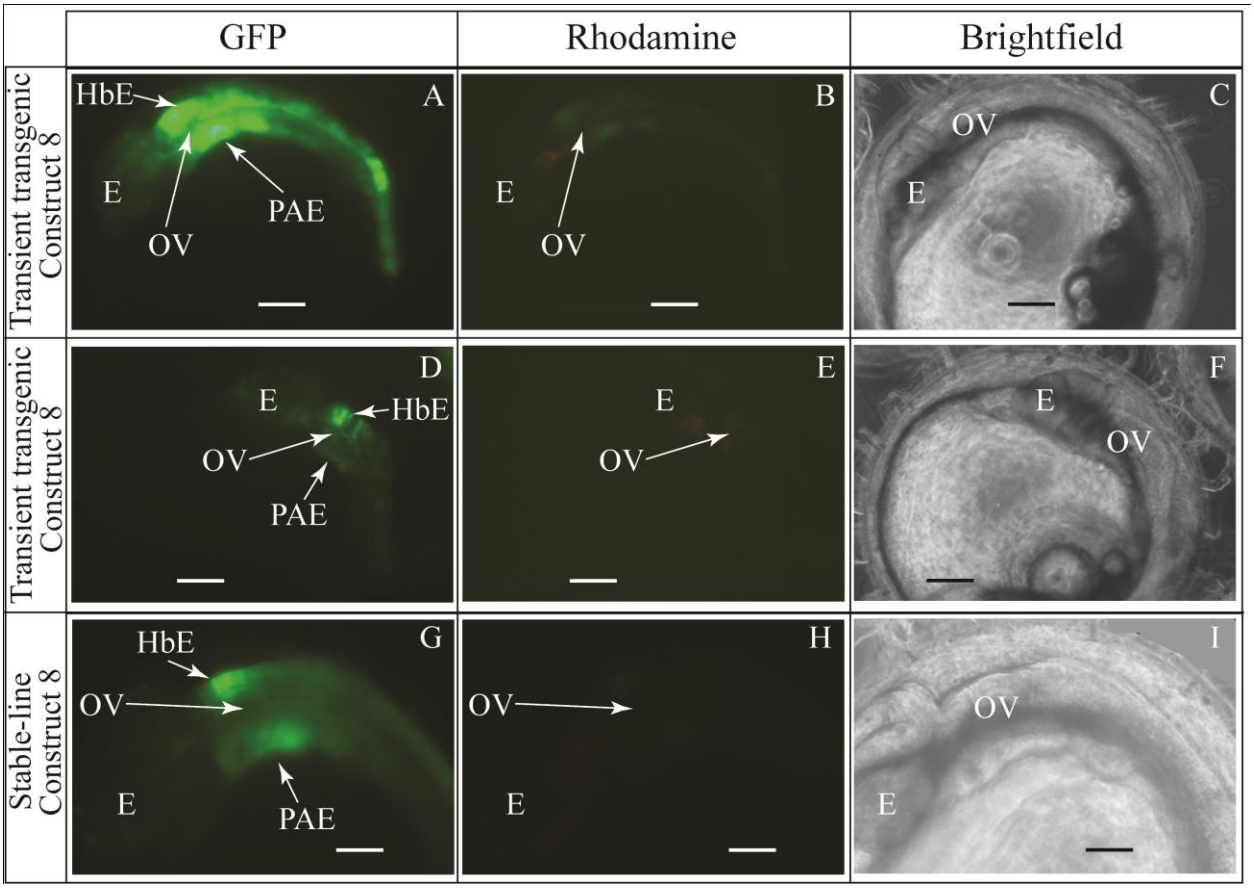


Fig. 3-6. Transient (A-F) and stable-line (G-I) transgenic data from the medaka *ψhoxa2b* r3/5ER (Construct #8). (A-I) Pictures of transient transgenic embryos at stage 29/30 (72-84 hpf) were taken using GFP (A, D and G), rhodamine (B, E and H) and brightfield (C, F and I) filters. Transient transgenic analyses show varying degrees of *eGFP* signal between embryos (compare A and D). All embryos are still in their chorions and are positioned with their anterior sides to the left and their lateral sides to the reader.

generated with construct #8 showed that *eGFP* was expressed in r3-7 of the hindbrain, the migratory CNCCs of the hyoid and post-otic streams, the post-migratory CNCCs in PA2 and the posterior pharyngeal arches and the chondrogenic CNCCs in PA2 and the posterior arches (Fig. 3-7A-C). These results show that the medaka *ψhoxa2b* r3/5ER is functionally active and possesses the capability of directing reporter gene expression in the hindbrain and in the pharyngeal arches, despite the fact that medaka *ψhoxa2b* expression is present in noncanonical *Hox* PG2 expression domains but absent of the hindbrain and pharyngeal arches (Davis et al., 2008). Interestingly, in transgenic reporter gene assays conducted in chicken embryos, a very similar construct to that tested here in the homologous medaka system, directed reporter gene expression in r3 and r5 but not in the CNCCs or the pharyngeal arches (Tümpel et al., 2006).

A comparison of expression in transient and stable-line transgenics generated with a series of nested deletion constructs extending from the 5'- and 3'-ends of the medaka *ψhoxa2b* r3/5ER was performed (Constructs #9-13). These constructs represented a set of nested deletions beginning from the 5'- and 3'-ends of a sequence that appeared to be orthologous to one that was shown to be responsible for directing reporter gene expression in r3 and r5 of the hindbrain in the mouse (*Hoxa2* r3/5ER) and r4 and the CNCCs for medaka *hoxa2a* (Table 3-2). These deletions eliminated regions corresponding to previously mapped enhancer elements (Tümpel et al., 2006). A high percentage of transient transgenic embryos generated with Construct #9, a 197 bp construct that spanned from genomic bp positions -868 to -671 and was devoid of Krox20 and BoxA sequences but retatined RE4, RE3, RE2 and RE5 sequences (Table 3-2), showed *eGFP* expression in the hindbrain (90%) and the pharyngeal arches (90%) (Table 3-2). Stable-line

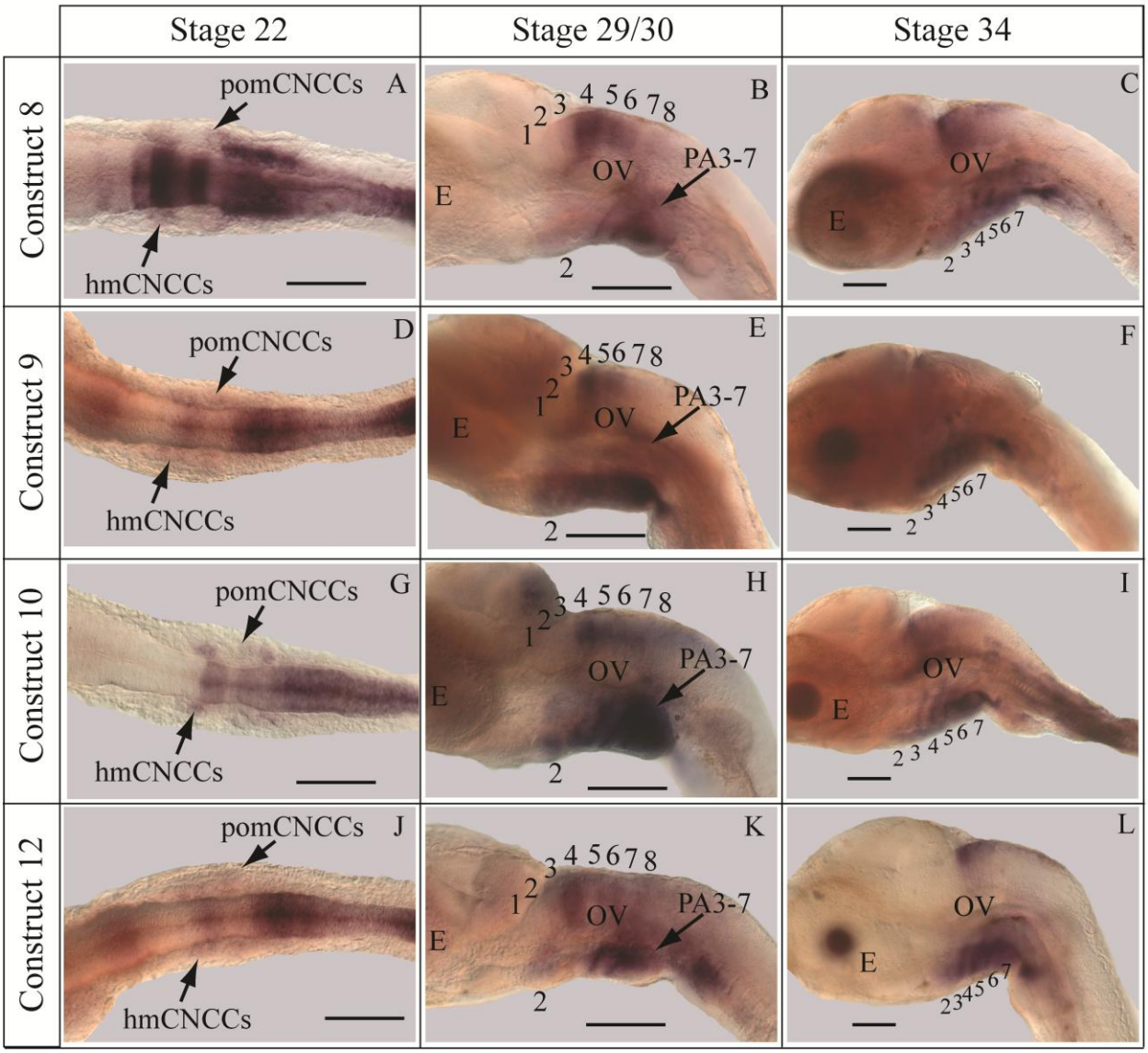


Fig.3-7. Whole-mount *in situ* hybridization of *eGFP* in stable-line *ψhoxa2b* r3/5ER transgenic medaka embryos generated with Construct #8 (A, B and C), Construct #9 (D, E and F), Construct #10 (G, H and I) and construct #12 (J, K and L) at stages 22 (9 s) (A, D, G and J), 29/30 (72-84 hpf) (B, E, H and K) and 34 (121 hpf) (C, F, I and L). (A, D, G and J) Embryos were mounted with their anterior sides facing left and their dorsal sides facing the reader. (B, C, E, F, H, I, K and L) Embryos were mounted with their anterior sides facing left and lateral sides facing the reader. (B, E, H and K) Images are magnified to show rhombomere

placement. Rhombomere numbers are indicated by black numbers above the dorsal sides of the embryos. Pharyngeal arches are indicated by black numbers below the ventral sides of the embryos. E, eye; hmCNCCs, hyoid migratory cranial neural crest cells; OV, otic vesicle; PA, pharyngeal arch; pomCNCCs, post-otic migratory cranial neural crest cells. Scale bars equal 0.1 mm.

transgenic embryos generated with Construct #9 showed a similar *eGFP* expression pattern to stable-line transgenic embryos generated with Construct #8, wherein reporter gene expression was observed in r3-7 of the hindbrain, the migratory CNCCs of the hyoid and post-otic streams, the post-migratory CNCCs in PA2 and the posterior pharyngeal arches and the chondrogenic CNCCs of PA2 and the posterior arches (Fig. 3-7D-F). A high percentage of transient transgenic embryos generated with Construct #10, a 132 bp construct that spanned from genomic bp positions -803 to -671 and that was devoid of the Krox20, BoxA and RE4 sequences but retained the RE3, RE2 and RE5 sequences, showed *eGFP* expression in the hindbrain (87%) and the pharyngeal arches (87%) (Table 3-2). Stable-line transgenic embryos generated with Construct #10 showed a similar *eGFP* expression pattern to stable-line embryos generated with Constructs #8 and 9, which included expression in r3-r7 of the hindbrain, the migratory CNCCs of the hyoid and post-otic streams and the post-migratory and chondrogenic CNCCs in PA2 and the posterior arches (Fig. 3-7G-I). A lower percentage of transient transgenic medaka embryos generated with Construct #11, a 94 bp construct that spanned from genomic bp positions -765 to -671 and that was devoid of Krox20, BoxA, RE4 and RE3 but retained RE2 and RE5, showed *eGFP* expression in the hindbrain (52%) and pharyngeal arches (52%) in a lower percentage of transgenic embryos when compared to embryos generated with Constructs #8, 9 or 10 (Table 3-2). Further, in comparison to embryos generated with Constructs #8, 9 and 10, *eGFP* expression was barely visible in embryos generated with Construct #11 (data not shown). Unfortunately, no stable-line transgenic embryos generated with Construct #11 showed any detectable *eGFP* expression in the hindbrain or pharyngeal arches (see Appendix C). A high percentage of transient transgenic embryos generated with Construct #12, a 353 bp construct that spanned from genomic bp positions -1068 to -715 and that is devoid of RE5 but retained Krox20, BoxA, RE4,

RE5 and RE2, showed *eGFP* expression in the hindbrain (80%) and the pharyngeal arches (80%) (Table 3-2). Stable-line transgenic embryos generated with Construct #12, showed a similar *eGFP* expression pattern to stable-line transgenic embryos generated with Constructs #8, 9 and 10, wherein *eGFP* expression was observed in r3-r7 of the hindbrain and the migratory CNCCs of the hyoid and post-otic streams as well as the post-migratory and chondrogenic CNCCs of PA2 and the posterior phrarygneal arches (Fig. 3-7J-L). A low percentage of transient transgenic embryos generated with Construct #13, a 309 bp construct that spanned from genomic bp positions -1068 to -759 and was devoid of the RE2 and RE5 sequences but retained the Krox20, BoxA, RE4 and RE3 sequences, showed *eGFP* expression in the hindbrain (14%) and pharyngeal arches (23%) (Table 3-2). Further, in comparison to embryos generated with Constructs #8, 9, 10 and 12, *eGFP* expression in embryos generated with Construct #13 was significantly reduced and barely visible (data not shown). Unfortunately, we did not obtain any stable-line transgenic embryos generated with Construct #13 (Appendix C). Overall, our nested deletion constructs of the medaka *ψhoxa2b* r3/5ER showed that sequence elements orthologous to Krox20, BoxA, RE4 and RE5 of the mouse *Hoxa2* r3/5ER were not required to direct reporter gene expression in the hindbrain and CNCCs. Remarkably, our analysis showed that the sequence that was responsible for directing *eGFP* in the hindbrain and CNCCs spanned from genomic bp positions -803 to -715 of the medaka *ψhoxa2b* r3/5ER and contained the elements that are orthologous to the RE3 and RE2 sequences of the mouse *Hoxa2* r3/5ER. Further, this sequence was observed to be orthologous to the region of the medaka *hoxa2a* r4/CNCC-specifying element (Fig. 3-5).

A comparative genomic sequence analysis of the 88 bp DNA sequence fragment that spanned from -803 to -715 upstream of medaka *ψhoxa2b* showed that this sequence, like the

paralogous *hoxa2a* sequence of medaka and orthologous *Hoxa2* genomic sequences of other vertebrates contains conserved Hox/Pbx, Prep/Meis and LEF-1 transcription factor binding sites (see Fig. 3-5). While these sites may be involved in driving *eGFP* expression in r4 and the CNCCs, other sites in this DNA sequence are involved in driving expression in r3, r5, r6 and r7. We observed 33 bp differences between the paralogous sequences of the r4/CNCC-specifying sequence element for medaka *hoxa2a* and the r3-7/CNCC-specifying sequence element for medaka *ψhoxa2b* (Fig. 3-8). It is possible that these substitutions between the medaka *hoxa2a* and *ψhoxa2b* genomic sequences have allowed the *ψhoxa2b* sequence, but not the *hoxa2a* sequence, to be receptive to transcription factors expressed within r3, r5, r6 and r7 of the hindbrain. An analysis of these sequences in the JASPAR software program showed the presence of several Sox binding elements within the *ψhoxa2b* r3-7/CNCC-specifying element but not in the *hoxa2a* r4/CNCC-specifying element. Sox proteins have been shown to be involved in driving *Hox* gene expression in several rhombomeres of the hindbrain (Tümpel et al., 2009). The base pair mutations between the the 88-bp sequence upstream of medaka *ψhoxa2b* and the paralogous r4/CNCC-specifying element may have occurred due to relaxation of selective pressures on the *ψhoxa2b* genomic sequence after the inactivation of the *hoxa2b* gene in the lineage leading to medaka (Davis et al., 2008). However, it must be reiterated that medaka *ψhoxa2b* is not expressed in the hindbrain or pharyngeal arches. Therefore the r3-7/CNCC-specifying element upstream of medaka *ψhoxa2b* may be redirected by other *cis*-regulatory sequences to drive expression of medaka *ψhoxa2b* in the noncanonical *Hox* PG2 expession domains.

Fig. 3-8. Sequence alignment of the medaka *hoxa2a* r4/CNCC and the *ψhoxa2b* r3-7/CNCC specifying elements. The sequence corresponding to the medaka *hoxa2a* r4/CNCC specifying element is denoted by "a" in parentheses. The sequence corresponding to the medaka *ψhoxa2b* r3-7/CNCC specifying element is denoted by "b" in parentheses. Numbers correspond to genomic base pair positions relative to the ATG start site of *hoxa2a* and *ψhoxa2b*. Base pairs colored in yellow correspond to complete conservation at particular sites across both sequences examined. Black boxed regions correspond to transcription factor binding sites identified in this study. LEF, lymphoid enhancer binding factor.

Discussion

The use of Medaka in Reporter Gene Expression Analyses

The *Tol2* transposon system has been used for the generation of transgenic vertebrate lines for several osteichthyans, including zebrafish, *Xenopus*, chicken and mouse (Kawakami, 2007). In this study we showed that the *Tol2* transposon system can be used for both transient and stable-line transgenic analyses of genomic enhancer regions in the Japanese medaka. Although the *Tol2* element is found within the genome of medaka, our control experiments using constructs containing the *Xenopus* eFI-αS promoter upstream of *eGFP* showed that this system works similarly in medaka to that of other osteichthyans. Specifically, medaka zygotes that were co-microinjected with constructs containing the *Xenopus* eFI-αS promoter and transposon mRNA showed strong reporter gene expression throughout the body whereas embryos that were microinjected solely with constructs containing the *Xenopus* eFI-αS promoter were not able to direct reporter gene expression. Similar results were observed by Koga and Hori (1999) who failed to detect reporter gene expression from *Tol2* in the absence of treatment with exogenous transposon mRNA.

Our reporter gene expression results showed that both r3/5ERs of medaka *hoxa2a* and *ψhoxa2b* are functional but that they have diverged from one another in their capacity to direct gene expression in the medaka hindbrain and pharyngeal arches. The medaka *hoxa2a* r3/5ER-directed reporter gene expression in r4 of the hindbrain and the CNCCs of PA2-7 while the medaka *ψhoxa2b* r3/5ER directed reporter gene expression in r3-r7 of the hindbrain and PA2-7. These reporter gene expression results underscore the importance of using a homologous model system for analyzing *cis*-regulatory element control of gene expression during embryonic development. Although the medaka *hoxa2a* r3/5ER did not direct reporter gene expression in r3

or r5 of the hindbrain in transgenic medaka embryos in this study or chicken embryos in the study by Tümpel et al. (2006), the medaka *hoxa2a* r3/5ER was shown to direct reporter gene expression in r4 and the CNCCs of medaka embryos but not in chicken embryos. Further, although the medaka *ψhoxa2b* r3/5ER directed reporter gene expression in r3 and r5 in both medaka and chicken embryos, there was no reporter gene expression observed in r4, r6 and r7 of the hindbrain or in the CNCCs in chicken embryos (Tümpel et al., 2006). The lack of medaka *hoxa2a* and *ψhoxa2b* r3/5ER-driven reporter gene expression in r4 and the CNCCs and in r4, r6, r7 and the CNCCs, respectively, of chick embryos suggests that these enhancer regions were not efficient in utilizing the *trans*-acting factors that were present in the heterologous chick model system. In support of this hypothesis, Tümpel et al. (2002) showed that the r3/5ER of mouse *Hoxa2* was able to generate strong reporter gene expression in r3, r5 and the CNCCs in mouse embryos but no reporter gene expression was detected for the mouse *Hoxa2* r3/5ER in the hindbrain or CNCCs of heterologous chick embryos. Conversely, the r3/5ER of chick *Hoxa2* was able to direct reporter gene expression in r3 and r5 of both chick and mouse embryos (Tümpel et al., 2002). These results from Tümpel et al. (2002) along with our reporter gene expression results of the medaka *hoxa2a* and *ψhoxa2b* r3/5ERs in medaka embryos suggest that either the functional nature of the sequence elements of r3/5ER, the transcription factors of the genetic regulatory networks that interact with these sequences or a combination of both have diverged greatly among evolutionarily divergent osteichthyans. Future analyses that use heterologous model systems for studying *cis*-regulatory element evolution must take into account the possibility of divergence of expression systems between evolutionarily divergent species. Unfortunately, the difficulty of acquiring embryos in suitable quantities on a rountine basis and under controlled conditions for many animal species makes homologous reporter gene

expression analyses impossible to use for studying *cis*-regulatory element control among a broad range of animals.

Medaka *Hoxa2a*-Directed Expression in the Hindbrain

In this chapter we showed, using reporter gene assays and whole-mount *in situ* hybridization, that the r3/5ER upstream of medaka *hoxa2a* does not direct expression in r3 and r5 of the hindbrain as expected based on expression detected *in vivo* using whole mount *in situ* hybridization with *hoxa2a* directed riboprobes, but instead directs expression in r4. We found these results interesting since this region of DNA was shown to be conserved structurally with orthologous genomic regions in evolutionarily divergent osteichthyans (Tümpel et al., 2006 and 2007; J. D. Raincrow, 2010, Ph.D. Thesis, Rutgers University) in which they have been shown to direct reporter gene expression in r3 and r5 at least in mouse and chicken transgenic hosts (Frasch et al., 1995; Nonchev et al., 1996a and b; Maconochie et al., 2001). The absence of reporter gene expression in r3 and r5 in transgenic medaka embros suggests several possibilities for the functional nature of the medaka *hox2a* r3/5ER. The r3/5ER of medaka *hoxa2a* tested in our experiments may not be exclusively responsible for directing *hoxa2a* expression in r3 or r5. Our analysis showed that the presence of Krox20, BoxA, RE4, RE3, RE2 and RE5 sequences within the medaka *hoxa2a* r3/5ER did not contribute to reporter gene expression in the r3 and r5 of the medaka hindbrain. Similar constructs containing the same regulatory elements of the medaka *hoxa2a* r3/5ER did not direct reporter gene expression in transient transgenic analyses performed in chicken embryos (Tümpel et al., 2006). Interestingly, orthologous genomic sequences of the mouse *Hoxa2* r3/5ER were shown to be functionally active in driving expression in r3 and r5 in stable-line transgenic analyses using mouse embryos (Frasch et al., 1995; Nonchev et al., 1996a and b; Maconochie et al., 2001). Since whole-mount *in situ*

hybridization results showed that medaka *hoxa2a* is expressed in r3 and r5 (Davis et al., 2008), the possibility that the medaka *hoxa2a* r3/5ER does not potentiate *hoxa2a* expression in these rhombomeres would suggest that other regions of genomic DNA sequence outside of the putative r3/5ER of medaka *hoxa2a* may be responsible for r3- and r5-directed expression. To test this hypothesis, reporter gene expression analyses of genomic DNA regions that exclude the medaka *hoxa2a* r3/5ER sequences tested in this study must be performed.

An alternative explanation for the lack of medaka r3/5ER-driven reporter gene expression in r3 and r5 in transgenic medaka embryos is that the genomic DNA sequence tested in this study requires the cooperation of other surrounding sequences within the medaka *hoxa3a-a2a* intergenic region or even outside the intergenic region. Our reporter gene analysis of the medaka *hoxa2a* r3/5ER in this study utilized a 531 bp genomic DNA sequence that spanned from genomic bp positions -1778 to -1247 (with respect to the medaka *hoxa2a* ATG translational start site) and included all of the *cis*-regulatory elements of the medaka *hoxa2a* r3/5ER that were tested in the chick embryonic model system (Tümpel et al., 2006). These elements include a Krox20 binding sequence and sequence elements that pertain to BoxA, RE4, RE3, RE2 and RE5. Functional genomic analyses of the mouse *Hoxa2* r3/5ER included regulatory elements orthologous to those mentioned above as well as a RE1 sequence element located upstream of Krox20 (Maconochie et al., 2001). The deletion of the RE1 sequence element from the mouse *Hoxa2* r3/5ER resulted in the loss of reporter gene expression in r3 of stable-line transgenic mouse embryos (Maconochie et al., 2001). Interestingly, no sequence element orthologous to the RE1 sequence was located in any teleost *hoxa3a-a2a* intergenic sequences analyzed by Tümpel et al. (2006). In order to determine if the 531 bp genomic sequence that corresponds to the medaka *hoxa2a* r3/5ER functions in conjunction with flanking genomic sequences to direct

reporter gene expression in r3 and r5, functional genomic assays that utilize constructs containing the 531 bp fragment and flanking genomic sequences in transient and stable-line transgenic medaka embryos must be performed. We attempted to generate transgenic medaka embryos with constructs containing 1800 bp of the medaka *hoxa3a-hoxa2a* intergenic sequence. This sequence contained the r3/5ER analyzed in this study as well as surrounding sequences and resulted in few transient transgenic medaka embryos with positive *eGFP* expression in the hindbrain that included r4 as well as other rhombomeres (data not shown). Unfortunately, we were not able to generate any stable-line transgenic medaka embryos with this construct.

Our reporter gene expression and whole-mount *in situ* hybridization analyses are the first to show the presence of sequences located in the vertebrate *Hoxa3-a2* intergenic regions that are involved in directing reporter gene expression in r4. The *eGFP* expression in r4 of medaka was authenticated using antisense *eGFP* riboprobes and antisense ribpoprobes of specific rhombomeric molecular markers. These molecular markers included medaka *hoxb1a*, which is expressed exclusively in r4 of the hindbrain (Hurley et al., 2007), and medaka *hoxd3a*, which is expressed in r6, r7 and r8 (Davis et al., 2010). Microinjection of a series of nested deletion constructs derived from a 531 bp fragment that encompassed a previously mapped r3/5 enhancer element showed that the r4/CNCC-specifying element lies within a 89 bp fragment that is located upstream of medaka *hoxa2a* and spans from genomic bp positions -1392 to -1303. Further, comparative genomic sequence analyses showed the presence of transcription factor binding sites, Prep/Meis and Hox/Pbx, located within the r4-specifying element. Hox/Pbx and Prep/Meis binding sites are important sequences that are involved in directing *Hox* expression in r4 (e.g.: Ferretti et al., 2000; Tümpel et al., 2006 and 2007). Further, these sites were shown to be highly conserved in *Hoxa2* genomic sequences across several evolutionarily divergent vertebrates,

including shark, bichir, *Latimeria*, teleosts and tetrapods. Future analyses using gel-shift, chromatin immunoprecipitation (ChIP) and site-directed mutagenesis assays must be performed to determine if the Hox/Pbx and Prep/Meis sites within the r4/CNCC-specifying element upstream of medaka *hoxa2a* are functional and responsible for directing medaka *hoxa2a* in r4. Further, reporter gene analyses must be performed to determine if the orthologous regions across evolutionarily divergent teleost *hoxa3a-a2a* and *hoxa9b-a2b* intergenic sequences drive reporter gene expression in r4. However, since many teleost models are intractable for generating stable-line transgenic embryos, such as striped bass, heterologous reporter gene assays using stable-line transgenic medaka and zebrafish models must be performed. Such analyses would provide an excellent experiment to determine if both the *cis*-regulatory elements and *trans*-acting factors that drive *hoxa2a* and *a2b* genes are conserved across evolutionarily divergent teleosts.

Based on functional results from reporter gene assays of the mouse *Hoxa2* and medaka *hoxa2a* r3/5ERs it is reasonable to conclude that there has been evolutionary divergence in these genomic sequences and systems responsible for controlling expression of the cognate coding sequences between the medaka and mouse lineages. It is interesting that the functional genomic analyes of the mouse *Hoxa2* r3/5ER performed by Maconochie et al. (2001) did not reveal any sequences involved in directing reporter gene expression in r4, especially since the mouse r3/5ER contains an orthologous sequence that is conserved with the r4/CNCC-specifying element of medaka *hoxa2a* (see Fig. 3-5). Further, genomic sequences upstream of mouse *Hoxa2* and medaka *hoxa2a* share conserved Hox/Pbx and Prep/Meis sites. Since homologous reporter gene studies of the mouse *Hoxa2* r3/5ER did not show expression in r4, it is possible that mouse has evolved sequence modulators that restrict the Hox/Pbx and Prep/Meis sites within the r3/5ER from directing expression in r4. Interestingly, the chicken *Hoxa2* r3/5ER also

contains Hox/Pbx and Prep/Meis binding sites that are conserved with medaka *hoxa2a* (see Fig. 3-5). However, homologous reporter gene studies of the chicken r3/5ER within chick embryos showed that this genomic region only potentiates expression in r3 and r5 (Tümpel et al., 2002), which suggests that the chicken *Hoxa2* r3/5ER possesses similar mechanisms to mouse in repressing its function in directing expression in r4. Recent reporter gene expression analyses of the osteichthyan *Evx1* enhancer region showed that the deletion of a conserved 25-bp sequence within the enhancer resulted in the expansion of *eGFP* expression from occurring specifically in the V0 spinal interneurons to a widespread pattern throughout the neural tube (Suster et al., 2011). These results suggested the presence of repressive elements within the *Evx1* enhancer that restrict *Evx1* from being expressed throughout the entire neural tube (Suster et al., 2011). Similar repressive elements within the tetrapod *Hoxa2* r3/5ER may function to restrict the r3/5ER from potentiating expression within r4. An alternative interpretation is that there is a molecular mechanism in medaka and/or other teleosts that redirect the r3/5 elements, such as RE3 and RE2, to specify *hoxa2a* expression in r4. Functional genomic analyses of r3/5ERs in species lineages that predate the evolutionary split of ray-finned from lobe-finned fishes must be performed in order to determine the ancestral function of the vertebrate *Hoxa2* r3/5ER. Since the shark *Hoxa2* gene is expressed in r2-5 of the hindbrain and the shark *Hoxa2* r3/5ER has been shown to be conserved with the *Hoxa2* r3/5ERs of vertebrates (Tümpel et al., 2002; Oulion et al., 2011) and possesses conserved Hox/Pbx and Prep/Meis sites that are orthologous to the binding sites of the medaka *hoxa2a* r4-specifying element (see Fig. 3-5), this would be an excellent model to use to test the ancestral function of the vertebrate *Hoxa2* r3/5ER.

Our comparative genomic analysis of the medaka *hoxa2a* r4/CNCC-specifying element with orthologous sequences in evolutionarily divergent vertebrates showed the presense of a

conserved LEF-1 binding site within the medaka *hoxa2a* r4/CNCC-specifying element and other orthologous genomic *Hoxa2* sequences from evolutionarily divergent vertebrates. The LEF-1 protein contains a HMG domain that has been shown to induce bends in the DNA helix, thus allowing LEF-1 to serve as an architectural factor that can potentiate the formation of unique chromatin configurations (Giese et al., 1992). Provided that LEF-1 is expressed in r4 of medaka, this transcription factor may serve architecturally to allow for increased interaction between the Hox/Pbx and Prep/Meis transcription factor binding sites between the medaka r4/CNCC-specifying element and the Hox/Pbx and Prep/Meis binding sites within the medaka *hoxa2a* intronic DNA that were identified by Tümpel et al. (2006). However, it must be stressed that heterologous reporter gene assays of the Hox/Pbx and Prep/Meis sites within the medaka *hoxa2a* intronic DNA did not direct reporter gene expression in r4 of chicken embryos (Tümpel et al., 2006). Future homologous functional genomic assays must be performed in order to determine if the intronic Hox/Pbx and Prep/Meis binding sites are functional for medaka *hoxa2a* and if these sites function in conjunction with LEF-1 and the Hox/Pbx and Prep/Meis sites in the r4/CNCC-specifying element upstream of medaka *hoxa2a*.

***Hoxa2a*-Directed Expression in the Cranial Neural Crest Cells**

Beyond the hindbrain expression directed activity of the medaka *hoxa2a* r3/5ER, this genomic DNA region was also shown to possess functional activity that directs *hoxa2a* expression in the CNCCs migrating to and populating PA2 and the posterior arches. Our whole-mount *in situ* hybridization results from stable-line transgenic medaka embryos have shown that the medaka *hoxa2a* r3/5ER drives reporter gene expression in the migratory CNCCs and is involved in maintaining *hoxa2a* expression in post-migratory CNCCs in PA2 and the posterior arches up until the chondrogenic stages of visceral skeletal element development. Several

functional genetic studies have shown that *Hox* PG2 gene expression that persists late into post-migratory CNCC stages of development is necessary for the proper patterning of the cranio-facial skeletal elements (Hunter and Prince, 2002; Baltzinger et al., 2005; Santagati et al., 2005; Crump et al., 2006; Le Pabic et al., 2010). Therefore, it is possible that the genomic sequences of the medaka r3/5ER are utilized *in vivo* to maintain medaka *hoxa2a* expression in the CNCCs until the chondrogenic stages of the bony elements arising from PA2 and the posterior pharyngeal arches.

The *eGFP* expression patterns in the post-migratory CNCCs that were directed by the medaka *hoxa2a* r3/5ER in this study did not fully phenocopy the post-migratory CNCC expression patterns shown by medaka *hoxa2a* whole-mount *in situ* hybridization experiments performed by Davis et al. (2008). For instance, we observed very low levels of *eGFP* expression within PA2 but strong *eGFP* expression in the posterior arches in stable-line transgenic medaka embryos generated with Constructs #1, 2 and 5 (see Fig. 3-4). Further, we observed strong *eGFP* expression in PA2 and the posterior pharyngeal arches in stable-line transgenic embryos generated with the 89-bp r4/CNCC-specifying element (Construct #7) (see Fig. 3-4). However, *eGFP* expression driven by the r4/CNCC-specifying element was restricted to the ventral domain of PA2. By contrast, whole-mount *in situ* hybridization analyses showed that medaka *hoxa2a* is expressed robustly throughout the ventral and dorsal domains of PA2 (Davis et al., 2008). Thus, our reporter gene expression results suggest that the genomic sequences analyzed in this study, as well as other surrounding sequences, are necessary for medaka *hoxa2a* to be expressed and ultimately function in patterning the jaw support elements that are derived from PA2 and the pharyngeal jaw apparatus from the posterior pharyngeal arches. Future functional genomic

studies using the medaka *hoxa2a* r3/5ER and flanking genomic sequences will be required to fully phenocopy the medaka *hoxa2a* expression pattern in the post-migratory CNCCs.

Our functional genomic analysis showed that the 89-bp r4/CNCC-specifying element was responsible for directing reporter gene expression in the CNCCs of the ventral domain of PA2 and the posterior pharyngeal arches. Further, comparative genomic sequence analyses suggest that this expression was driven by Hox/Pbx and Prep/Meis sites located within the 89-bp r4/CNCC-specifying element. Previous functional genomic studies of the mouse *Hoxb3-b2* intergenic sequences have shown that Hox/Pbx and Prep/Meis binding sites contain the capability to direct *Hox* gene expression in the CNCCs (Ferretti et al., 200). Thus, our functional genomic analyses of the medaka r4/CNCC-specifying element and comparative genomic sequence results suggest that there are auto- and/or cross-regulatory interactions of *Hox* genes within the CNCCs of the ventral domain of PA2 and in the posterior pharyngeal arches of medaka and other teleosts. Functional genetic analyses of tilapia *Hox* PG2 genes showed that the independent knockdowns of *hoxa2a*, *a2b* and *b2a* resulted in changed expression levels of *Hox* PG2 genes in PA2 and the posterior pharyngeal arches, which suggested that *Hox* PG2 genes undergo auto- and cross-regulation in the CNCCs of PA2 and the posterior pharyngeal arches (Le Pabic et al., 2010). Furthermore, dual knockdowns of zebrafish *hoxa2b* and *b2a* resulted in readily observable morphant phenotypes in the bony elements arising from the ventral domain of PA2 but not in the dorsal domain, which is suggestive of a higher sensitivity of *Hox* gene activity in the ventral domain of PA2 than in the dorsal (Hunter and Prince, 2002). Coupled with the aforementioned results of previous functional genetic studies in mouse, tilapia and zebrafish, the presence of conserved Hox/Pbx and Prep/Meis binding sites in the r4/CNCC-specifying element suggests that these elements are required for *Hox* gene activity in these pharyngeal arch domains.

Future reporter gene analyses using site-directed mutagenesis of the Hox/Pbx and Prep/Meis binding sites within the r4/CNCC-specifying element of the medaka *hoxa3a-a2a* intergenic region should be performed to determine if these regulatory elements control reporter gene expression in the ventral domain of PA2 and the posterior pharyngeal arches of stable-line transgenic medaka embryos.

The conservation in sequence of the Hox/Pbx and Prep/Meis regulatory elements in the r4/CNCC-specifying element of medaka with orthologous sequences of evolutionarily divergent vertebrates suggests that these elements function similarly across vertebrates in directing *Hoxa2* genes in the CNCCs. In support, medaka *hoxa2a* shows a similar CNCC expression pattern to *Hoxa2* of tetrapods, *hoxa2a* and *a2b* of striped bass and tilapia and *hoxa2b* zebrafish (Gendron-Maguire et al., 1993; Rijli et al., 1993; Grammatopoulus et al., 2000; Pasqualetti et al., 2000; Hunter and Prince, 2002; Baltzinger et al., 2005; Scemama et al., 2006; Le Pabic et al., 2007; Davis et al., 2008). Future functional genomic analyses in evolutionarily divergent vertebrates will help to shed light on the conservation of the r4/CNCC-specifying element. Provided that there are similar CNCC-specific *trans*-acting factors among medaka, tilapia, striped bass and fugu, the medaka may serve as a useful model for understanding the functional nature of the *hoxa2a*-specific genomic sequences among evolutionarily divergent acanthopterygians teleosts. Interestingly, the Hox/Pbx and Prep/Meis sites were not identified by Maconochie et al. (1999) when they examined *cis*-regulatory elements responsible for directing mouse *Hoxa2* expression in the CNCCs. However, Maconochie et al. (1999) did define a NC3 element that was necessary for directing CNCC expression. This element lies between the RE3 and RE2 sequences of the mouse *Hoxa2* r3/5ER and is conserved with the orthologous sequence of the medaka *hoxa2a* r4/CNCC-specifying element (Tümpel et al., 2002) (see Fig. 3-5). Further, the NC3 element

defined in mouse contains a Hox/Pbx element that is conserved with the Hox/Pbx sequence of the medaka *hoxa2a* r4/CNCC-specifying element (Fig. 3-5). Therefore, it is quite possible that the mouse NC3 element functions similarly to the medaka *hoxa2a* r4/CNCC-specifying element.

Beyond the 89 bp r4/CNCC-specifying element that is required for directing *eGFP* expression in the ventral domain of PA2 and the posterior pharyngeal arches, our functional genomic results showed that there are flanking sequence elements that are involved in repressing the activity of the r4/CNCC-specifying element from directing robust reporter gene expression in the post-migratory and chondrogenic CNCCs in PA2. Stable-line transgenic medaka embryos generated with DNA constructs that encompassed the entire medaka *hoxa2a* r3/5ER (-1778 to -1247, Construct #1, Table 3-2) but that excluded sequences upstream of RE4 (-1468 to -1247, Construct #2, Table 3-2) and downstream of RE2 (-1778 to -1303, Construct #5, Table 3-2) all yielded *eGFP* expression in the migratory CNCCs of the hyoid and post-otic streams and strong *eGFP* expression in the post-migratory CNCCs of the posterior arches (Fig. 4A-I). However, none of stable-line transgenic embryos generated using Constructs #1, 2 or 5 showed robust *eGFP* expression in the post-migratory CNCCs in PA2. The repressed reporter gene expression activity in PA2 but not in migratory CNCCs of the hyoid and and post-otic streams or in the post-migratory CNCCs of the posterior pharyngeal arches suggests that there are differential regulatory element interactions mediated by sequences within the 531 bp the *hoxa2a* r3/5ER but outside the 89 bp r4/CNCC-specifying element that direct medaka *hoxa2a* expression in these embryonic CNCC domains. Interestingly, none of the CNCC-specific *cis*-regulatory elements upstream of mouse *Hoxa2* were shown to repress reporter gene expression in PA2 (Maconochie et al., 1999). These discrepant results between medaka *hoxa2a* and mouse *Hoxa2* suggest that the CNCC-directing activities of the *cis*-regulatory elements have diverged between these

evolutionarily divergent osteichthyans and argue that the regulatory systems that mediate expression in the various developmental compartments have shifted their specifying activity by modulating the interplay among elements located within clusters of transcription elements embedded within 'specific' enhancers.

Functional Nature of the Medaka *ψHoxa2b* r3/5 Enhancer Region

Our results from transient and stable-line transgenesis and whole-mount *in situ* hybridization analyses using anti-*eGFP* riboprobes have shown that the r3/5ER upstream of medaka *ψhoxa2b* is functional and is able to direct robust gene expression in the hindbrain and pharyngeal arch embryonic compartments. These results were completely contrary to our expectations based on whole-mount *in situ* hybridization analyses of medaka *ψhoxa2b* which showed that this pseudogene is expressed strongly in noncanonical *Hox* PG2 expression domains but not at all in the characteristic hindbrain and CNCC compartments. These seemingly contradictory results suggest that regions of genomic sequences surrounding the medaka *ψhoxa2b* r3/5 enhancer region function to redirect the activity of the medaka *ψhoxa2b* r3/5 enhancer region to mediate expression in the noncanonical *Hox* PG2 embryonic compartments. Future reporter gene expression analyses that involve microinjection of constructs containing *hoxa9b-ψhoxa2b* intergenic sequences into stable-line transgenic medaka that already contain the entire *ψhoxa2b* r3/5ER may shed light on how surrounding genomic sequences affect the function of the *ψhoxa2b* r3/5ER.

Our functional genomic and whole-mount *in situ* hybridization assays of stable-line transgenic medaka embryos generated with constructs containing the medaka *ψhoxa2b* r3/5ER show that this region functions divergently in directing hindbrain expression from the orthologous r3/5ERs of mouse *Hoxa2* and medaka *hoxa2a*. In mouse the r3/5ER was shown to

direct strong reporter gene expression in r3 and r5 and the deletion or mutagenesis of Krox20 and BoxA sequences from the mouse r3/5ER resulted in the loss of reporter gene expression in these rhombomeres. Further, the deletion of sequences corresponding to RE2, RE3 and RE5 in mouse resulted in the loss of reporter gene expression in r3 and the deletion of RE4 resulted in the loss of r5 expression (Maconochie et al., 2001; Tümpel et al., 2006). By contrast, stable-line transgenic medaka generated with nested deletion constructs of the medaka *ψhoxa2b* r3/5ER in which the sequences orthologous to Krox20, BoxA, RE4 and RE5 were deleted were able to maintain *eGFP* expression in r3-7 of the hindbrain, PA2 and the posterior pharyngeal arches. Interestingly, our analyses of clones representing a series of nested deletions showed that the sequence that is required to direct reporter gene expression in r3-r7 and PA2 and the posterior pharyngeal arches spans from genomic bp positions -803 to -715 and is paralogous to the r4/CNCC-specifying element upstream of medaka *hoxa2a*. Comparative genomic analysis of the medaka *ψhoxa2b* r3-7/CNCC-specifying element revealed the presence of conserved Hox/Pbx, Prep/Meis and LEF-1 binding sites (see Fig. 3-5). However, this sequence was shown to be divergent from the medaka *hoxa2a* r4/CNCC-specifying element by 33 bp (Fig. 3-8). It is possible that these substitutions between the medaka *hoxa2a* and *ψhoxa2b* genomic sequences have allowed the *ψhoxa2b* sequence, but not the *hoxa2a* sequence, to be receptive to transcription factors expressed within r3, r5, r6 and r7 of the hindbrain. Interestingly, our analyses of these sequences using the JASPAR software program showed the presence of several Sox binding elements within the *ψhoxa2b* r3-7/CNCC-specifying element but not in the *hoxa2a* r4/CNCC-specifying element (see Fig. 3-8). Sox proteins have been shown to be involved in driving *Hox* gene expression in several rhombomeres of the hindbrain (Tümpel et al., 2009), so it is possible that these elements are involved in directing reporter gene expression in r3, r5, r6 and

r7. The base pair mutations between the the 88-bp sequence upstream of medaka *ψhoxa2b* and the paralogous r4/CNCC-specifying element may have occurred due to relaxation of selective pressures on the *ψhoxa2b* genomic sequence after the inactivation of the *hoxa2b* gene in the lineage leading to medaka (Davis et al., 2008).

Our reporter gene expression results of the medaka *ψhoxa2b* r3/5ER coupled with previous expression pattern results of medaka *ψhoxa2b* during embryonic development may provide an excellent example of how noncoding sequences contribute to morphological evolution. An increasing body of evidence in the field of evolutionary and developmental biology has shown that mutational changes in conserved noncoding sequences can affect expression patterns of developmentally important genes, and the co-option of such genes in divergent embryonic domains can lead to morphological novelties (see Caroll, 2008). If surrounding sequences interact with the medaka *ψhoxa2b* r3/5ER and cause it to redirect the expression of *ψhoxa2b* in noncanonical *Hox* PG2 domains, this heterotopic shift in expression may have allowed medaka *ψhoxa2b* to be co-opted in developmental compartments other than the characteristic hindbrain and pharyngeal arch compartments. This, however, is assuming that the medaka *ψhoxa2b* transcript gives rise to a functional product that can affect developmental specification. Based on *in silico* translation of the likely mRNA product derived from our *ψhoxa2b* cDNA clones, it is clear that this pseudogene cannot generate a functional Hox protein product (Davis et al., 2008). However, it can generate a truncated protein that possesses a hexapeptide that is conserved with orthologous hexapeptides of *hoxa2b* genes of evolutionarily divergent teleosts (Davis et al., 2008). Since the hexapeptide of *Hox* genes are known to mediate interactions with other transcription factors, especially *Pbx*, there is the tantalizing possibility that the truncated product of *ψhoxa2b* translation could interact with *Pbx* and influence

transcription in developmental compartments in which it is expressed. In the case of medaka which is a representative of the beloniform fishes (an order that includes the flying fishes) these mechanisms may be responsible for generating divergent morphological characters specific to beloniforms. Of particular interest in this regard is the fact that we have documented expression of *ψhoxa2b* in the subterminal region of developing pectoral fin bud that corresponds to the progress zone for fin bud elongation. Given the location of expression of the pseudogene and its potential to interact with *Hox* genes products in the fin bud progress zone, it is possible that the co-option of *ψhoxa2b* expression into this compartment may have influenced the evolution of elongated pectoral fins in the lineage leading to flying fish. In order to test these hypotheses, functional genomic analyses of the *hoxa2b* r3/5ER of other beloniform fishes must be tested within the medaka. Further, phylogenetic analyses must be performed in order to understand if the *hoxa2b* inactivation is specific to the lineage leading to medaka or if it occurred earlier in the beloniform radiation to include other fishes.

Conclusions

In conclusion, our reporter gene analyses showed that the medaka *hoxa2a* and *ψhoxa2b* r3/5ERs are functionally divergent from one another and both are divergent from the r3/5ER of mouse *Hoxa2*. While both r3/5ERs were shown to direct reporter gene expression in the CNCCs, the medaka *hoxa2a* r3/5ER directed expression in r4 of the hindbrain while the medaka *ψhoxa2b* directed expression in r3-r7. These results are different from those observed for the mouse *Hoxa2* r3/5ER, which directed reporter gene expression in r3 and r5. Surprisingly, they are also different from heterologous reporter gene expression results of the medaka *hoxa2a* and *ψhoxa2b* r3/5ERs when they were tested in chicken embryos (Tümpel et al., 2006). In chicken, the medaka *hoxa2a* r3/5ER did not direct expression in the hindbrain or pharyngeal arches while

the *ψhoxa2b* r3/5ER directed reporter gene expression in just r3 and r5 of the hindbrain. Overall, our results underscore the importance of using homologous systems for analyzing *cis*-regulatory elements that direct gene expression during embryonic development.

CHAPTER 4: SPATIO-TEMPORAL PATTERNS OF *HOX* PARALOG GROUP 3-6 EXPRESSION DURING JAPANESE MEDAKA (*Oryzias latipes*) EMBRYONIC DEVELOPMENT

Introduction

Hox genes are a family of evolutionarily-related developmental regulatory genes that serve as critical genetic determinants of regional tissue identities along the anterior–posterior (A-P) axis of animal species (McGinnis and Krumlauf, 1992). They are organized in clusters in the chordate genome and are expressed along the A-P axis during animal embryonic development collinear with their physical location within a cluster (Holland and Garcia-Fernandez, 1996; Ferrier et al., 2000; Powers and Amemiya, 2004). Multiple whole-genome duplications have expanded the total number of *Hox* clusters to 4 in tetrapods, 7–8 in most teleosts and even 13 in salmoniformes (Stellwag, 1999; Amores et al., 2004; Moghadam et al., 2005; Hoegg et al., 2007; Mungpakdee et al., 2008a). Post-genome duplication independent gene loss has resulted in *Hox* gene clusters and paralog groups (PGs) that differ in gene numbers across evolutionarily divergent species depending on the historical timing of gene losses relative to genome duplications (Amores et al., 2004; Le Pabic et al., 2007; Davis et al., 2008).

The bony derivatives arising from the pharyngeal arches (PA) in teleosts aid in feeding and breathing and include the oral jaws arising from PA1, the oral jaw support structures arising from PA2 and the pharyngeal jaw apparatus, which originates from the posterior pharyngeal arches (PA3-7) (Schaeffer and Rosen, 1961; Kimmel et al., 2001). Many of the bony elements, especially those derived from the posterior arches, have been shown to differ in structure across evolutionarily divergent teleosts (e.g., Le Pabic et al., 2009). The difference in structure of pharyngeal jaw components among these teleosts may be related to alteration of nested

expression patterns of *Hox* genes in the posterior pharyngeal arches, which have been shown experimentally to function both cell autonomously and non cell autonomously in patterning postmigratory cranial neural crest cells (CNCCs) into specific cartilaginous structures within the pharyngeal arches (Santagati et al., 2005; Crump et al., 2006; Minoux et al., 2009). Unfortunately, very few studies have examined *Hox* gene expression patterns in the posterior pharyngeal arches of teleosts (Miller et al., 2004; Le Pabic et al., 2007, 2009; Davis et al., 2008), and of these, none were conducted using a comprehensive collection of the *Hox* genes from paralog groups 3 through 6 over a developmental period that extended into the chondrogenic stages of posterior pharyngeal arch development. Until the patterns of anterior expressing *Hox* genes are observed and documented for divergent teleost species, little will be understood regarding how *Hox* genes have contributed to the morphological diversity in the posterior pharyngeal arch derivatives in teleosts.

Here we report the expression patterns of the *Hox* PG3-6 genes in the Japanese medaka (*Oryzias latipes*) within the pharyngeal arches and the neural tube and compare them to the expression patterns of their strict orthologs in other osteichthyans. This study will help to serve as a basis for the expression profile of teleost *Hox* PG3-6 genes in the migrating CNCCs and the pharyngeal arches and neural tube during later developmental stages. We show that each posterior arch of medaka consists of a unique combinatorial code of nested *Hox* gene expression. We analyzed the expression patterns of these genes at developmental stages 22 (nine somites), 29/30 (74–82 hpf) and 34 (121 hpf). At stage 22, the post-otic CNCCs are observed migrating from rhombomere (r) 6 and r7 of the hindbrain to the posterior pharyngeal arches, as shown by the CNCC marker, *dlx2a* (data not shown). At stage 29/30, the undifferentiated cranial neural crest cells of medaka have segregated into six morphologically distinct pharyngeal arches, PA1,

2, 3, 4, 5, 6/7, also shown by the marker *dlx2a* (data not shown). Further, at stage 29/30, the rhombomeres are easily distinguished morphologically without the aid of rhombomere-specific molecular markers. By stage 34, there are seven pharyngeal arches which undergo chondrogenesis of the cranio-facial skeletal elements (Langille and Hall, 1987; Davis et al., 2008). The pharyngeal arches are easily distinguished morphologically at stages 29/30 and 34. This study will serve as an index for comparative expression profiling of teleost *Hox* PG3-6 genes in the migrating CNCCs and the pharyngeal arches and neural tube during later developmental stages.

Materials and Methods

Medaka *Hox* PG3-6 Partial cDNA Cloning

Medaka *Hox* PG3-6 partial cDNAs were generated from RT-PCR using total RNA isolated from stage 29/30 medaka embryos according to the manufacturer's procedure (Totally RNA_, Applied Biosystems, Foster City, CA). The primers used for the amplification of all medaka *Hox* PG3-6 and *dlx2a* partial cDNAs are listed in Table 5-1 and were designed based on published medaka genomic sequences (Accession numbers: AB232918, AB232920, AB232921, AB232922, AB232923, AB232924 and NM001104820) (Kurosawa et al., 2006; Stock et al., 2006). The PCR products generated from RT-PCR-mediated amplification of medaka embryonic total RNA were cloned in pCR II vectors (Invitrogen, Carlsbad, CA), according to the manufacturer's instructions. Confirmation and orientation of PCR products corresponding to inserts from plasmid cDNA clones were determined by restriction endonuclease digestion and DNA sequencing using dideoxyterminator sequencing chemistry (Big Dye v. 3.0, Applied

Table 4-1. Primers used for riboprobe production

Gene	Forward primer (5' to 3')	Reverse primer (5' to 3')	Amplicon Length
hoxa3a	GAGATGGCGGAGGGCTGC	TTTCATCCTGCGGTTCTG	477 bp
hoxb3a	CTCCTCCACAGTCTCCAAGTC	GGGACACCAACTAACATACAG	745 bp
hoxb3b	CAGCAGGCACCGACGCAG	GGCTTCAGAGGAGGGGAG	536 bp
hoxc3a	CTGGATGACCCTTCGGAC	GGGATAATGATGACCACC	466 bp
hoxd3a	GCAACCTATTACGACAACTC	AGTTCTGTCGGGTCTCCTTC	406 bp
hoxa4a	GCGATTACTACGAGCGAC	TTTCCACTTCATCCTCCG	518 bp
hoxb4a	GACTCCCTCTACCACCCCCAC	AGATGTCCAGAGGGGCGGTTC	611 bp
hoxc4a	TATTTGATGGAGTCTAAC	TTTCATCCTGCGGTTCTG	624 bp
hoxd4a	GGGCTCGGACTACTACAG	TTGATTTGCCTCTCGGAC	477 bp
hoxd4b	TACGCAGAGCCGCAGTTC	ACTTTTACTTGCCGTTCG	569 bp
hoxa5a	AGCAAACGAGCAATACAG	TTCTCCAGTTCCAGGGTC	573 bp
hoxb5a	TGTGAACTCGCTGTCGGGGC	ATCTGAGGTGTGTGGGGGTC	636 bp
hoxb5b	TCTTTCGGCTACAACTAC	TTCCCATCAGGTCCAGTC	491 bp
hoxc5a	ATTTACTCCTGTCCCGTC	CTTCATCCGTCTGTTCTG	525 bp
hoxb6a	TCTTATTTTGTGAACCCATC	TTCAGCAAACAGTCCCTTCC	471 bp
hoxb6b	CAGCCTCTGTTTGTAACT	TTTTTTCCGCCTGTTCCT	391 bp
hoxc6a	TTTCGTGCCATTTATCCG	CTCTCCTTCTTCCATTTC	565 bp
dlx2a	CAACCAGATTACCTCAAACAG	AGATGCGTGGTAGAGTTCGTC	732 bp

Biosystems, Foster City, CA).

Whole-Mount *In Situ* Hybridizaton

Cultivation of medaka fish and collection, raising, anesthetizing, fixation and dehydration of medaka embryos were performed as described in Davis et al. (2008). Medaka embryos were developmentally staged according to Iwamatsu (2004). Whole-mount *in situ* hybridization was performed according to Davis et al. (2008). All experiments used digoxigenin (DIG)-labeled sense and antisense riboprobes that were produced and purified according to Scemama et al. (2006). Sense riboprobes were used in control experiments to assess nonspecific binding. Development of DIG-labeled probe signal, examination of embryos and digital photography of embryos was performed as described in Scemama et al. (2006). In comparing medaka *Hox* PG3-6 gene expression patterns to those of other teleosts, morphological features, including the midbrain/hindbrain boundary, rhombomeres (r), otic vesicles (OV), pectoral fins (PF), pharyngeal arches (PA) and somites within developing embryos were used as morphological landmarks.

Results

Medaka *Hox* PG3 Gene Expression Patterns

There are five medaka *Hox* PG3 genes: *hoxa3a*, *b3a*, *b3b*, *c3a* and *d3a* (Kurosawa et al., 2006). This gene complement is divergent from that of zebrafish, which has lost *hoxb3b*, and from *Takifugu rubripes* and *Spheroides nephelus*, which both have lost *hoxc3a* (Amores et al., 1998, 2004). Of all the medaka *Hox* PG3-6 genes analyzed in this study, only medaka *hoxa3a* and *b3a* were observed to be expressed in the post-otic CNCCs migrating from the hindbrain toward the posterior pharyngeal arches at stage 22 (Fig. 4-1A and B and data not shown). At developmental stages 29/30 and 34, medaka *hoxa3a* was observed to be strongly expressed

throughout PA3-7 (Fig. 4-2A and B). Interestingly, *hoxa3a* was expressed exclusively within the mesenchymal region of arches 3-7. At stage 29/30, medaka *hoxa3a* was also observed to be strongly expressed in rhombomeres (r) 5–7, but more weakly in r8 (Fig. 4-2A). By comparison, similar hindbrain expression patterns were observed for *Hoxa3* of mouse, *hoxa3a* of *Takifugu* and Nile tilapia and *hoxa3aa* and *b* of Atlantic salmon (Amores et al., 2004; Mungpakdee et al., 2008b; Le Pabic et al., 2009; Tümpel et al., 2009). A similar pharyngeal arch expression pattern was observed for Nile tilapia *hoxa3a* (Le Pabic et al., 2009).

Medaka *hoxb3a* was observed to be expressed weakly in PA4 but more strongly in PA5-7 at stage 29/30 (Fig. 4-2C). By stage 34, *hoxb3a* expression expanded to include PA3 (Fig. 4-2D). Further, at stage 34 *hoxb3a* was strongly expressed throughout PA3-7. Medaka *hoxb3a* was also observed to be expressed in r4–r8, with the most robust expression occurring in r7, at stage 29/30 (Fig. 4-2C). The medaka *hoxb3a* spatial expression pattern in the hindbrain matched *hoxb3a* of Nile tilapia and *hoxb3aa* and *b* of Atlantic salmon but was divergent from *Hoxb3* of mouse and *hoxb3a* of zebrafish, which had anterior limits of expression in r5, and *Takifugu* in which the anterior limit of expression occurred in r3 (Prince et al., 1998; Amores et al., 2004; Mungpakdee et al., 2008b; Le Pabic et al., 2009; Tümpel et al., 2009). Similar spatial expression patterns in the posterior arches were observed for *hoxb3a* of Nile tilapia and zebrafish (Miller et al., 2004; Le Pabic et al., 2009).

In contrast to the hindbrain and pharyngeal arch expression patterns of *hoxb3a*, medaka *hoxb3b* expression was observed in PA6/7 at stage 29/30 (Fig. 4-2E) but was restricted to PA7 at stage 34 (Fig. 4-2F). In the hindbrain *hoxb3b* was observed solely in r4 at stage 29/30, which

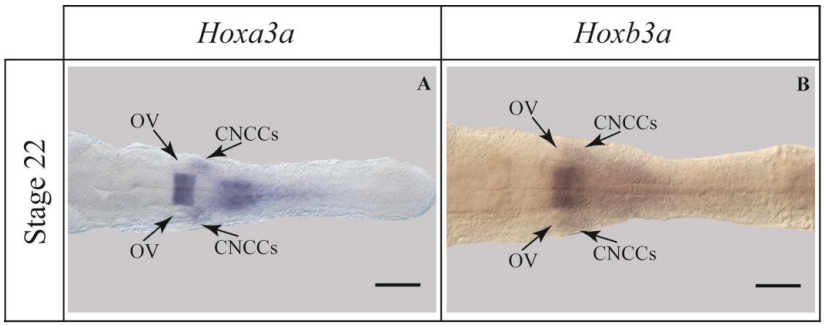


Fig. 4-1. Whole-mount *in situ* hybridization analysis of medaka *hoxa3a* (A) and *hoxb3a* (B) gene expression at stage 22 (9s). All embryos were mounted with their anterior sides to the left and their dorsal sides toward the reader. CNCCs, cranial neural crest cells; OV, otic vesicle. Scale bars equal 0.1 mm.

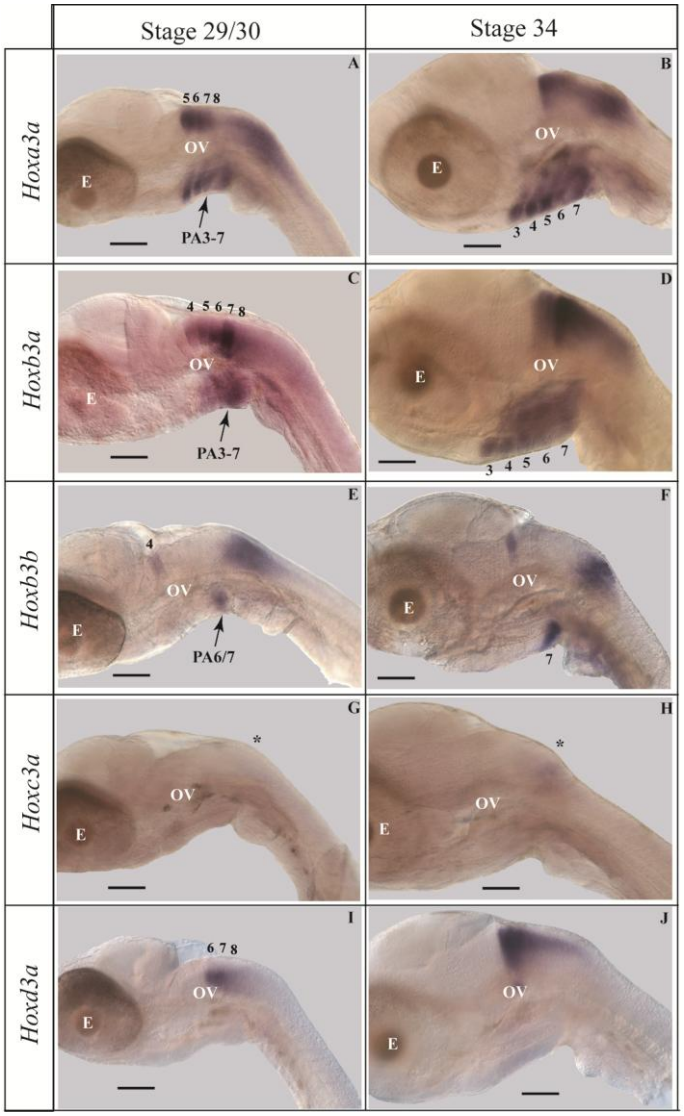


Fig. 4-2. Whole-mount *in situ* hybridization analysis of medaka *hoxa3a* (A and B), *hoxb3a* (C and D), *hoxb3b* (E and F), *hoxc3a* (G and H) and *hoxd3a* (I and J) at stages 29/30 (74–82 hpf) (A, C, E, G and I) and 34 (121 hpf) (B, D, F, H and J). All embryos were mounted with their anterior sides to the left and their lateral sides toward the reader. Rhombomere numbers are indicated by black numbers above the dorsal sides of the embryos. Pharyngeal arch numbers are indicated by black numbers below the ventral sides of the embryos. Asterisk (*) indicates *hoxc3a* expression in neural tube (G and H). E, eye; OV, otic vesicle; PA, pharyngeal arch. Scale bars equal 0.1 mm.

matched the pattern documented for *hoxb3ba* of Atlantic salmon (Mungpakdee et al., 2008b). Interestingly, like *hoxb3a*, *hoxb3b* of medaka was shown to have a more anterior limit of expression than *Hoxb3* in mouse, which was expressed with an anterior limit at the r4/r5 boundary (reviewed by Tümpel et al., 2009).

Medaka *hoxc3a* expression was not observed in the hindbrain or pharyngeal arches, although it was observed to be faintly expressed in the neural tube immediately posterior to the hindbrain (see asterisk (*) in Fig. 4-2G and H). By comparison, Atlantic salmon *hoxc3aa* and *b* were expressed in the neural tube differently from *hoxc3a* of medaka with anterior limits for both genes in r7 of the hindbrain (Mungpakdee et al., 2008b).

Like medaka *hoxc3a*, *hoxd3a* was not observed to be expressed in the pharyngeal arches (Fig. 4-2I and J), although *hoxd3a* showed hindbrain expression within r6–r8 at stage 29/30 (Fig. 4-2I). The hindbrain expression of medaka *hoxd3a* was observed to be similar to *hoxd3a* of Nile tilapia and zebrafish, but divergent from *Hoxd3* of mouse, *hoxd3a* of *Takifugu* and *hoxd3aa* and *b* of Atlantic salmon, which showed an anterior limit of expression at r5 (Prince et al., 1998; Amores et al., 2004; Mungpakdee et al., 2008b; Le Pabic et al., 2009; Tümpel et al., 2009). For the pharyngeal arches, *hoxd3a* of Nile tilapia differed from *hoxd3a* of medaka in that it was observed to be expressed in PA4 and 5 (Le Pabic et al., 2009).

Medaka *Hox* PG4 Gene Expression Patterns

There are five medaka *Hox* PG4 genes: *hoxa4a*, *b4a*, *c4a*, *d4a* and *d4b* (Kurosawa et al., 2006). This gene complement is identical to that of *Takifugu* and *S. nephalus* but divergent from zebrafish, which lacks *hoxd4b* (Amores et al., 2004). All five genes were expressed in the hindbrain and pharyngeal arches at both developmental stages 29/30 and 34. Medaka *hoxa4a* was expressed in PA5-7 at stages 29/30 and 34, with expression being the faintest in PA5 (Fig.

4-3A and B). It was also expressed in the neural tube with an anterior limit at the r7/r8 boundary (Fig. 4-3A and B). Similar *hoxa4a* hindbrain expression patterns were observed for *Takifugu* and Nile Tilapia (Amores et al., 2004; Le Pabic et al., 2009). However, medaka *hoxa4a* expression in the hindbrain was observed to be divergent from *Hoxa4* in mouse, which was expressed with an anterior limit at the r6/r7 boundary (Tümpel et al., 2009). Further, medaka *hoxa4a* showed more anterior expression in the pharyngeal arches than that of Nile tilapia, which was only expressed in PA6 and 7 (Le Pabic et al., 2009).

Medaka *hoxb4a* was expressed in PA5-7 with the strongest expression observed in the unsegmented primordia of PA6/7 at stage 29/30 (Fig. 4-3C). By stage 34 the expression became robust in the dorsal domains of PA5 and 6 and both the dorsal and ventral domains of PA7 (Fig. 4-3D). Medaka hoxb4a was also expressed in the hindbrain with an anterior limit at the r6/r7 boundary (Fig. 4-3C), which was similar to the hindbrain expression patterns of *Hoxb4* in mouse and *hoxb4a* in zebrafish (Prince et al., 1998; Tümpel et al., 2009). Interestingly, the pharyngeal arch expression pattern of medaka *hoxb4a* diverged from zebrafish *hoxb4a*, which was expressed in PA4-7 (Miller et al., 2004).

Medaka *hoxc4a* was expressed faintly in the unsegmented primordial of PA6/7 at stage 29/30 (Fig. 4-3E). By stage 34, *hoxc4a* was observed to be expressed faintly in PA7 (Fig. 4-3F). Medaka *hoxc4a* was expressed more robustly in the hindbrain and posterior neural tube, where it showed an anterior limit of expression at the r6/r7 boundary (Fig. 4-3E). A similar spatio-temporal expression pattern in the pharyngeal arches was observed for tilapia *hoxc4a* (Le Pabic et al., 2009). However, medaka *hoxc4a* showed a divergent hindbrain expression pattern from

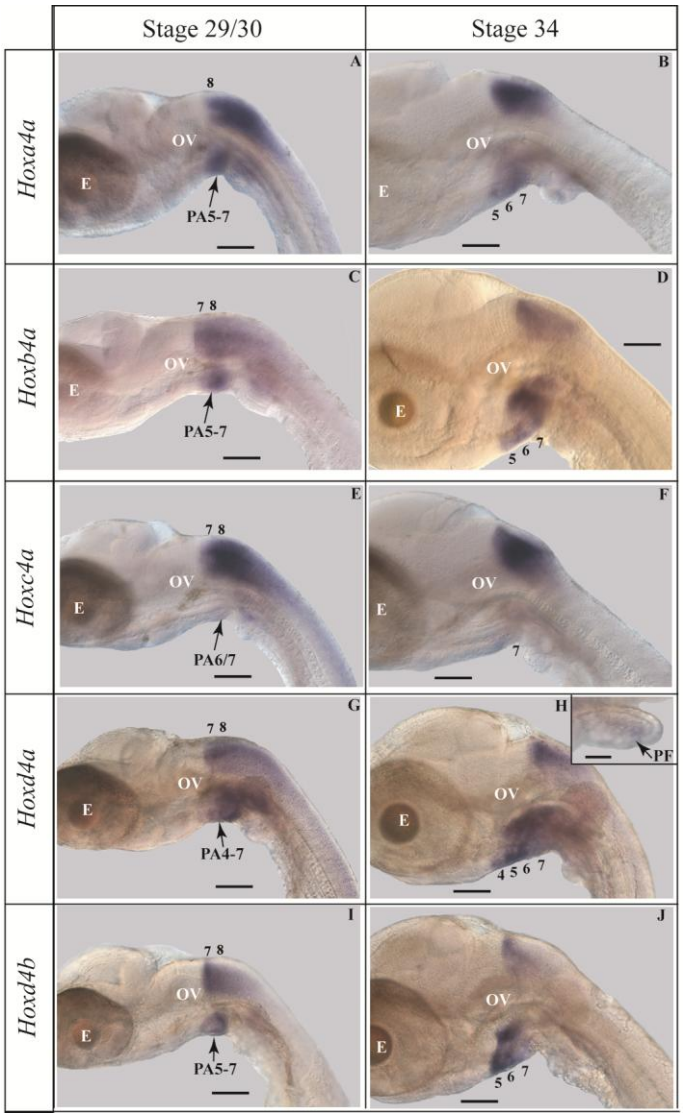


Fig. 4-3. Whole-mount *in situ* hybridization analysis of medaka *hoxa4a* (A and B), *hoxb4a* (C and D), *hoxc4a* (E and F), *hoxd4a* (G and H) and *hoxd4b* (I and J) at stages29/30 (74–82 hpf) (A, C, E, G and I) and 34 (121 hpf) (B, D, F, H and J). All embryos were mounted with their anterior sides to the left and their lateral sides toward the reader. Rhombomere numbers are indicated by black numbers above the dorsal sides of the embryos. Pharyngeal arch numbers are indicated by black numbers below the ventral sides of the embryos. E, eye; OV, otic vesicle; PA, pharyngeal arch; PF, pectoral fin. Scale bars equal 0.1 mm.

that of *hoxc4a* of tilapia and *Takifugu*, which both showed an anterior limit of expression at the r7/r8 boundary, and *hoxc4aa and b* of Atlantic salmon, which were both expressed in the neural tube posterior to the hindbrain (Amores et al., 2004; Mungpakdee et al., 2008b; Le Pabic et al., 2009).

Medaka *hoxd4a* and *d4b* were expressed in similar patterns in the hindbrain and neural tube with anterior limits of expression at the r6/r7 boundary (Fig. 4-3G and I). However, these duplicates showed divergent expression patterns in the pharyngeal arches, such that *hoxd4a* was expressed in PA4-7 with the faintest levels of expression in PA4 at stages 29/30 and 34 (Fig. 4-3G and H) while hoxd4b was expressed in PA5-7 at stages 29/30 and 34 with the faintest levels of expression in PA7 at stage 34 (Fig. 4-3I and J). Similar hindbrain expression patterns to medaka *hoxd4a* were observed for *Hoxd4* of mouse and *hoxd4a* of *Takifugu* and tilapia (Amores et al., 2004; Le Pabic et al., 2009; Tümpel et al., 2009). By contrast, *hoxd4b* of *Takifugu* and tilapia differed from that of medaka, such that both genes showed an anterior limit of expression at the r7/r8 boundary (Amores et al., 2004; Le Pabic et al., 2009). For the pharyngeal arches, a similar expression pattern to medaka *hoxd4a* was observed in tilapia (Le Pabic et al., 2009). By contrast, medaka *hoxd4b* showed a divergent pharyngeal arch expression pattern from tilapia *hoxd4b*, which was observed to be expressed in PA5-7 during early developmental stages but restricted to PA5 at chondrogenesis (Le Pabic et al., 2009). In addition to the divergent expression patterns of medaka *hoxd4a* and *d4b* within the pharyngeal arches, these genes also showed divergence in expression at the level of the pectoral fin buds, such that *hoxd4a*, but not *hoxd4b*, expression was observed in these embryonic domains at stage 34 (Fig. 4-3H inset), which suggests that there are divergent *cis*-regulatory elements between these medaka *Hox* gene duplicates that are involved in directing gene expression in the pectoral fin buds.

Medaka *Hox* PG5 Gene Expression Patterns

The medaka *Hox* PG5 gene complement is composed of four genes, *hoxa5a*, *b5a*, *b5b* and *c5a*, and is identical to the *Hox* PG5 gene complements of *Takifugu, S. nephalus* and zebrafish (Amores et al., 2004; Kurosawa et al., 2006). All *Hox* PG5 genes of medaka were observed to be expressed in the neural tube posterior to the hindbrain (Fig. 4-4A–H). In the pharyngeal arches, medaka *hoxa5a* was expressed strongly in PA6-7 at stages 29/30 and 34 (Fig. 4-4A and B). *Hoxb5a* and *b5b* were both expressed weakly in the unsegmented primordia of PA6/7 at stage 29/30 (Fig. 4-4C and E). By stage 34, medaka *hoxb5a* expression expanded robustly to include PA5-7 (Fig. 4-4D), whereas *hoxb5b* was restricted to PA7 with the most robust expression observed in the dorsal domain of this arch (Fig. 4-4F). Medaka *hoxc5a* was not expressed in the pharyngeal arches (Fig. 4-4G and H). A similar *hoxc5a* expression pattern in the neural tube was observed in tilapia (Le Pabic et al., 2009).

Medaka *Hox* PG6 Gene Expression Patterns

The medaka *Hox* PG6 gene complement is composed of three genes, *hoxb6a*, *b6b* and *c6a*, and is identical to the *Hox* PG6 complements of *Takifugu, S. nephalus* and zebrafish (Amores et al., 2004; Kurosawa et al., 2006). Each of the medaka *Hox* PG6 genes were expressed in the neural tube posterior to the hindbrain (Fig. 4-5A–F). Medaka *hoxb6a* was not expressed in the pharyngeal arches at stage 29/30 (Fig. 4-5A), but weak expression was detected in the ventral domain of PA7 at stage 34 (Fig. 4-5B). By comparison, medaka *hoxb6b* showed weak expression in the unsegmented primordial of PA6/7 at stage 29/30 (Fig. 4-5C) and was restricted to PA7 at stage 34, where it was expressed faintly in the dorsal domain and more

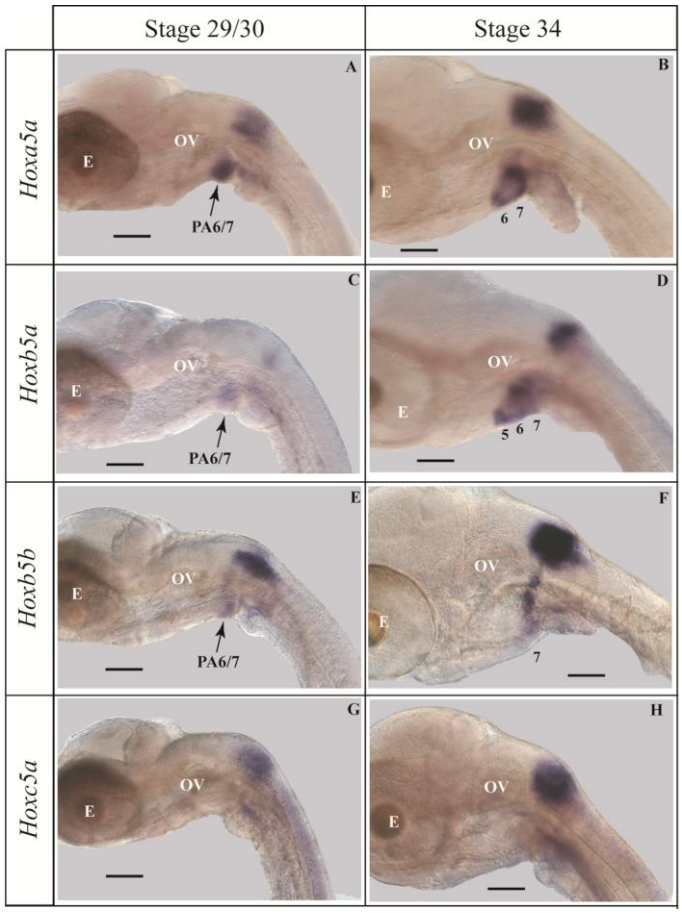


Fig. 4-4. Whole-mount *in situ* hybridization analysis of medaka *hoxa5a* (A and B), *hoxb5a* (C and D), *hoxb5b* (E and F) and *hoxc5a* (G and H) at stages 29/30 (74–82 hpf) (A, C, E and G) and 34 (121 hpf) (B, D, F and H). All embryos were mounted with their anterior sides to the left and their lateral sides toward the reader. Pharyngeal arch numbers are indicated by black numbers below the ventral sides of the embryos. E, eye; OV, otic vesicle; PA, pharyngeal arch. Scale bars equal 0.1 mm.

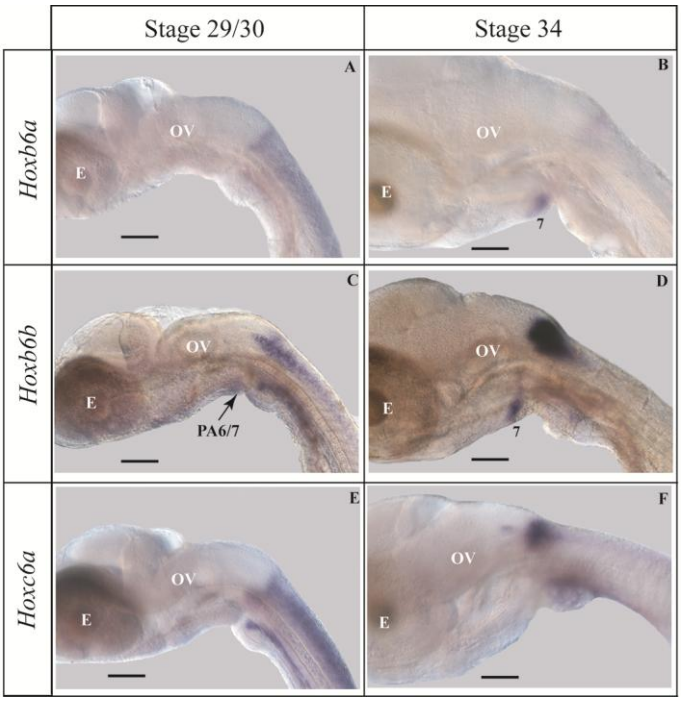


Fig. 4-5. Whole-mount *in situ* hybridization analysis of medaka *hoxb6a* (A and B), *hoxb6b* (C and D) and *hoxc6a* (E and F) at stages 29/30 (74–82 hpf) (A, C and E) and 34 (121 hpf) (B, D and F). All embryos were mounted with their anterior sides to the left and their lateral sides toward the reader. Pharyngeal arch numbers are indicated by black numbers below the ventral sides of the embryos. E, eye; OV, otic vesicle; PA, pharyngeal arch. Scale bars equal 0.1 mm.

robustly in the ventral domain (Fig. 4-5D). *Hoxc6a* was not observed to be expressed in the pharyngeal arches (Fig. 4-5E and F). A similar *hoxc6a* expression pattern in the neural tube was observed in tilapia (Le Pabic et al., 2009).

Discussion

Evolution of Pharyngeal Arch Expression in Medaka

In this paper, we provided evidence for an extensive nested expression pattern for themedaka *Hox* PG3-6 genes within the posterior pharyngeal arches prior to and during chondrogenesis of post-migratory CNCCs into their derivative bony elements. Our results show that the medaka *Hox* genes are expressed with distinct A–P boundaries in the pharyngeal arches, the patterns of which, in many cases, change in position in a developmental stage-dependent manner. Thus, as in tilapia, the use of *Hox* genes as molecular markers of the medaka posterior arches is feasible but stage-dependent (Le Pabic et al., 2009). By combining the results of this study with those reported by us previously for *Hox* PG2 genes (Davis et al., 2008), we provide a combinatorial code of expressed medaka *Hox* genes that specifies the identity of each of the posterior pharyngeal arches (3–7) during the chondrogenesis of the post-migratory CNCCs into their bony derivatives (stage 34) (Fig. 4-6). The characteristics of this code are that at stage 34, PA3 expresses *hoxa2a*, *b2a*, *a3a* and *b3a*. PA4 expresses *hoxa2a*, *b2a*, *a3a*, *b3a* and *d4a*. PA5 expresses *hoxa2a*, *b2a*, *a3a*, *b3a*, *a4a*, *b4a*, *d4a*, *d4b* and *b5a*. PA6 expresses *hoxa2a*, *b2a*, *a3a*, *b3a*, *a4a*, *b4a*, *d4a*, *d4b*, *a5a* and *b5a*. PA7 expresses *hoxa2a*, *b2a*, *a3a*, *b3a*, *b3b*, *a4a*, *b4a*, *c4a*, *d4a*, *d4b*, *a5a*, *b5a*, *b5b*, *b6a* and *b6b*, and is the only arch to express *hoxb3b*, *c4a*, *b5b*, *b6a* and *b6b*. Provided that the *Hox* PG2-6 gene expression patterns in the posterior pharyngeal arches

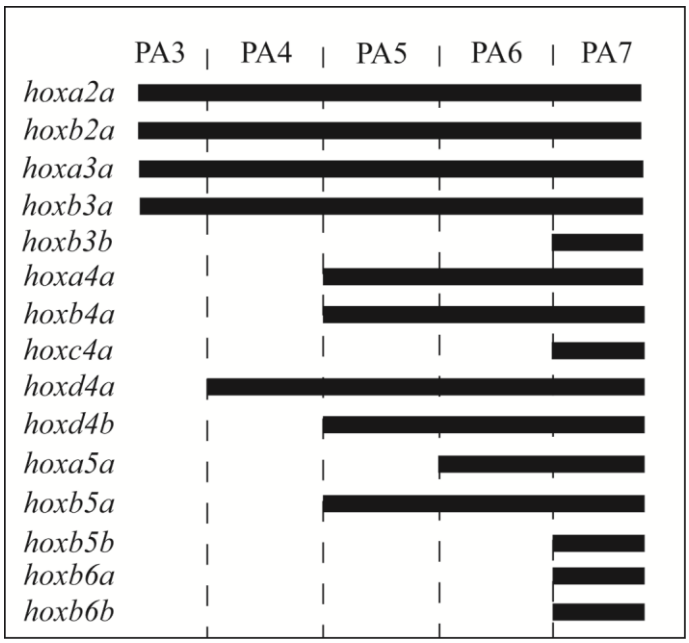


Fig. 4-6. Combinatorial code of *Hox* PG2-6 gene expression in the posterior pharyngeal arches of medaka at stage 34 (121 hpf).

exhibited by medaka are similar to those of other closely related beloniform fishes but not to more evolutionarily divergent teleosts, it would be tantalizing to suggest that the expression of the medaka *Hox* PG2-6 genes in the posterior pharyngeal arches may have, in part, resulted in the autopomorphic pharyngeal jaw apparatus structures characteristic of beloniform fishes, including the reduction in size of the second and third epibranchials (arising from the dorsal regions of the fourth and fifth PAs), the great expansion of the articular surface of the fourth epibranchial (arising from the dorsal region of PA6) and the presence of large ventral flanges on the fifth ceratohyal (arising from PA7) (Rosen and Parenti, 1981; Parenti, 1987). Interestingly, several *Hox* genes in tilapia, but not medaka, were observed to be expressed in the pharyngeal arches including *hoxa2b* which was shown to be expressed in PA2 and the posterior arches and *hoxd3a* which was shown to be expressed in PA4 and 5 (Le Pabic et al., 2007, 2009). However, it must be stressed that a more complete investigation of pharyngeal arch *Hox* expression in tilapia and other evolutionarily divergent teleosts is necessary to better predict the role played by *Hox* genes in shaping divergent morphologies derived from the posterior pharyngeal arches.

The stage-dependent medaka *Hox* gene expression patterns reported in this study are suggestive of dynamic *cis*-regulatory circuitry that direct *Hox* genes in the migratory and post-migratory CNCCs. Taking into account the whole-mount *in situ* hybridization results in Davis et al. (2008) we only observed the expression of four medaka *Hox* genes, *hoxa2a*, *b2a*, *a3a* and *b3a*, in the post-otic CNCC stream and throughout all of the posterior pharyngeal arches, which suggests that their expression in these embryonic domains may be caused by *cis*-regulatory sequences that are similar to those that direct mouse *Hoxa2* expression in the post-otic CNCC stream and the pharyngeal arches (Maconochie et al., 1999). By contrast, all of the other medaka *Hox* PG3-6 genes analyzed in this study were only observed to be expressed in post-migratory

CNCCs at variable A-P boundaries in the posterior pharyngeal arches (see Fig. 4-6), which suggests that post-migratory CNCC *Hox* gene expression may be regulated differentially between individual arches. By comparison, several *cis*-regulatory sequences have been shown to be associated with directing or repressing *Hox* gene expression in specific rhombomeres of the hindbrain, such as Krox20-binding sequences which direct the expression of *Hoxa2* and *b2* in r3 and r5 but repress *Hoxb1* from being expressed in these rhombomeres and Hox/Pbx- and Prep/Meis-binding sequences which have been shown to direct the expression of *Hoxb1*, *a2* and *b2* in r4 (reviewed by Tümpel et al., 2009). Given that individual pharyngeal arches and rhombomeres give rise to distinct morphological derivatives, it is likely that the medaka *Hox* PG3-6 gene expression patterns in the post-migratory CNCCs in the posterior phraryngeal arches are patterned by distinct *cis*-regulatory modules in a similar fashion to the rhombomeres of the hindbrain.

Evolution of Hindbrain Expression in Teleosts

In addition to the medaka *Hox* PG3-6 expression patterns in the pharyngeal arches, we observed the spatial expression patterns of these genes in the neural tube at stage 29/30. Taking into account the medaka *hoxb1a* and *b1b* expression patterns reported by Hurley et al. (2007) and the *Hox* PG2 expression patterns reported by Davis et al. (2008), we present the spatial expression patterns of medaka *Hox* PG1-4 in the hindbrain during later developmental stages (stage 29/30 and on) (Fig. 4-7). r2 expresses *hoxa2a* alone. r3 expresses *hoxa2a* and *b2a*. r4 expresses *hoxb1a*, *b1b*, *a2a*, *b2a*, *b3a* and *b3b*. r5 expresses *hoxa2a*, *b2a*, *a3a* and *b3a*. r6 expresses *hoxa2a*, *b2a*, *a3a*, *b3a* and *d3a*. r7 expresses *hoxa2a*, *b2a*, *a3a*, *b3a*, *d3a*, *b4a*, *c4a*,

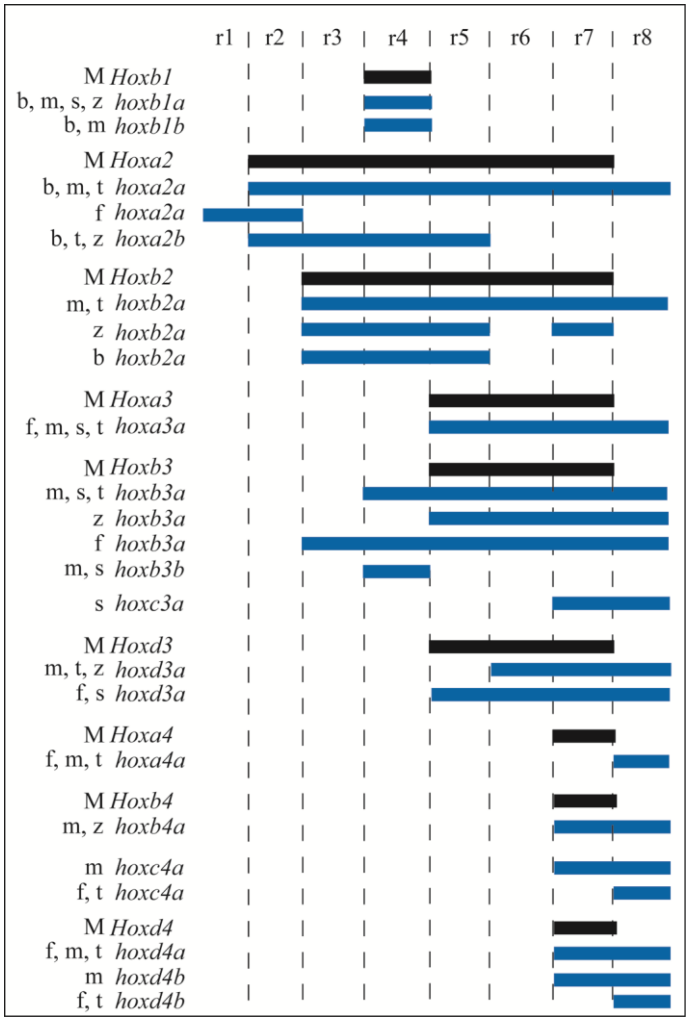


Fig. 4-7. Comparison of mouse and teleost *Hox* PG1-4 gene expression in the rhombomeres of the developing hindbrain. Black bars correspond to mouse *Hox* gene expression. Blue bars correspond to teleost *Hox* gene expression. b, striped bass; f, *Takifugu*; m, medaka; M, mouse; s, salmon; t, tilapia; z, zebrafish.

d4a and *d4b*. r8 expresses *hoxa2a*, *b2a*, *a3a*, *b3a*, *d3a*, *a4a*, *b4a*, *c4a*, *d4a* and *d4b*, which is indicative that the more posterior rhombomeres express ever increasing numbers of *Hox* paralog 2–6 genes. *Hoxc3a* and all of the *Hox* PG5-6 genes of medaka are expressed in the neural tube with anterior limits caudally to the hindbrain, results that are consistent with the expression patterns documented for *Hox* PG5-6 genes of tetrapods (see Tümpel et al., 2009). Interestingly, many of the medaka *Hox* PG2-4 genes exhibit divergent spatial expression patterns in the hindbrain from their strict orthologs in several evolutionarily divergent teleosts, including zebrafish (*hoxb2a* and *b3a*), striped bass (*hoxb2a*), tilapia (*hoxc4a* and *d4b*), *Takifugu* (*hoxa2a*, *b3a*, *d3a*, *c4a* and *d4b*) and salmon (*hoxc3a* and *d3a*) (Prince et al., 1998; Scemama et al., 2002, 2006; Amores et al., 2004; Le Pabic et al., 2007; Davis et al., 2008; Mungpakdee et al., 2009a, b) (Fig. 4-7). Further, in comparison to the *Hox* PG1-4 spatial expression patterns in the hindbrain of tetrapods (summarized in Tümpel et al., 2009), several orthologous *Hox* genes of teleosts exhibit divergent anterior limits of expression in the hindbrain, including *hoxa2a* (*Takifugu*), *b3a* (medaka, tilapia and salmon), *b3b* (medaka and salmon), *d3a* (medaka, zebrafish and tilapia), *a4a* (medaka, tilapia and *Takifugu*) and *d4b* (tilapia and *Takifugu*) (Fig. 4-7). The *cis*-regulatory circuitry responsible for directing tetrapod *Hox* PG1-4 gene expression in the hindbrain is summarized in Tümpel et al. (2009). Based on the divergent anterior limits of expression of *Hox* PG2-4 genes in the hindbrain exhibited by several teleosts in comparison to mouse, it can be deduced that several of the *cis*-regulatory modules that direct *Hox* genes in specific rhombomeres of the hindbrain have diverged between tetrapods and teleosts. Given that the hindbrain expression patterns of *Hox* PG2-4 genes appear to be largely conserved in tetrapods but highly divergent in teleosts, future functional genomic studies in teleosts are necessary in

order to understand how the ray-finned fish-specific whole-genome duplication resulted in divergent *cis*-regulatory activity in *Hox* genes of teleosts.

CHAPTER 5: CONCLUDING REMARKS AND FUTURE DIRECTIONS

In this thesis, I investigated the gene complement, expression and regulation of *Hox* paralog group 2 (PG2) genes in the Japanese medaka. I found that the *Hox* PG2 gene complement of medaka is divergent from other osteichthyans as it is composed solely of *hoxa2a* and *b2a* (see Chapter 2). Further, these genes were shown to be expressed in conserved patterns in the developing hindbrain and pharyngeal arches relative to orthologous *Hoxa2* and *b2* genes of evolutionarily divergent osteichthyans. Interestingly, I found that medaka *hoxa2b* has deteriorated to a pseudogene, *ψhoxa2b*, as a result of coding sequence mutations that generated premature termination codons in exon 2 upstream of the homeodomain encoding region. The absence of a homeodomain in translation products from the *ψhoxa2b* pseudogene argues that the truncated product would not function as a classical *Hox* gene despite its retention of a functional hexapeptide sequence. It is interesting that the transcript derived from the *in vivo* expression of the *ψhoxa2b* pseudogene was expressed in noncanonical *Hox* PG2 expression domains, including the caudal-most region of the embryonic trunk, the ventral-most aspect of the neural tube and the distal mesenchyme of the pectoral fin buds, which if translated could have a role in interactions with proteins like Pbx, that bind the expressed hexapeptide.

Reporter gene analyses of the r3/5 enhancer regions (r3/5ER) upstream of medaka *hoxa2a* and *ψhoxa2b* in their homologuous host showed that these genomic DNA regions have diverged in function from each other as well as the r3/5ER upstream of mouse *Hoxa2*. The mouse *Hoxa2* r3/5ER has been shown to direct *Hoxa2* expression in rhombomere (r) 3 and r5 of the hindbrain and the migratory cranial neural crest cells (CNCCs) delaminating from the hindbrain to enter the 2nd (PA2) and the posterior pharyngeal arches. Although the orthologous genomic region upstream of medaka *hoxa2a* was shown to direct reporter gene expression in the

migratory and post-migratory CNCCs entering and populating PA2 and the posterior pharyngeal arches, it directed expression in r4 of the hindbrain but not in r3 or r5. Further, transgenic medaka fish generated with nested deletion constructs of the medaka *hoxa2a* r3/5ER showed that the r4 and CNCC driven reporter gene expression required the presence of, at most, an 89 bp fragment of genomic DNA that encompassed the regulatory elements orthologous to the RE2 and RE3 sequences of the mouse *Hoxa2* r3/5ER. I termed this DNA sequence fragment as the medaka *hoxa2a* r4/CNCC-specifying element. A comparative genomic analysis of this 89 bp fragment with orthologous genomic sequences upstream of *Hoxa2* genes of several evolutionarily divergent osteichthyans, including shark, *Latimeria*, mouse, human, chicken, bichir and several teleosts, revealed the presence of conserved Hox/Pbx and Prep/Meis binding sites for all sequences analyzed (see Chapter 3). Hox/Pbx and Prep/Meis binding sites have been shown to be integral in directing *Hox* gene expression in r4 of the hindbrain (Ferretti et al., 2000; Tümpel et al., 2006 and 2007) and in the CNCCs (Ferretti et al., 2000). Interestingly, the orthologous region of DNA of the mouse *Hoxa2* r3/5ER did not drive reporter gene expression in r4 in transgenic mouse embryos (Maconochie et al., 2001), which suggests divergence in function of the r3/5ERs between mouse *Hoxa2* and medaka *hoxa2a*. The r3/5ER upstream of medaka *ψhoxa2b* was shown to direct reporter gene expression in r3-7 of the hindbrain and the migratory and post-migratory CNCCs entering and populating PA2 and the posterior pharyngeal arches despite the fact that the transcript from this pseudogene is expressed in noncanonical domains. Further, transgenic medaka fish generated with nested deletion constructs of the medaka *ψhoxa2b* r3/5ER showed that the hindbrain and CNCC driven reporter gene expression required the presence of, at most, an 88 bp fragment of genomic DNA that is paralogous to the medaka *hoxa2a* r4/CNCC-specifying element (see Chapter 3). This DNA fragment was

therefore termed the medaka *ψhoxa2b* r3-7/CNCC-specifying element. These results suggest an evolutionary divergence in function of the r3/5ERs of mouse *Hoxa2*, medaka *hoxa2a* and medaka *ψhoxa2b* and question the ancestral function of the *Hoxa2* r3/5ER prior to the evolutionary divergence of ray-finned fishes and lobe-finned fishes. A functional genomic analysis of the orthologous genomic DNA upstream of shark *Hoxa2* may help to shed light on the ancestral nature of the *Hoxa2* r3/5ER. Likewise, further functional genomic analyses of r3/5ERs in other teleosts must be performed in order to determine if the functional nature of the medaka *hoxa2a* and *ψhoxa2b* r3/5ERs are representative of all teleost *hoxa2a* and *hoxa2b* genes.

In addition to the expression and regulation of medaka *hoxa2a* and *ψhoxa2b*, I investigated the nested expression patterns of medaka *Hox* PG3-6 genes in the developing posterior pharyngeal arches. This study is the first of its kind and provides a basis for comparative studies of posterior arch-expressing *Hox* genes in evolutionarily divergent teleosts. The bony structures of the pharyngeal jaw apparatus have been shown to be variable in morphology across evolutionarily divergent teleosts (e.g.: Langille and Hall, 1987; Parenti, 1987; Le Pabic et al., 2009). *Hox* genes occupy a significant role in patterning the bony architecture arising from the pharyngeal arches (Gendron-Maguire et al., 1993; Rijli et al., 1993; Grammatopoulus et al., 2000; Hunter and Prince, 2002; Baltzinger et al., 2005; Crump et al., 2006; Minoux et al., 2009; Le Pabic et al., 2010). The nested expression patterns of *Hox* genes along the anterior-posterior axis of developing embryos has given rise to the *Hox* code hypothesis, which states that a given combination of *Hox* gene products is necessary for specifying segmental identities along the A-P axis (Krumlauf, 1994). Given that medaka and other beloniform fishes possess autopomorphic bony characteristics arising from the posterior arches, it is possible that the nested *Hox* expression patterns presented in Chapter 4 represent a

Hox code that is specific to beloniform fishes. The documentation of *Hox* PG3-6 gene expression patterns in other teleosts that possess divergent posterior pharyngeal arch-derived bony characters from medaka must be performed to increase our understanding of how *Hox* codes function among evolutionarily divergent teleosts.

Contribution of My Thesis

My dissertation research provides several contributions to the field-of evolutionary and developmental biology. First, it adds *Hox* PG2 gene expression data from an acanthopterygian teleost with a unique *Hox* PG2 gene complement to a list of characterized *Hox* PG2 gene complements and expression patterns of evolutionarily divergent vertebrates. Variation in *Hox* PG2 gene complements in evolutionarily divergent vertebrates may have, in part, resulted in divergent gene expression patterns in the pharyngeal arches. Divergence in expression is mirrored by variable *Hox* PG2 gene function in patterning the bony structures arising from the pharyngeal arches. In chapter 2, I showed that medaka has just two functional *Hox* PG2 genes that are expressed in PA2, *hoxa2a* and *b2a*. Interestingly, these expression patterns are similar to *hoxa2b* and *b2a* of zebrafish (Hunter and Prince, 2002). In zebrafish, *hoxa2b* and *b2a* function redundantly in patterning PA2 (Hunter and Prince, 2002) and the similar *Hox* PG2 gene expression patterns between medaka and zebrafish is suggestive of redundancy in function of medaka *Hox* PG2 genes in specifying PA2. Functional redundancy of the medaka *Hox* PG2 genes in specifying PA2 would be divergent from tetrapods, wherein *Hoxa2* alone patterns the bony arch morphologies of PA2 and tilapia, where *hoxa2a*, *a2b* and *b2a* appear to function in a cross-regulatory manner to specify the identity of PA2 (Gendron-Maguire et al., 1993; Rijli et al., 1993; Grammatopoulus et al., 2000; Hunter and Prince, 2002; Baltzinger et al., 2005; Le Pabic et al., 2010). Functional analyses of the medaka *Hox* PG2 genes may shed light on their

activity of patterning the unique bony morphology arising from PA2. Medaka, like other beloniform fishes, does not possess an interhyal, a small cartilaginous structure that facilitates buccopharyngeal expansion in other evolutionarily divergent teleosts, like zebrafish and tilapia (Schaeffer and Rosen, 1961). Second, this research provides genetic evidence of a gene inactivation event with medaka *ψhoxa2b* (shown in Chapter 2). Mutations of *cis*-regulatory elements specific to medaka *ψhoxa2b* may have resulted in its aberrant expression patterns during embryonic development. Medaka *ψhoxa2b* is expressed in noncanonical *Hox* PG2 domains, specifically the caudal-most region of the embryonic trunk, the ventral-most aspect of the neural tube and the distal mesenchyme of the pectoral fin buds. Third, this thesis research shows the divergence in function of paralogous genomic sequences specific to duplicated genes. In Chapter 3, I showed that the r3/5ER of medaka *hoxa2a* drives expression in r4 while the orthologous region upstream of *ψhoxa2b* drives expression in r3-r7. Interestingly, both of these enhancer regions were also shown to have diverged in function from orthologous genomic sequences upstream of mouse *Hoxa2*, which direct *Hoxa2* expression in r3 and r5 of the hindbrain (Maconochie et al., 2001). My functional and comparative genomic analyses in Chapter 3 show that our understanding of the functional nature of orthologous enhancer regions in evolutionarily divergent lineages is limited and should be explored further (discussed in more detail below). Fourth, this research shows the importance of using homologous systems for performing reporter gene assays. Data from this thesis using transient and stable-line homologous reporter gene assays to study the r3/5ERs of medaka *hoxa2a* and *ψhoxa2b* does not agree with data using the same genomic sequences when they were tested in chick embryos in a transient transgenic electroporation assay (Tümpel et al., 2006) (see Chapter 3). This suggests that there is either an evolutionary divergence of the *cis*-regulatory elements of the r3/5ER

between teleosts and chicken, divergence in the *trans*-acting factors that bind to these sequences or a combination of both. Lastly, this research provides the first comprehensive examination of nested *Hox* gene expression patterns in the posterior pharyngeal arches of a teleost fish (Chapter 4). There is much evolutionary divergence in the bony structures of the pharyngeal jaw apparatus that arises from the posterior pharyngeal arches in teleosts. *Hox* genes have been shown to be major contributors in patterning the morphology of the bony structures arising from the pharyngeal arches. Divergent *Hox* codes, or specific combinations of *Hox* gene products that pattern unique identities to segments along the developing anterior-posterior axis (Krumlauf, 1994), may be responsible for patterning the morphological variation of the pharyngeal arch-derived bony structures among evolutionarily divergent teleosts.

Hox PG2 Gene-Dependent Pharyngeal Arch Specification in Medaka

In Chapter 2 of my thesis, I described the expression patterns of medaka *hoxa2a* and *b2a*. These *Hox* PG2 genes were shown to be expressed in PA2 up until the chondrogenic stage of the developing bony PA2-derivatives. Functional genetic studies of *Hox* PG2 genes in evolutionarily divergent osteichthyans have shown that when these genes are expressed until late in development they serve as selector genes in patterning the identity of PA2 (Gendron-Maguire et al., 1993; Rijli et al., 1993; Grammatopoulus et al., 2000; Hunter and Prince, 2002; Baltzinger et al., 2005; Le Pabic et al., 2010). The expression patterns of medaka *hoxa2a* and *b2a* are remarkably similar to *hoxa2b* and *b2a* of zebrafish with respect to PA2 development, and this suggests that medaka *hoxa2a* and *b2a* function redundantly in patterning PA2. If this is the case, then it suggests that the ray-finned fish-specific whole genome duplication event and post-genome duplication independent gene loss has channeled *Hox* PG2 gene function differently between teleosts that have retained three genes and those that have retained just two. Evidence

for differential gene function partitioning among divergent *Hox* PG2 gene complements is shown from functional *Hox* PG2 studies in zebrafish and tilapia. While zebrafish *hoxa2b* and *b2a* function redundantly as selectors of PA2 identity, knockdown studies of each of the three genes within tilapia, *hoxa2a*, *a2b* and *b2a*, show that each gene has independent selector gene function in PA2 (Hunter and Prince, 2002; Le Pabic et al., 2010). Further, evidence of cross-regulatory activity was observed between the three *Hox* PG2 genes of tilapia within the developing pharyngeal arches but not between *hoxa2b* and *b2a* of zebrafish (Hunter and Prince, 2002; Le Pabic et al., 2010). Interestingly, recent evidence from a PCR survey of the *Hox* clusters in the goldfish (*Carassius auratus auratus*) has shown that this lineage has retained *hoxa2a*. Since both goldfish and zebrafish are members of the Cypriniformes, the absence of *hoxa2a* in zebrafish (Amores et al., 1998), but not in the goldfish (Luo et al., 2007) and in members of the superorder Acanthopterygii, suggests that the loss of *hoxa2a* occurred in the lineage leading to zebrafish after the divergence of the zebrafish and goldfish lineages. If the expression patterns of the goldfish *Hox* PG2 genes are similar to those of other three *Hox* PG2 gene retaining teleosts, such as tilapia and striped bass, then this would suggest that the three goldfish *Hox* PG2 genes function similarly to those of tilapia. Future functional genetic studies of teleosts containing two and three *Hox* PG2 genes must be performed to understand how independent *Hox* gene loss channels the functional nature in patterning the identity of PA2.

Medaka *Hox* PG2 Gene Complement

In Chapter 2 of my thesis, I showed that medaka *hoxa2a* and *b2a* show expression patterns in the hindbrain and pharyngeal arches that are conserved relative to other *Hoxa2* and *b2* genes in evolutionarily divergent vertebrates. I also showed in Chapter 2 that medaka *hoxa2b* is a transcribed pseudogene, *ψhoxa2b*, that is expressed in noncanonical *Hox* PG2 domains, namely

the caudal-most region of the embryonic trunk, the ventral-most aspect of the neural tube and the distal mesenchyme of the pectoral fin buds. Future analyses using PCR surveys must be performed to determine the phylogenetic timing of the loss of *hoxa2b*. Such analyses would shed light on whether this gene loss event was specific to the lineage leading to medaka or whether it occurred more globally in the Beloniformes and other closely related sister-orders within the Acanthopterygii. Of particular interest for the timing of the loss of this gene lies in the flying fishes, which are beloniform fishes that are closely related to medaka (Lovejoy et al., 2004). Did the *hoxa2b* gene inactivation event occur prior to the split of the lineages leading to medaka and flying fishes or did it occur after this split in the lineage leading to medaka? Provided that the *hoxa2b* orthologs between medaka and flying fish share similar expression domains in the pectoral fins, it would be interesting to determine if *hoxa2b* of flying fish is involved in patterning the elongated pectoral fins. An increasing body of evidence in the field of evolutionary and developmental biology has shown that mutational changes in conserved noncoding sequences can affect expression patterns of developmentally important genes, and the co-option of such genes in divergent embryonic domains can lead to morphological novelties (see Caroll, 2008). If surrounding sequences interact with the medaka *ψhoxa2b* r3/5ER and cause it to redirect the expression of *ψhoxa2b* in noncanonical *Hox* PG2 domains, this heterotopic shift in expression may have allowed medaka *ψhoxa2b* to be co-opted in developmental compartments other than the characteristic hindbrain and pharyngeal arch compartments

While medaka *ψhoxa2b* does not contain a functional homeodomain, it does contain a conserved hexapeptide motif in exon 1. Thus, it would be interesting to determine if this transcript yields a functional product that contributes to the patterning of the embryonic

compartments in which it is expressed. A translatable medaka *ψhoxa2b* product with a functional hexapeptide motif could potentially compete with other homeodomain-containing proteins for co-activator proteins, such as *Pbx*, which aid in increasing *Hox* gene product binding specificity to target sequences of downstream effector genes. Future analyses using oligonucleotide morpholino-mediated knockdowns of medaka *ψhoxa2b* should be considered to test whether this transcript yields a functional product. Also, since the medaka *ψhoxa2b* r3/5ER has been shown to possess the capability of directing reporter gene expression in the hindbrain and pharyngeal arches (shown in Chapter 3), directed misexpression of this pseudogene in the hindbrain and pharyngeal arches using the medaka *ψhoxa2b* r3/5ER and the *Tol2* transposon system may aid our understanding of whether this transcript yields a functional product.

Medaka *Hoxa2a* and *ψHoxa2b* r3/5ER Function

In Chapter 3 of this thesis, I showed that the r3/5ERs of medaka *hoxa2a* and *ψhoxa2b* have diverged in function both from each other and from the r3/5ER of mouse. While the mouse r3/5ER has been shown to drive reporter gene expression in r3, r5 and the pharyngeal arches, I showed that the medaka *hoxa2a* r3/5ER directs reporter gene expression in r4 and the pharyngeal arches while the medaka *ψhoxa2b* r3/5ER directs reporter gene expression in r3-7 of the hindbrain and in the pharyngeal arches. Further, I showed that the medaka *hoxa2a* r4/CNCC and *ψhoxa2b* r3-7/CNCC-specifying elements are orthologous to each other and both are orthologous to a sequence embedded in the mouse *Hoxa2* r3/5ER. This region corresponds to a sequence containing the mouse RE3 and RE2 sequence elements (see Fig. 3-5). Comparative genomic sequence analyses showed that the *hoxa2a* r4/CNCC and the *ψhoxa2b* r3-7/CNCC-specifying elements are conserved between genomic sequences upstream of *Hoxa2* in vertebrates, including shark, *Latimeria*, tetrapods, bichir and teleosts. These conserved regions of sequence were

further shown to contain Hox/Pbx and Prep/Meis binding sites, which have been shown to function in directing *Hox* expression in r4 and the CNCCs (Ferretti et al., 2000; Tümpel et al., 2006 and 2007).

Based on functional results from reporter gene assays of the mouse *Hoxa2* and medaka *hoxa2a* and *ψhoxa2b* r3/5ERs it is reasonable to conclude that there has been evolutionary divergence in these genomic sequences and systems responsible for controlling expression of the cognate coding sequences between the medaka and mouse lineages. It is interesting that the functional genomic analyses of the mouse *Hoxa2* r3/5ER performed by Maconochie et al. (2001) did not reveal any sequences involved in driving reporter gene expression in r4, especially since the mouse *Hoxa2* r3/5ER contains an orthologous sequence that is conserved with the r4/CNCC and r3-7/CNCC-specifying elements of medaka *hoxa2a* and *ψhoxa2b*, respectively (see Fig. 3-5). Further, genomic sequences upstream of mouse *Hoxa2* and medaka *hoxa2a* share conserved Hox/Pbx and Prep/Meis sites. Since homologous reporter gene studies of the mouse *Hoxa2* r3/5ER did not show expression in r4, it is possible that the mouse has evolved sequence modulators that restrict the Hox/Pbx and Prep/Meis sites within the r3/5ER from directing expression in r4. Interestingly, the chicken *Hoxa2* r3/5ER also contains Hox/Pbx and Prep/Meis binding sites that are conserved with medaka *hoxa2a*. However, homologous reporter gene studies of the chicken *Hoxa2* r3/5ER within chick embryos showed that this genomic region only potentiates expression in r3 and r5 (Tümpel et al., 2002), which suggests that the chicken *Hoxa2* r3/5ER possesses similar mechanisms to mouse in repressing its function in directing expression in r4. An alternative interpretation is that there is a molecular mechanism in medaka and/or other teleosts that redirect the r3/5 elements, such as RE3 and RE2, to specify *hoxa2a* expression in r4. Functional genomic analyses of r3/5ERs in species lineages that predate the evolutionary

split of ray-finned from lobe-finned fishes must be performed in order to determine the ancestral function of the vertebrate *Hoxa2* r3/5ER. Since the shark *Hoxa2* r3/5ER has been shown to be conserved with the *Hoxa2* r3/5ERs of vertebrates (Tümpel et al., 2002) and possesses conserved Hox/Pbx and Prep/Meis sites that are orthologous to the binding sites of the medaka *hoxa2a* r4/CNCC-specifying element, this would be an excellent model to use to test the ancestral function of the vertebrate *Hoxa2* r3/5ER.

Our functional genomic results and comparative genomic analyses of the medaka *hoxa2a* and *ψhoxa2b* r3/5ERs show that our knowledge of how genetic regulatory networks function to pattern morphological features is limited. Our contemporary thinking is that kernels, or conserved subcircuits of regulatory genes that interact with each other to pattern specific embryonic compartments (Davidson, 2006), are conserved across evolutionarily divergent species if they show the presence of multiple conserved *cis*-regulatory sequences among divergent species lineages. Comparative genomic analyses have shown that several sequence elements of the *Hoxa2* r3/5ER are conserved across evolutionarily divergent vertebrates from shark to tetrapods and teleosts (Tumpel et al., 2002 and 2006; this study). Based on the kernel hypothesis postulated by Davidson (2006), the medaka *hoxa2a* and *ψhoxa2b* r3/5ERs would be expected to direct reporter gene expression in r3 and r5 of their homologous host and within heterologous model systems. However, the medaka *hoxa2a* r3/5ER sequence directed expression solely in r4 of the hindbrain when it was tested in the homologous host but did not direct expression in the hindbrain of chicken embryos. On the contrary, the presence of conserved Hox/Pbx and Prep/Meis binding sites across orthologous *Hoxa2* sequences from evolutionarily divergent vertebrates would suggest that there would be a conserved r4/CNCC-specifying element upstream of *Hoxa2* genes throughout vertebrates. However, functional

genomic analyses of the mouse and chicken *Hoxa2* r3/5ERs did not show reporter gene expression in r4, despite the fact that these enhancer modules contain conserved Hox/Pbx and Prep/Meis binding sites (Maconochie et al., 2001; Tümpel et al., 2002). These analyses clearly show that the *cis*-regulatory sequences, the *trans*-acting factors or a combination of both have diverged between the r3/5ERs and the r4/CNCC-specifying elements of medaka *hoxa2a* and tetrapod *Hoxa2*. These analyses also show that more caution should be used when generating hypotheses of the functional nature of enhancers based on comparative genomic analyses.

In contrast to the medaka *hoxa2a* r4/CNCC-specifying element, a medaka *ψhoxa2b* r3-7/CNCC-specifying element was observed in this study. These results are interesting because *in vivo* expression of medaka *ψhoxa2b* is found mainly in noncanonical *Hox* PG2 domains (see Chapter 2). These results suggest that flanking genomic sequences surrounding the medaka *ψhoxa2b* r3/5ER function to redirect the *ψhoxa2b* r3/5ER to drive *ψhoxa2b* expression in the noncanonical *Hox* PG2 domains. Identifying the mechanism of redirection of the medaka *ψhoxa2b* r3-7/CNCC-specifying element during medaka *ψhoxa2b* expression will aid our understanding of how enhancers evolve between species. As mentioned above, the genomic sequences containing the Hox/Pbx and Prep/Meis sites upstream of *Hoxa2* in mouse and chicken were not observed to direct expression in r4 (Maconochie et al., 2001; Tümpel et al., 2002). It is quite possible that similar mechanisms that restrict *ψhoxa2b* expression from the hindbrain and pharyngeal arches and direct it in the pectoral fins, trunk and ventral-most aspect of the neural tube operate in the *Hoxa2* r3/5ER of mouse and chicken to direct expression in r3 and r5 but restrict it from r4.

Reporter Gene Expression Assays in Homologous and Heterologous Developmental Systems

In Chapter 3 of this thesis, I showed that using medaka as a model system for analyzing the r3/5ERs upstream of medaka *hoxa2a* and *ψhoxa2b* produced different results from those documented by Tümpel et al. (2006) wherein these sequences were tested in chicken embryos. These reporter gene expression results underscore the importance of using a homologous model system for analyzing *cis*-regulatory element control on gene expression during embryonic development. Although the medaka *hoxa2a* r3/5ER did not direct reporter gene expression in r3 or r5 of the hindbrain in transgenic medaka embryos in this study or chicken embryos in the study by Tümpel et al. (2006), the medaka *hoxa2a* r3/5ER was shown to direct reporter gene expression in r4 and the CNCCs of medaka embryos but not in chicken embryos. Further, although the medaka *ψhoxa2b* r3/5ER directed reporter gene expression in r3 and r5 in both medaka and chicken embryos, there was no reporter gene expression observed in r4, r6 and r7 of the hindbrain or in the CNCCs in chicken embryos (Tümpel et al., 2006). The lack of medaka *hoxa2a* and *ψhoxa2b* r3/5ER-driven reporter gene expression in r4 and the CNCCs and in r4, r6, r7 and the CNCCs, respectively, of chick embryos suggests that these enhancer regions were not efficient in utilizing the *trans*-acting factors that were present in the heterologous chick model system. In support of this hypothesis, Tümpel et al. (2002) showed that the r3/5ER of mouse *Hoxa2* was able to generate strong reporter gene expression in r3, r5 and the CNCCs in mouse embryos but no reporter gene expression was detected for the mouse *Hoxa2* r3/5ER in the hindbrain or CNCCs of heterologous chick embryos. Conversely, the r3/5ER of chick *Hoxa2* was able to direct reporter gene expression in r3 and r5 of both chick and mouse embryos (Tümpel et al., 2002). These results from Tümpel et al. (2002) along with our reporter gene expression results of the medaka *hoxa2a* and *ψhoxa2b* r3/5ERs in medaka embryos suggest that either the functional nature of the sequence elements of r3/5ER, the transcription factors of the

genetic regulatory networks that interact with these sequences or a combination of both have diverged greatly among evolutionarily divergent osteichthyans. Future analyses that use heterologous model systems for studying *cis*-regulatory element evolution must take into account the possibility of divergence of expression systems between evolutionarily divergent species. Unfortunately, the difficulty of acquiring embryos in suitable quantities on a rountine basis and under controlled conditions for many animal species makes homologous reporter gene expression analyses impossible to use for studying *cis*-regulatory element control among a broad range of animals.

Although caution should be exercised in drawing conclusions from expression studies conducted using heterologous reporter gene systems, these analyses should be performed to evaluate the degree of conservation of the genetic regulatory networks across evolutionarily divergent teleosts that exhibit divergent *Hox* PG2 gene complements but show conserved *Hox* PG2 gene expression patterns in the hindbrain and pharyngeal arches. Zebrafish and medaka would be excellent candidates for this type of study. Zebrafish has retained *hoxa2b* but lost *hoxa2a* and medaka has retained *hoxa2a* but lost *hoxa2b*. Both functional *hoxa2* genes of zebrafish and medaka show conserved expression patterns in the hindbrain and CNCCs of the pharyngeal arches. Further, the Hox/Pbx and Prep/Meis binding sites within the medaka *hoxa2a* r4/CNCC-specifying element were shown to be conserved in the orthologous genomic sequence of zebrafish *hoxa2b*. A cross-examination of the r3/5ERs and r4/CNCC-specifying elements of zebrafish *hoxa2b* and medaka *hoxa2a* would provide an excellent basis for the evolution of genetic regulatory networks and regulatory enhancer modules between evolutionarily divergent osteichthyan species.

Medaka *Hox* Expression Patterns in the Posterior Pharyngeal Arches

Hox genes occupy a significant role in patterning the bony architecture arising from the pharyngeal arches (Gendron-Maguire et al., 1993; Rijli et al., 1993; Grammatopoulus et al., 2000; Hunter and Prince, 2002; Baltzinger et al., 2005; Crump et al., 2006; Minoux et al., 2009; Le Pabic et al., 2010). The nested expression patterns of *Hox* genes along the anterior-posterior axis of developing embryos has given rise to the *Hox* code hypothesis, which states that a given combination of *Hox* gene products is necessary for specifying segmental identities along the A-P axis (Krumlauf, 1994). In Chapter 4 of this thesis, I provide the first comprehensive analysis of *Hox* PG3-6 gene expression patterns in the posterior pharyngeal arches of a teleost. Given that medaka and other beloniform fishes possess autopomorphic bony characteristics specific to the pharyngeal jaw apparatus that arise from the posterior arches, such as the reduction in size of the second and third epibranchials (arising from the dorsal regions of the fourth and fifth PAs), the great expansion of the articular surface of the fourth epibranchial (arising from the dorsal region of PA6) and the presence of large ventral flanges on the fifth ceratohyal (arising from PA7) (Rosen and Parenti, 1981; Parenti, 1987), it is possible that the nested *Hox* gene expression patterns presented in Chapter 4 represent a *Hox* code that is specific to beloniform fishes. More complete investigations of nested *Hox* PG3-6 gene expression patterns in evolutionarily divergent teleosts, including tilapia and zebrafish, must be performed to better predict the role played by overlapping expression of *Hox* genes in shaping divergent morphological features derived from the posterior pharyngeal arches.

LITERATURE CITED

Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, Postlethwait JH 1998. Zebrafish *hox* clusters and vertebrate genome evolution. Science 282:1711-1714.

Amores A, Suzuki T, Yan YL, Pomeroy J, Singer A, Amemiya C, Postlethwait JH. 2004. Developmental roles of pufferfish *Hox* clusters and genome evolution in ray-fin fish. Genome Res 14:1-10.

Aparicio S, Hawker K, Cottage A, Mikawa Y, Zuo L, Venkatesh B, Chen E, Krumlauf R, Brenner S. 1997. Organization of the *Fugu rubripes* Hox clusters: evidence for continuing evolution of vertebrate Hox complexes. Nature Genetics 16(1):79-83.

Amemiya CT., Powers TP, Prohaska SJ, Grimwood J, Schmutz J, Dickson M, Miyake T, Schoenborn MA, Myers, RM, Ruddle, FH, Stadler PF. 2010. Complete *Hox* cluster characterization of the coelacanth provides further evidence for slow evolution of its genome. Proc Natl Acad Sci 107:3622-3627.

Baltzinger M, Ori M, Pasqualetti M, Nardi I, Rijli FM. 2005. *Hoxa2* knockdown in *Xenopus* results in hyoid to mandibular homeosis. Dev Dyn 234:858-867.

Barrow JR, Capecchi MR. 1996. Targeted disruption of the Hoxb-2 locus in mice interferes with expression of Hoxb-1 and Hoxb-4. Development 122:3817-3828.

Barrow JR, Stadler HS, Capecchi MR. 2000. Roles of Hoxa1 and Hoxa2 in patterning the early hindbrain of the mouse. Development 127:933-944.

Carroll SB. 2008. Evo-Devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134:25-36.

Chandrasekhar A. 2004. Turning heads: development of vertebrate branchiomotor neurons. Dev Dyn 229:143-161.

Chiu C, Dewar K, Wagner GP, Takahashi K, Ruddle F, Ledje C, Bartsch P, Scemama J, Stellwag EJ, Fried C, Prohaska SJ, Stadler PF, Amemiya CT. 2004. Bicher *HoxA* cluster sequence reveals surprising trends in ray-finned fish genomic evolution. Genome Res 14:11-17.

Church DM, Goodstadt L, Hillier LW, Zody MC, Goldstein S, She X, Bult CJ, Agarwala R, Cherry JL, DiCuccio M, Hlavina W, Kapustin Y, Meric P, Maglott D, Birtle Z, Marques AC, Graves T, Zhou S, Teague B, Potamousis K, Churas C, Place M, Herschleb J, Runnheim R, Forrest D, Amos-Landgraf J, Schwartz DC, Cheng Z, Lindblad-Toh K, Eichler EE, Ponting CP. 2009. Lineage-specific biology revealed by a finished genome assembly of the mouse. PLOS Biol. 7:E1000112.

Condie BG, Capecchi MR. 1993. Mice homozygous for a targeted disruption of *Hoxd-3* (*Hox-4.1*) exhibit anterior transformations of the first and second cervical vertebrae, the atlas and the axis. Development 119:579-595.

Couly G, Creuzet S, Bennaceur S, Vincent C, Le Douarin NM. 2002. Interactions between Hox-negative cephalic neural crest cells and the foregut endoderm in patterning the facial skeleton in the vertebrate head. Development 129:1061-1073.

Creuzet S, Couly G, Le Douarin NM. 2005. Patterning the neural crest derivatives during development of the vertebrate head: insights from avian studies. J Anat 207:447-459.

Crump JG, Swartz ME, Eberhart JK, Kimmel CB. 2006. Moz-dependent Hox expression controls segment-specific fate maps of skeletal precursors in the face. Development 133:2661-2669.

Davenne M, Maconochie MK, Neun R, Pattyn A, Chambon P, Krumlauf R, Rijli FM. 1999. Hoxa2 and Hoxb2 control dorsoventral patterns of neuronal development in the rostral hindbrain. Neuron 22:677-691.

Davis A, Scemama JL, Stellwag EJ. 2008. Japanese medaka *Hox* paralog group 2: insights into the evolution of *Hox* PG2 gene composition and expression in the Osteichthyes. J Exp Zool (Mol Dev Evol) 310B:623-641.

Davis A, Stellwag EJ. 2010. Spatio-temporal patterns of *Hox* paralog group 3-6 gene expression during Japanese medaka (*Oryzias latipes*) embryonic development. Gene Expr Patterns 10:244-250.

Davidson EH. 2006. The regulatory genome: gene regulatory networks in development and evolution. Elsevier, Burlington, MA. 289pp.

Ferretti E, Marshall H, Popperl H, Maconochie M, Krumlauf R, Blasi F. 2000. Segmental expression of Hoxb2 in r4 requires two separate sites that integrate cooperative interactions between Prep1, Pbx and Hox proteins. Development 127:155-166.

Ferrier DE, Minguillon C, Holland PW, Garcia-Fernandez J. 2000. The amphioxus Hox cluster: deuterostome posterior flexibility and Hox14. Evolution & Development 2:284-293.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531-1545.

Frasch M, Chen X, Lufkin T. 1995. Evolutionary-conserved enhancers direct region-specific expression of the murine *Hoxa-1* and *Hoxa-2* loci in both mice and Drosophila. Development 121:957-974.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. Vista: computational tools for comparative genomics. Nucleic Acids Res 32(Web Serve Issue):W273-9.

Gavalas A, Davenne M, Lumsden A, Chambon P, Rijli FM. 1997. Role of *Hoxa-2* in axon pathfinding and rostral hindbrain patterning. Development 124:3693-3702.

Gavalas A, Ruhrberg C, Livet J, Henderson CE, Krumlauf R. 2003. Neuronal defects in the hindbrain of *Hoxa1*, *Hoxb1* and *Hoxb2* mutants reflect regulatory interactions among these *Hox* genes. Development:dev.00802.

Gendron-Maguire M, Mallo M, Zhang M, Gridley T. 1993. *Hoxa-2* mutant mice exhibit homeotic transformation of skeletal elements derived from cranial neural crest. Cell 75:1317-1331.

Giese H, Owen M. 1992. The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. Cell. 69:185-195.

Grammatopoulos GA, Bell E, Toole L, Lumsden A, Tucker AS. 2000. Homeotic transformation of branchial arch identity after Hoxa2 overexpression. Development 127:2355-5365.

Haldane, JBS. 1933. The par t played by recurrent mutation in evolution. Am. Nat. 67:5-9.

Hoegg S, Boore JL, Kuehl JV, Meyer A. 2007. Comparative phylogenomic analyses of teleost fish Hox gene clusters: lessons from the cichlid fish *Astatotilapia burtoni*. BMC Genomics 8:317.

Holland PW, Garcia-Fernandez J. 1996. Hox genes and chordate evolution. Developmental Biology (Orlando) 173:382-395.

Hunter M, Prince VE. 2002. Zebrafish Hox paralogue group 2 genes function redundantly as selector genes to pattern the second pharyngeal arch. Dev Biol 247:367-389.

Hurley IA, Scemama J, Prince VE. 2007. Consequences of Hoxb1 duplication in teleost fish. Evol Dev 9:540-554.

Hurley IA, Mueller RL, Dunn KA, Schmidt EJ, Friedman M, Ho RK, Prince VE, Ziheng Y, Thomas MG, Coates MI. 2007. A new time-scale for ray-finned fish evolution. Proc Biol Sci 274:489-498.

Inohaya K, Yasumasu S, Ishimaru M, Ohyama A, Iuchi I, Yamagami K. 1995. Temporal and spatial patterns of gene expression for the hatching enzyme in the teleost embryo, Oryzias latipes. Dev Biol 171:374-385.

Inohaya K, Yasumasu S, Yasumasu I, Iuchi I, Yamagami K. 1999. Analysis of the origin and development of hatching gland cells by transplantation of the embryonic shield in the fish, Oryzias latipes. Dev Growth Differ 41:557-566.

Iwamatsu T. 2004. Stages of normal development in the medaka *Oryzias latipes*. Mech Dev 12:605-618.

Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthouard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biemont C, Skalli Z, Cattolico L, Poulain J, De Berardinis V, Cruaud C, Duprat S, Brottier P, Coutanceau JP, Gouzy J, Parra G, Lardier G, Chapple C, McKernan KJ, McEwan P, Bosak S, Kellis M, Volff JN, Guigo R, Zody MC, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, Robinson-Rechavi M, Laudet V, Schachter V, Quetier F, Saurin W, Scarpelli C, Wincker P, Lander ES, Weissenbach J, Roest Crollius H. 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431:946-957.

Kage T, Takeda H, Yasuda T, Maruyama K, Yamamoto N, Yoshimoto M, Araki K, Inohaya K, Okamoto H, Yasumasu S, Watanabe K, Ito H, Ishikawa Y. 2004. Morphogenesis and regionalization of the medaka embryonic brain. J Comp Neurol 476:219-239.

Karlsson J, Hofsten JV, Olsson PE. 2001. Generating transparent zebrafish: a refined method to improve detection of gene expression during embryonic development. Marin Biotechnol 3:522-527.

Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y, Jindo T, Kobayashi D, Shimada A, Toyoda A, Kuroki Y, Fujiyama A, Sasaki T, Shimizu A, Asakawa S, Shimizu N, Hashimoto S, Yang J, Lee Y, Matsushima K, Sugano S, Sakaizumi M, Narita T, Ohishi K, Haga S, Ohta F, Nomoto H, Nogata K, Morishita T, Endo T, Shin-I T, Takeda H, Morishita S, Kohara Y. 2007. The medaka draft genome and insights into vertebrate genome evolution. Nature 447:714-719.

Kawakami K. 2007. *Tol2*: a versatile gene transfer vector in vertebrates. Genome Biology. 8:S7.

Kim CB, Amemiya C, Bailey W, Kawasaki K, Mezey J, Miller W, Minoshima S, Shimizu N, Wagner G, Ruddle F. 2000. *Hox* cluster genomics in the horn shark, *Heterodontus francisci*. Proc Natl Acad Sci 97:1655-1660.

Kimmel CB, Miller CT, Kruze G, Ullmann B, BreMiller RA, Larison KD, Snyder HC. 1998. The Shaping of Pharyngeal Cartilages during Early Development of the Zebrafish,. Developmental Biology 203:245-263.

Kimmel CB, Miller CT, Keynes RJ. 2001. Neural crest patterning and the Evolution of the jaw. J Anat 199:105-120.

Kirchen RV, West WR. 1999. The Japanese Medaka: Its Care and Development. Carolina Biological Supply Co., Burlington, NC. 36pp.

Krumlauf R. 1994. Mouse *Hox* genetic functions. Curr Op Gen Dev 3:621-625.

Kurosawa G, Takamatsu N, Takahashi M, Sumitomo M, Sanaka E, Yamada K, Nishii K, Matsuda M, Asakawa S, Ishiguro H, Miura K, Kurosawa Y, Shimizu N, Kohara Y, Hori H. 2006. Organization and structure of hox gene loci in medaka genome and comparison with those of pufferfish and zebrafish genomes. Gene 370:75-82.

Lampe X, Samad OA, Guiguen A, Matis C, Remacle S, Picard JJ, Rijli FM, Rezsohazy R. 2008. An ultraconserved Hox-Pbx responsive element resides in the coding sequence of *Hoxa2* and is active in rhombomere 4. Nucleic Acids Res 36:3214-3225.

Lang M, Hadzhiev Y, Siegel N, Amemiya CT, Parada C, Strähle U, Becker M, Müller F, Meyer A. 2010. Conservation of *shh cis*-regulatory architecture of the coelacanth is consistent with its ancestral phylogenetic position. EVODEVO 1:1-11.

Langille RM, Hall BK. 1987. Development of the head skeleton of the Japanese medaka, *Oryzias latipes* (Teleostei). J Morphol 193:135-158.

LaRonde-LeBlanc NA, Wolberger C. 2003. Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. Genes Dev 17:2060-2072.

Le Pabic P, Stellwag EJ, Brothers SN, Scemama JL. 2007. Comparative analysis of Hox paralog group 2 gene expression during Nile tilapia (Oreochromis niloticus) embryonic development. Dev Genes Evol 217(11-12):749-758.

Le Pabic P, Scemama JL, Stellwag EJ. Evolutionary divergence of pharyngeal arch specification in teleosts. www.sicb.org/meetings/2008/schedule/abstractdetails.php3?id=521.

Le Pabic P, Stellwag EJ, Scemama J. 2009. Embryonic development and skeletogenesis of the pharyngeal jaw apparatus in the cichlid Nile tilapia (*Oreochromis niloticus*). Anat Rec 292:1780-1800.

Le Pabic P, Scemama J, Stellwag EJ. 2010. Role of *Hox* PG2 genes in Nile tilapia pharyngeal arch specification: implications for gnathostome pharyngeal arch evolution. Evol Dev 12:45-60.

Lee AP, Koh EGL, Tay A, Brenner S, Venkatesh B, 2006. Highly conserved syntenic blocks at the vertebrate *Hox* loci and conserved regulatory elements within and outside *Hox* gene clusters. Proc Natl Acad Sci 103:6994-6999.

Lovejoy NR, Iranpour M, Collette BR. 2004. Phylogeny and jaw ontogeny of beloniform fishes. Integr Comp Biol 44:366-377.

Luo J, Stadler PF, He SP, Meyer A. 2007. PCR survey of *Hox* genes in the the goldfish *Carassius auratus auratus*. J Exp Zool (Mol Dev Evol) 308B:250-258.

Lynch M, Conery JS. 2003. The evolutionary demography of duplicate genes. J Struct Funct Genomics 3:35-44.

Maconochie, M.K., Krishnamurthy, R., Nonchev, S., Meier, P., Manzanares, M., Mitchell, P.J., Krumlauf, R. 1999. Regulation of *Hoxa2* in cranial neural crest cells involves members of the *AP-2* family. Development. 126:1483-1494.

Maconochie MK, Nonchev S, Manzanares M, Marshall H, Krumlauf R. 2001. Differences in Krox20-dependent regulation of Hoxa2 and Hoxb2 during hindbrain development. Dev Biol 233(2):468-481.

Malaga-Trillo E, Meyer A. 2001. Genome duplications and accelerated evolution of *Hox* genes and cluster architecture in teleost fishes. Am Zool 41:676-686.

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

Metscher BD, Ahlberg, PE. 1999. Zebrafish in context: uses of a laboratory model in comparative studies. Dev Biol 210:1-14.

Miller CT, Maves L, Kimmel CB. 2004. moz regulates Hox expression and pharyngeal segmental identity in zebrafish. Development 131:2443-2461.

Minoux M, Antonarakis GS, Kmita M, Duboule D, Rijli FM. 2009. Rostral and caudal pharyngeal arches share a common neural crest ground pattern. Development 136:637-645.

Miya M, Takeshima H, Endo H, Ishiguro NB, Inoue JG, Mukai T, Satoh TP, Yamaguchi M, Kawaguchi A, Mabuchi K, Shirai SM, Nishida M. 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. Mol Phylogenet Evol 26:121-138.

Moens C, Yan Y, Appel B, Force A, Kimmel C. 1996. valentino: a zebrafish gene required for normal hindbrain segmentation. Development 122:3981-3990.

Moghadam H, Ferguson M, Danzmann R. 2005. Evolution of Hox Clusters in Salmonidae: A Comparative Analysis Between Atlantic Salmon (*Salmo salar*) and Rainbow Trout (*Oncorhynchus mykiss*). Journal of Molecular Evolution 61:636.

Mungpakdee S, Seo H, Angotzi AR, Dong X, Akalin A, Chourrout D. 2008a. Differential evolution of the 13 Atlantic salmon hox clusters. Mol Biol Evol 25:1333-1343.

Mungpakdee S, Seo H, Chourrout D. 2008b. Spatio-temporal expression patterns of anterior *Hox* genes in Atlantic salmon (*Salmo salar*). Gene Expr Patterns 8:508-514.

Naruse K, Tanaka M, Mita K, Shima A, Postlethwait J, Mitani H. 2004. A medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping. Genome research 14:820-828.

Nelson JS. 1994. Fishes of the world, third ed. Wiley, New York.

Nonchev S, Maconochie M, Vesque C, Aparicio S, Ariza-McNaughton L, Manzanares M, Maruthainar K, Kuroiwa A, Brenner S, Charnay P, Krumlauf R. 1996a. The conserved role of Krox-20 in directing Hox gene expression during vertebrate hindbrain segmentation. Proc Natl Acad Sci 93:9339-9345.

Nonchev S, Vesque C, Maconochie M, Seitanidou T, Ariza-McNaughton L, Frain M, Marshall H, Sham MH, Krumlauf R, Charnay P. 1996b. Segmental expression of Hoxa-2 in the hindbrain is directly regulated by Krox-20. Development 122:543-554.

Oosterwegel M, van de Wetering M, Timmerman J, Kruisbeek A, Destree O, Meijlink F, Clevers H. 1993. Differential expression of the HMG box factors *TCF-1* and *LEF-1* during murine embryogenesis. Development 118:439-448.

Oury F, Murakami Y, Renaud JS, Pasqualetti M, Charnay P, Ren SY, Rijli FM. 2006. Hoxa2- and rhombomere-dependent development of the mouse facial somatosensory map. Science 313(5792):1408-1413.

Oxendine SL, Cowden J, Hinton DE, Padilla S. 2006. Adapting the medaka embryo assay to a high-throughput approach for developmental toxicity testing. NeuroToxicology 27:840-845.

Oulion S, Debiais-Thibaud M, d'Aubenton-Carafa Y, Thermes C, Da Silva C, Bernard-Samain S, Gavory F, Wincker P, Mazan S, Casane D. 2010. Evolution of *Hox* gene clusters in gnathostomes: insights from a survey of a shark (*Scyliorhinus canicula*) transcriptome. Mol Biol Evol 27:2829-2838.

Oulion S, Borday-Birraux V, Debiais-Thibaud M, Mazan S, Laurenti P, Casane D. 2011. Evolution of repeated structures along the body axis of jawed vertebrates, insights from the *Scyliorhinus canicula* Hox code. Evol. Dev. 13:247-259.

Parenti LR. 1987. Phylogenetic aspects of tooth and jaw structure of the Medaka, *Oryzias latipes*, and other beloniform fishes. J Zool Lond 211:561-572.

Pasqualetti M, Ori M, Nardi I, Rijli FM. 2000. Ectopic *Hoxa2* induction after neural crest migration results in homeosis of jaw elements in *Xenopus*. Development 127:5367-5378.

Piper DE, Batchelor AH, Chang C-P, Cleary ML, Wolberger C. 1999. Structure of a HoxB1-Pbx1 Heterodimer Bound to DNA: Role of the Hexapeptide and a Fourth Homeodomain Helix in Complex Formation. Cell 96:587-597.

Powers TP, Amemiya CT. 2004. Evidence for a Hox14 paralog group in vertebrates. Curr Biol 14:R183-184.

Prince VE, Moens CB, Kimmel CB, Ho RK. 1998. Zebrafish *hox* genes: expression in the hindbrain region of wild-type and mutants of the segmentation gene, *valentino*. Development 125:393-406.

Prince V. 2002: The *Hox* paradox: more complex(es) than imagined. Dev Biol 249:1-15.

Raincrow JD. 2010. *Hox* cluster intergenic sequence evolution. Ph.D. Thesis. Rutgers University.

Reubens MC. 2009. Transient transgenic vector system for enhancer mapping. M.S. Thesis. East Carolina University.

Rijli FM, Mark M, Lakkaraju S, Dierich A, Dolle P, Chambon P. 1993. A homeotic transformation is generated in the rostral branchial region of the head by disruption of *Hoxa-2*, which acts as a selector gene. Cell 75:1333-1349.

Rosen DE and Parenti LR. 1981. Relationships of *Oryzias*, and the groups of atherinomorph fishes. Am Mus Novit 2719:1-25.

Sandelin A, Alkema W, Engstrom P, Wasserman WW, and Lenhard B. 2004. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res 32:D91-4.

Santagati F, Minoux M, Ren SY, Rijli FM. 2005. Temporal requirement of *Hoxa2* in cranial neural crest skeletal morphogenesis. Development 132:4927–4936.

Scemama JL, Hunter M, McCallum J, Prince V, Stellwag E. 2002. Evolutionary divergence of vertebrate *Hoxb2* expression patterns and transcriptional regulatory loci. J Exp Zool 294:285-299.

Scemama JL, Vernon JL, Stellwag E. 2006. Differential expression of Hoxa2a and Hoxa2b genes during striped bass embryonic development. . Gene Expr Patterns 6:843-848.

Schaeffer B, Rosen DE. 1961. Major adaptive levels in the evolution of the actinopterygian feeding mechanism. Am Zoologist 1:187-204.

Schilling TF, Kimmel CB. 1997. Musculoskeletal patterning in the pharyngeal segments of the zebrafish embryo. Development 124:2945-2960.

Schug J, Overton GC. 1997. TESS: transcription element serach software on the WWW. Technical report CBIL-TR-1997-1001-v0.0, of the computational biology and informaticslaboratory. School of Medicine, University of Pennsylvania.

Sham MH, Vesque C, Nonchev S, Marshall H, Frain M, Gupta RD, Whiting J, Wilkinson D, Charnay P, Krumlauf R. 1993. The zinc finger gene *Krox20* regulates *HoxB2* (*Hox2.8*) during hindbrain segmentation. Cell 72:183-196.

Shestopalov IA, Sinha S, Chen JK. 2007. Light-controlled gene silencing in zebrafish embryos. Nat. Chem. Biol. 3:650-651.

Steinke D, Salzburger W, Meyer, A. 2006. Novel relationships among ten fish model species based on a phylogenomic analysis using ESTs. J Mol Evol 62:772-784.

Stellwag EJ. 1999. Hox gene duplication in fish. Seminars in Cell and Developmental Biology 10:531-540.

Stock DW, Jackman WR, Trapani J. 2006. Developmental genetic mechanisms of evolutionary tooth loss in cypriniform fishes. Development 133:3127-3137.

Subramanian A, Kaufmann M, Morgenstern B. 2008. DIALIGN-TX: greedy and progressive approaches for segment-based multiple sequence alignment. http://www.almob.org/content/3/1/6

Suster ML, Kania A, Liao M, Asakawa K, Charron F, Kawakami K, Drapeau P. 2011. A novel conserved *evx1* enhancer links spinal interneuron morphology and *cis*-regulation from fish to mammals. Dev Biol 325:422-433.

Takamatsu N, Kurosawa G, Takahashi M, Inokuma R, Tanaka M, Kanamori A, Hori H. 2007. Duplicated Abd-B class genes in medaka hoxAa and hoxAb clusters exhibit differential expression patterns in pectoral fin buds. Dev Genes Evol 217:263-273.

Taylor JS, Raes J. 2004. Duplication and divergence: the evolution of new genes and old ideas. Annu. Rev. Genet 38:615-643.

Tomasini A J, Schuler AD, Zebala JA., Mayer AN. 2009. Photomorphs: A novel light-activated reagent for controlling gene expression in zebrafish. Genesis 47:736-743.

Tümpel S, Maconochie M, Wiedemann LM, Krumlauf R. 2002. Conservation and diversity in the cis-regulatory networks that integrate information controlling expression of Hoxa2 in hindbrain and cranial neural crest cells in vertebrates. Dev Biol 246:45-56.

Tümpel 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 103:5419-5424.

Tümpel S, Cambronero F, Ferretti E, Blasi F, Wiedemann LM, Krumlauf R. 2007. Expression of Hoxa2 in rhombomere 4 is regulated by a conserved cross-regulatory mechanism dependent upon Hoxb1. Dev Biol 302:646-660.

Tümpel S, Cambronero F, Sims C, Krumlauf R, Wiedemann LM. 2008. A regulatory module embedded in the coding region of *Hoxa2* controls expression in rhombomere 2. Proc Natl Acad Sci 105:20077-20082.

Urasaki A, Morvan G, Kawakami K. 2006. Functional dissection of the *Tol2* transposable element identified the minimal *cis*-sequences and a highly repetitive sequence in the subterminal region essential for transposition. Genetics 174:639-649.

van Genderen C, Okamura RM, Fariñas I, Quo R, Parslow TG, Bruhn L, Grosschedl R. 1994. Development of several organs that require epithelial-mesenchymal interactions is impaired in *LEF-1* defient mice. Genes Dev 8:2691-2703.

Vesque C, Maconochie M, Nonchev S, Ariza-McNaughton L, Kuroiwa A, Charnay P, Krumlauf R. 1996. *Hoxb-2* transcriptional activation in rhombomeres 3 and 5 requires an evolutionarily conserved cis-acting element in addition to the Krox-20 binding site. EMBO Journal 15:5383-5396.

Vieille-Grosjean I, Hunt P, Gulisano M, Boncinelli E, Thorogood P. 1997. Branchial HOX gene expression and human craniofacial development. Developmental Biology (Orlando) 183:49-60.

Voiculescu O, Taillebourg E, Pujades C, Kress C, Buart S, Charnay P, Schneider-Maunoury S. 2001. Hindbrain patterning: *Krox20* couples segmentation and specification of regional identity. Development 128:4967-4978.

Wittbrodt J, Shima A, Schartl M. 2002. Medaka-a model organism from the far east. Nat Rev Genet 3:53-64.

Woltering JM, Durston AJ. 2006. The zebrafish *hoxDb* cluster has been reduced to a singl microRNA. Nature 38:601-602.

Yan Y, Jowett T, Postlethwait JJ. 1998. Ectopic expression of *hoxb2* after retinoic acid treatment or mRNA injection: disruption of hindbrain and craniofacial morphogenesis in zebrafish embryos. Dev Dyn 213:370-385.

Yasutake J, Inohaya K, Kudo A. 2004. *Twist* functions in vertebral column formation in medaka, *Oryzias latipes*. Mech Dev 121:883-894.

APPENDIX A: IUPAC #D239

East Carolina University
1907-2007 CENTENNIAL

Animal Care and Use Committee
212 Ed Warren Life Sciences Building
East Carolina University
Greenville, NC 27834

252-744-2436 office
252-744-2355 fax

March 10, 2010

Ed Stellwag, Ph.D.
Department of Biology
Howell Science Complex
East Carolina University

Dear Dr. Stellwag:

Your Animal Use Protocol entitled, "Investigation of the Evolutionary Role of Jaw Patterning Genes in Japanese Medaka," (AUP #D239) was reviewed by this institution's Animal Care and Use Committee on 3/10/10. The following action was taken by the Committee:

"Approved as submitted"

A copy is enclosed for your laboratory files. Please be reminded that all animal procedures must be conducted as described in the approved Animal Use Protocol. Modifications of these procedures cannot be performed without prior approval of the ACUC. The Animal Welfare Act and Public Health Service Guidelines require the ACUC to suspend activities not in accordance with approved procedures and report such activities to the responsible University Official (Vice Chancellor for Health Sciences or Vice Chancellor for Academic Affairs) and appropriate federal Agencies.

Sincerely yours,

Robert G. Carroll, Ph.D.
Chairman, Animal Care and Use Committee

RGC/jd

enclosure

APPENDIX B: AN ANALYSIS OF THE EVOLUTION OF *HOX* PARALOG GROUP 2 GENE FUNCTION IN THE JAPANESE MEDAKA

Introduction

For my original doctoral research project at East Carolina University, I intended to provide an understanding for the mechanisms of the *Hox* paralog group 2 (PG2) gene functional specification on the 2nd pharyngeal arch (PA2)-derived structures in the Japanese medaka. I intended to conduct a comparative study among divergent species to elucidate how post-genome duplication evolution and divergent secondary gene loss has influenced the development of *Hox* PG2 gene-specified pharyngeal arch and hindbrain structures within the Osteichthyes.

Extensive variation has been observed among osteichthyan *Hox* PG2 gene ortholog and co-ortholog expression in the pharyngeal arches. In tetrapods *Hoxa2* orthologs are expressed in the cranial neural crest streams that migrate out of the r4 and r6/r7 to populate PA2 and the posterior pharyngeal arches, respectively (Fig. A-1A) (Sham et al., 1993; Prince and Lumsden, 1994; Vesque et al., 1996; Vieille-Grosjean et al., 1997; Prince et al., 1998; Yan et al., 1998; Maconochie et al., 1999; Grammatopoulos et al., 2000; Pasqualetti et al., 2000; Baltzinger et al., 2005). The expression of *Hoxa2* in the PA2 persists into the chondrogenic phase of PA2-derived cartilages. By contrast, *Hoxb2* of tetrapods is not expressed in the pharyngeal arches. In zebrafish, both *hoxa2b* and *b2a* are expressed into the chondrogenic phase in PA2 (Fig. A-1B). However, only *hoxa2b* of zebrafish is expressed in the posterior pharyngeal arches. In the Nile tilapia, a teleost that possesses three *Hox* PG2 genes, all three genes are expressed in PA2 and the posterior arches (Fig. A-1C). However, only *hoxa2a* remains expressed into the chondrogenic phase in these embryonic domains.

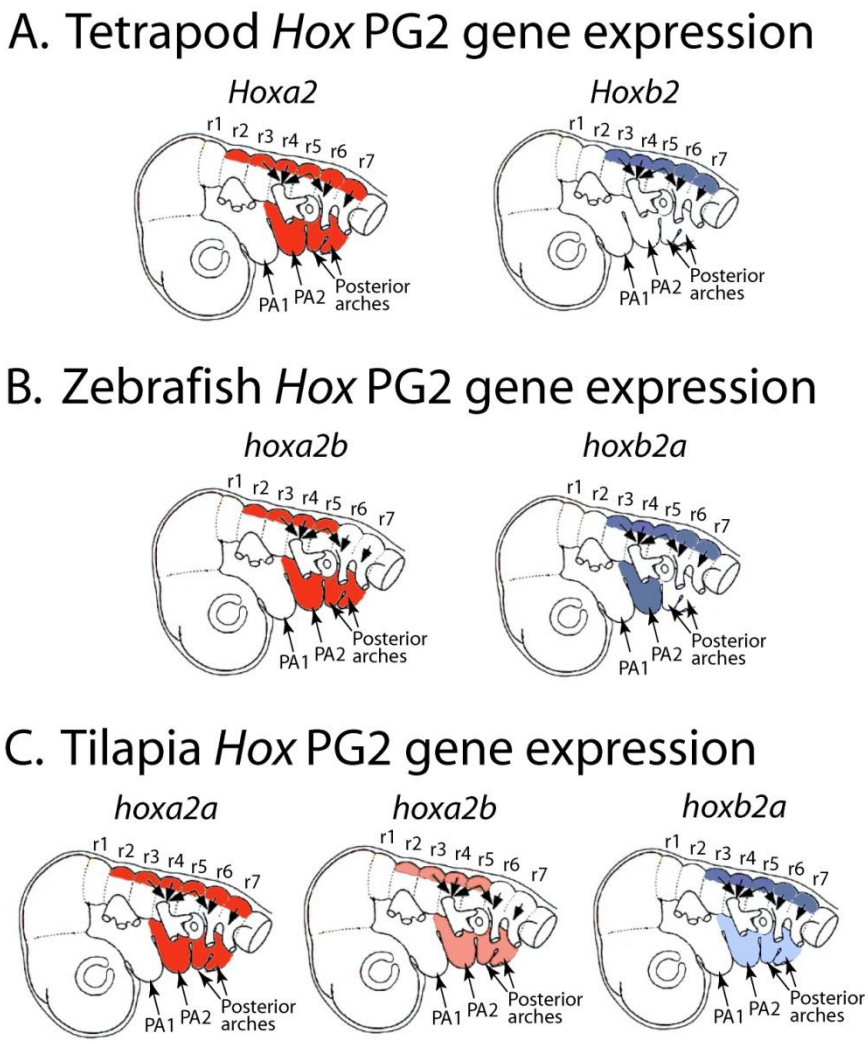


Fig. AB-1. *Hox* PG2 gene expression patterns in tetrapods (A), zebrafish (B) and tilapia (C). Dark red and dark blue colors in the pharyngeal arches correspond to *Hoxa2* and *b2* genes, respectively, that are expressed into the chondrogenic phase. Light red and light blue colors in the pharyngeal arches correspond to *Hoxa2* and *b2* genes, respectively, that are downregulated in expression prior to the chondrogenic phase. Embryos are oriented with their anterior ends facing left and their dorsal sides facing up. PA, pharyngeal arch; r, rhombomere. Arrows in rhombomeres that point downward denote neural crest cell migratory patterns.

The variation in pharyngeal arch-specific *Hox* PG2 gene expression patterns among osteichthyans is mirrored by variation in osteichthyan *Hox* PG2 gene function within the pharyngeal arches. In tetrapods, the *Hoxa2* loss-of-function resulted in anteriorizing homeotic transformations in which the PA2-derived skeletal elements were re-specified to those of PA1 (Gendron-Maguire et al., 1993; Rijli et al., 1993; Baltzinger et al., 2005) (Fig. A-2A). Further, ectopic expression of *Hoxa2* in PA1 of tetrapods resulted in posteriorizing homeotic transformations, wherein PA1-derived skeletal elements were transformed into PA2-like elements (Grammatopoulus et al., 2000; Pasqualetti et al., 2000) (Fig. A-2B). No homeotic transformations were observed in the pharyngeal arches in tetrapod *Hoxb2* null mutants or antisense morpholino-induced knock-down embryos (Barrow and Cappechi, 1996; Baltzinger et al., 2005). In zebrafish, both *hoxa2b* and *b2a* had to be knocked down in order to observe an anteriorizing homeotic transformation in PA2 (Hunter and Prince, 2002), which was interpreted to mean that these two genes functioned redundantly to pattern PA2 (Fig. A-2A). Ectopic expression of either *hoxa2b* or *b2a* caused posteriorizing homeotic transformations in zebrafish, which demonstrated that either gene could function in second arch specification (Fig. A-2B). In tilapia individual knockdowns of *hoxa2a*, *a2b* or *b2a* resulted in the anteriorizing homeotic transformations in PA2, which suggested that each tilapia *Hox* PG2 gene is able to function independently as a selector gene in patterning the PA2 identity (Fig. A-2A) (Le Pabic et al., 2010). Overall, these results have shown that the *Hox* PG2 selector gene activity in PA2 is a characteristic of *Hoxa2* in tetrapods, the redundant action of *hoxa2b* and *b2a* in zebrafish and a possible combination of *hoxa2a*, *a2b* and *b2a* in tilapia.

Results that I have obtained from cDNA cloning, comparative genomics and the study of

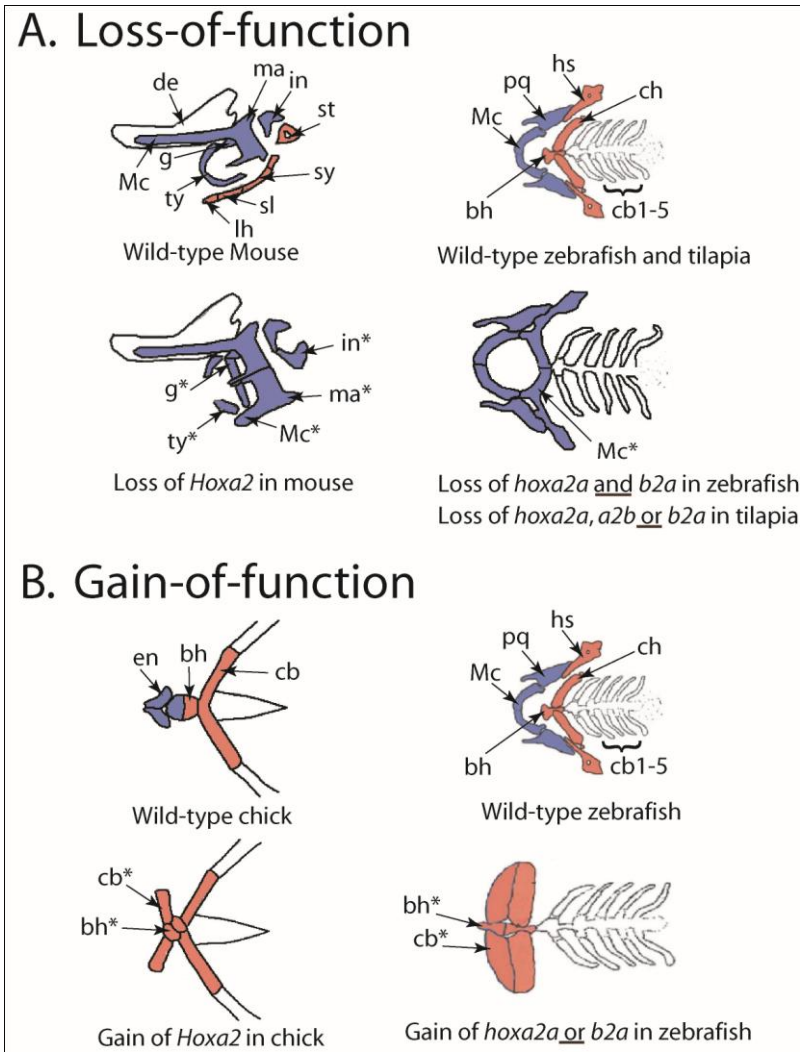


Fig. AB-2. *Hox* PG2 gene loss-of-function (A) and gain-of-function (B) in tetrapods and teleosts. Bones colored in blue represent PA1-derived bones. Bones colored in red represent PA2-derived bones. Bone labels followed by asterisks (*) represent duplicated structures. Bh, basihyal; cb, ceratobranchial; ch, ceratohyal; de, dentary; en, entoglossum; g, gonial bone; hs, hyosymplectic; in, incus; lh, lesser horn of the hyoid; ma, malleus; Mc, Meckel's cartilage; pq, palatoquadrate; sl, stylohyoid ligament; st, stapes; sy, styloid bone; ty, tympanic ring.

embryonic *Hox* PG2 gene expression patterns have revealed that medaka possessed just two functional *Hox* PG2 genes, *hoxa2a* and *b2a* (Davis et al., 2008; Chapter 2). Contrary to expectations based on previously reported evidence (Hunter and Prince, 2002; Scemama et al., 2006; Le Pabic et al., 2007), medaka *hoxa2b* was shown to be a transcribed pseudogene that is expressed in noncanonical *Hox* PG2 gene expression domains (Davis et al., 2008; Chapter 2). Despite uncharacteristic expression patterns for *hoxa2b*, whole-mount *in situ* hybridization assays showed that medaka *hoxa2a* and *b2a* share similar hindbrain expression patterns with the *Hoxa2* and *b2* genes of tetrapods and other teleosts (Davis et al., 2008; Chapter 2). By contrast, the medaka *Hox* PG2 gene expression patterns in PA2 were shown to be most related to *hoxa2b* and *b2a* patterns of PA2 expression in zebrafish, wherein both genes are expressed into the chondrogenic phase. These expression results suggest that the medaka *hoxa2a* and *b2a* genes function similarly to zebrafish *hoxa2b* and *b2a* in patterning PA2.

Based on my results from cloning and expression analyses of medaka *Hox* PG2 genes and from previous functional studies of *Hox* PG2 genes in osteichthyans, I hypothesized that the medaka *Hox* PG2 genes, *hoxa2a* and *b2a*, would function redundantly in patterning the identity of PA2 during embryonic development. I proposed to test this hypothesis through gene knockdown and rescue experiments. Based on previous results from tetrapods and zebrafish, I expected that knockdowns of both *hoxa2a* and *b2a* would create an anteriorizing homeotic transformation, wherein the identity of PA2 would be changed to that of PA1. Specifically, I expected the PA2-derived ceratohyal and hyosymplectic cartilages to be transformed into duplicate Meckel's and palatoquadrate cartilages, respectively. Further, I expected that these transformed cartilages would be in reverse orientation to their wild-type PA1-counterparts. Confirmation of homeotic transformations were to be assessed using direct morphological

observation and *in situ* hybridization with specific developmental markers. Specifically, I expected to observe changes in PA2-specific expression patterns of *bapx1* and *goosecoid*, such that they would resemble expression patterns observed in PA1. Homeotic transformations within PA2 were also to be verified with rescue studies. I expected that genetic rescue experiments would reverse the anteriorizing homeotic transformations in PA2 caused by the knockdown of both *hoxa2a* and *b2a* through the co-microinjection of modified *hoxa2a* or *b2a* mRNAs.

Contrary to my expectations, I did not observe any effects of morpholinos on the development of PA2-specified skeletal derivatives even with the microinjection of different combinations of morpholino concentrations. Although I did observe some morphological changes within the bony elements derived from the posterior pharyngeal arches when using a morpholino specific to the 5'-untranslated region (5'-UTR), these morphant phenotypes occurred when concentrations of the morpholino approached toxic levels. Therefore, these morphant phenotypes may have been the result of toxicity rather than a true knockdown of medaka *hoxa2a*. Further, these morphant phenotypes within the posterior pharyngeal arches were not recapitulated when I used a morpholino specific to the splice junction (SJ) of medaka *hoxa2a*. Ultimately, this project was abandoned and I began a new project that assayed the genomic regulation of medaka *hoxa2a* and *ψhoxa2b* expression in the hindbrain and pharyngeal arches (shown in Chapter 3).

Methods and Materials

Medaka *Hox* PG2 gene loss-of-function analyses were performed using anti-sense oligonucleotide morpholinos (MOs) directed toward the 5'-UTRs (OlaA2a5'UTRMO and OlaB2a5'UTRMO) and SJs (OlaA2aSJMO and OlaB2aSJMO) of the medaka *hoxa2a* and *b2a* mRNAs. Morpholino sequences are listed in Table A-1. The strategy of employing two MOs

per gene was used so that MOs would act at different sites on the mRNA that affect distinctly different molecular mechanisms of mRNA expression. I expected that MO-induced morphants that target distinct sites on the mRNA would generate identical or very similar phenotypes, which would support the validity of the knockdown phenotype. A relatively common concern about MOs is the possibility that MO-induced phenotypes will result from non-specific effects. The use of two MOs for each target gene provides independent corroboration of morphant phenotypes and provides validation of morphant phenotypes.

All MOs were diluted in 1X Yamamoto's Buffer and 1X phenol red at varying concentrations from 0.5 to 20 ng/nl. Control embryos were injected with 1X Yamamoto's buffer and 1X phenol red. Microinjections were performed as described in Chapter 3. All MOs were tested for generating toxic effects on embryos by microinjecting different concentration levels of MOs, separately or in specific combinations. Measurements were taken from wild-type hatchlings (i.e.: overall length of hatchling, distance between eyes, length of pectoral fins, etc.) to serve as a basis for comparison between intoxicated and non-intoxicated embryos. Hatchlings that deviated from the wild-type measurements or underwent gastrulation defects were deemed to be intoxicated and were discarded from the analysis. Normal MO-microinjected and control embryos were observed and assayed for pharyngeal arch-derived cartilage morphant phenotypes at the hatching stage (stage 39; Iwamatsu, 2004). Medaka hatchlings were fixed overnight in 95% ethanol at room temperature (RT), incubated in Alcian blue treatment (80% ethanol (95%), 20% acetic acid, 0.15 mg/ml alcian blue-tetrakis (methylpyridinium) chloride) at RT for at least

Table AB-1. Morpholinos used for loss-of-function study for medaka *Hox* PG2 genes.

Morpholino Name	Morpholino Sequence (5'-3')	Morpholino Targeting Site
OlaA2a5'UTRMO	TTCGTAATTACATCTCCGCTCCCTGA	Medaka *hoxa2a* 5'- UTR
OlaA2aSJMO	CGCGCACTTTTGAGTCACTCACCTT	Medaka *hoxa2a* splice donor
OlaA2a5'UTRMO	CTTCCTCAGCCTGGTTTCAGCCTCA	Medaka *hoxb2a* 5'- UTR
OlaB2aSJMO	TGGCTGGAATCTGATGCAAACAAAC	Medaka *hoxb2a* splice acceptor

3 days for coloration of cartilage, washed twice with 95% ethanol for 1 hr at RT, cleared of opaque tissue for several minutes using 1% KOH at RT and stored in 35% glycerol until observation and microdissection.

Results and Discussion

The morpholino toxicity analyses performed on medaka embryos showed that each of the four morpholinos affected the medaka development at different concentration levels. These levels ranged from 0.6 to 0.75 ng per injection for OlaA2a5'UTRMO (Fig. A-3), 1 ng per injection for OlaB2a5'UTRMO (Fig. A-4), 13 to 14 ng per injection for OlaA2aSJMO (Fig. A-5) and 8 to 9 ng per injection for OlaB2aSJMO (Fig. A-6). Out of all MOs tested, only the morpholino targeting the 5'UTR of medaka *hoxa2a* (OlaA2a5'UTRMO) generated morphant phenotypes specific to the cartilage elements derived from the pharyngeal arches. These phenotypes included loss of the PA3-derived ceratobranchials and reduction or loss of the PA2-derived opercle bone and branchiostegal rays (Fig. A-7). However, these morphant phenotypes only occurred at near toxic and toxic concentrations (0.5, 0.6 and 0.75 ng/injection). Further, these morphant phenotypes were not reproduced using the MO targeting the medaka *hoxa2a* splice junction site (OlaA2aSJMO). Therefore, these morphant phenotypes may have been a result of toxicity rather than specific activity of the OlaA2a5'UTRMO on the pharyngeal arch-derived structures.

Co-microinjection of MOs targeting both medaka *hoxa2a* and *b2a* mRNAs did not produce homeotic transformations of PA2-derived cartilages in medaka embryos. These MOs were co-microinjected in different combinations, including the MOs targeting the 5'-UTRs of

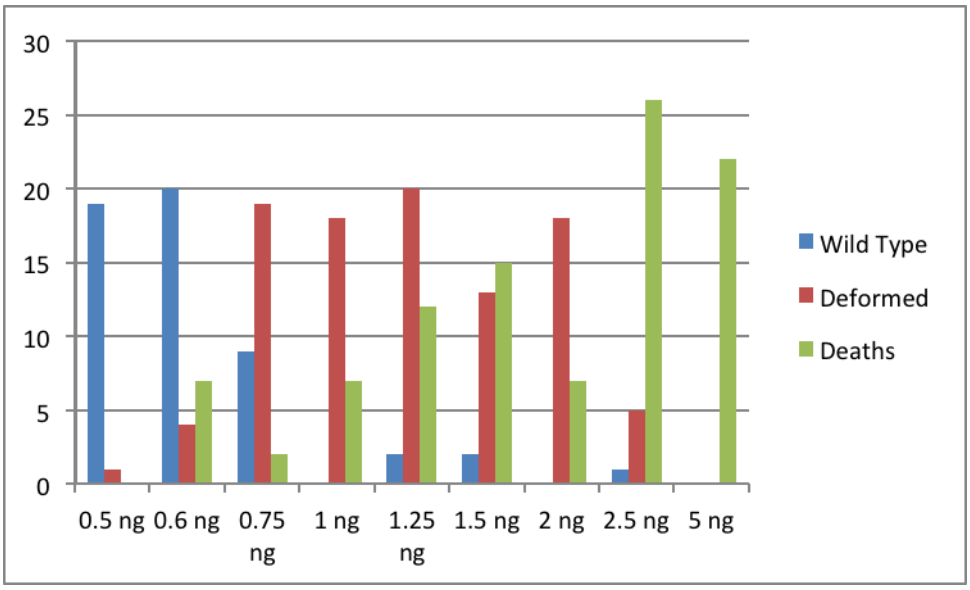


Fig. AB-3. Toxicity analysis of the OlaA2a5'UTRMO. Morpholino concentrations are shown on the X-axis. Numbers of injected embryos are shown on the Y-axis. Blue bars correspond to wild type embryos. Red bars correspond to embryos that were deformed. Green bars correspond to embryos that were killed in the analysis.

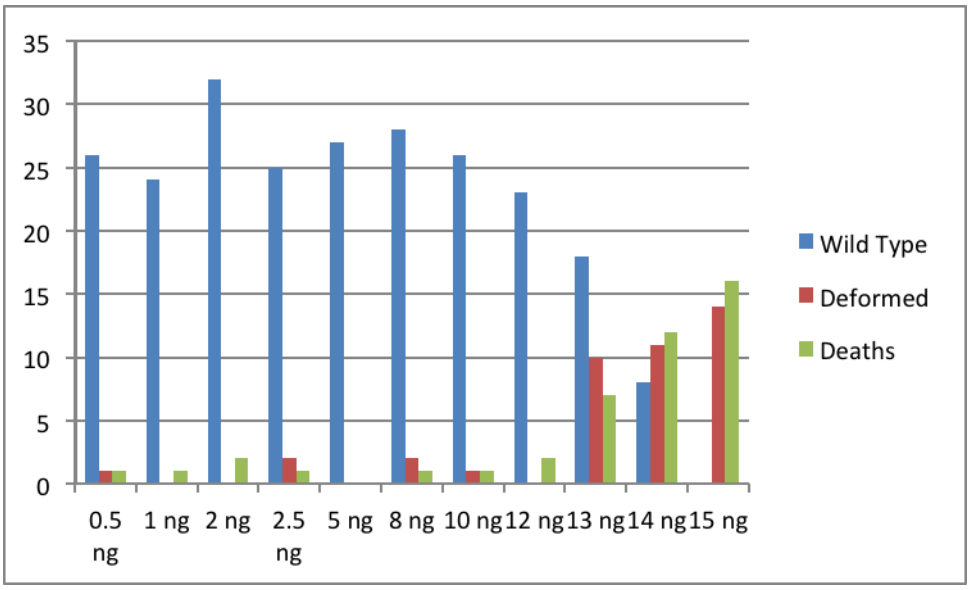


Fig. AB-4. Toxicity analysis of the OlaA2aSJMO. Morpholino concentrations are shown on the X-axis. Numbers of injected embryos are shown on the Y-axis. Blue bars correspond to wild type embryos. Red bars correspond to embryos that were deformed. Green bars correspond to embryos that were killed in the analysis.

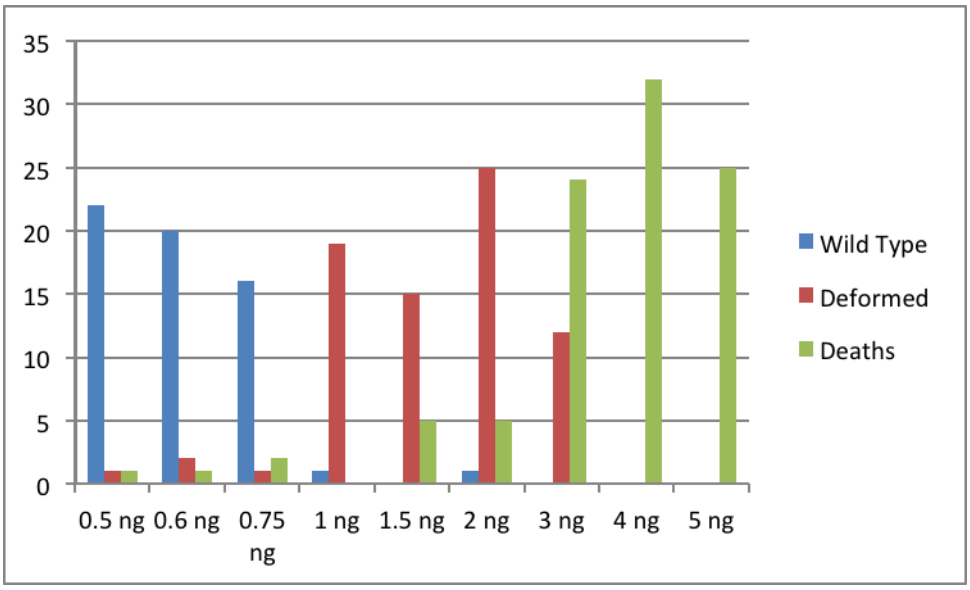


Fig. AB-5. Toxicity analysis of the OlaB2a5'UTRMO. Morpholino concentrations are shown on the X-axis. Numbers of injected embryos are shown on the Y-axis. Blue bars correspond to wild type embryos. Red bars correspond to embryos that were deformed. Green bars correspond to embryos that were killed in the analysis.

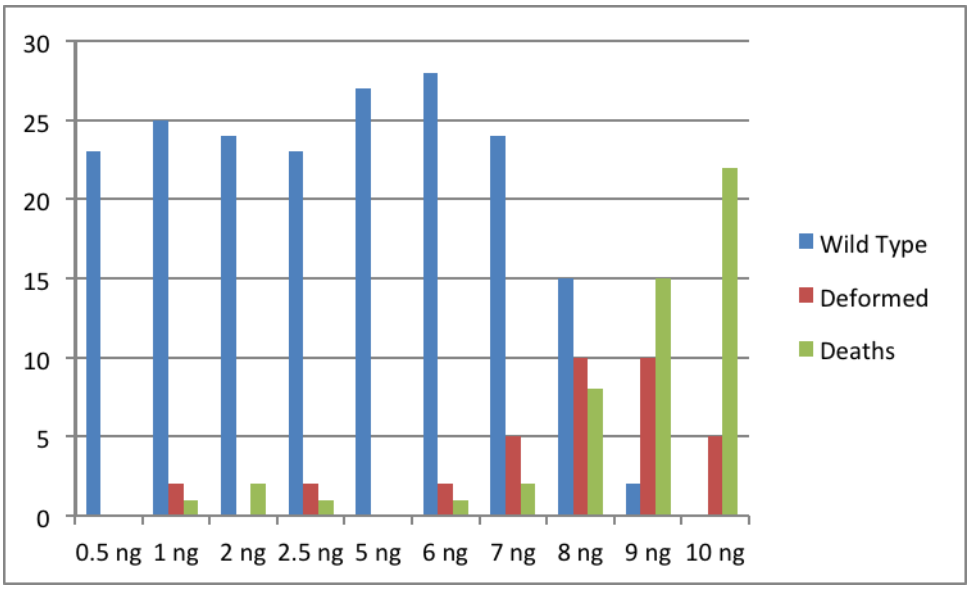


Fig. AB-6. **Toxicity analysis of the OlaB2aSJMO.** Morpholino concentrations are shown on the X-axis. Numbers of injected embryos are shown on the Y-axis. Blue bars correspond to wild type embryos. Red bars correspond to embryos that were deformed. Green bars correspond to embryos that were killed in the analysis.

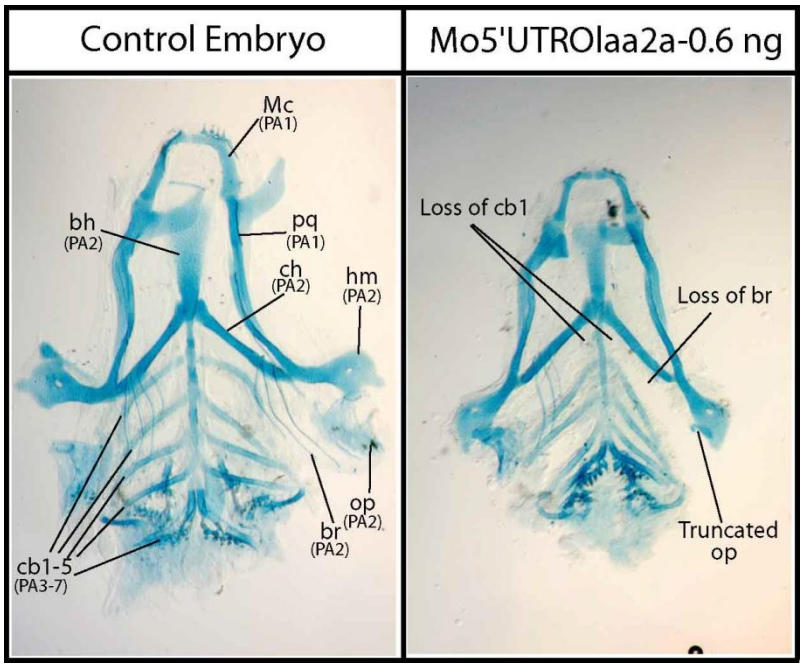


Fig. AB-7. Flatmounts of medaka pharyngeal skeletons from a control embryo (left) and an embryo injected with 0.6 ng of OlaA2a5'UTRMO (right). Flatmounts are oriented with their anterior ends facing up and their ventral sides facing the reader. Bh, basihyal; cb, ceratobranchial, ch; ceratohyal; hm, hyomandibular; Mc, Meckel's cartilage; pq, palatoquadrate; sy, symplectic.

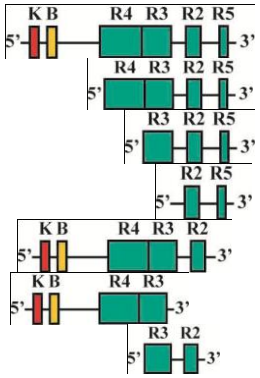
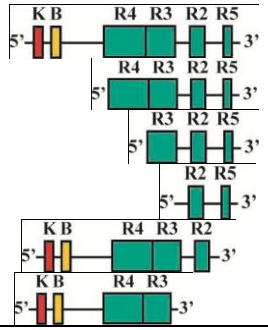
both *hoxa2a* and *b2a*, the MOs targeting the splice junction (SJ) sites of both *hoxa2a* and *b2a*, the MO targeting the 5'-UTR of *hoxa2a* with the MO targeting the SJ of *hoxb2a* and the MO targeting the SJ of *hoxa2a* with the MO targeting the 5'-UTR of *hoxb2a*. Successive concentrations of each MO were used in these analyses until microinjected embryos showed developmental defects due to toxic levels of MOs.

The failure of the MOs to influence a homeotic transformation in PA2 of medaka may have been due to the rate of development of the bony architecture arising from PA2 and the continuous expression of medaka *hoxa2a* and *b2a* up to this chondrogenic stage. Results from functional genetic studies in mouse have shown that continuous *Hox* PG2 gene expression into the chondrogenic phase of PA2 development is necessary for the proper patterning of the PA2-derived cartilages (Santagati et al., 2005). The chondrogenic phase at which the PA2 cartilages begin to form in medaka occurs at developmental stage 34 (121 hpf) (Langille and Hall, 1987; Iwamatsu, 2004). Interestingly, Yasutake et al. (2004) showed that MOs targeting the 5'-UTR of medaka *twist* were only able to knock down *eGFP* expression in *twist*-EGFP medaka transgenic lines for up to 72 hpf. After 72 hpf, *eGFP* expression in *twist*-EGFP transgenic lines exhibited normal expression levels (Yasutake et al., 2004). Therefore, it is possible that the MOs used in this study were not effective for knocking down the medaka *Hox* PG2 gene expression in PA2 after 72 hpf. In contrast to the medaka PA2 development, in zebrafish the PA2-derived bony elements begin to form at 53-57 hpf (Schilling and Kimmel, 1997). Likewise, in tilapia the PA2-derived cartilages begin to chondrify at 60 hpf (Le Pabic et al., 2009). Therefore, it is possible that the homeotic transformations in PA2 of zebrafish and tilapia caused by MOs targeting the 5'-UTRs and SJs of the *Hox* PG2 genes in these species were facilitated, in part, by a faster rate of development of the PA2-derived cartilages when compared to that of medaka (Hunter and

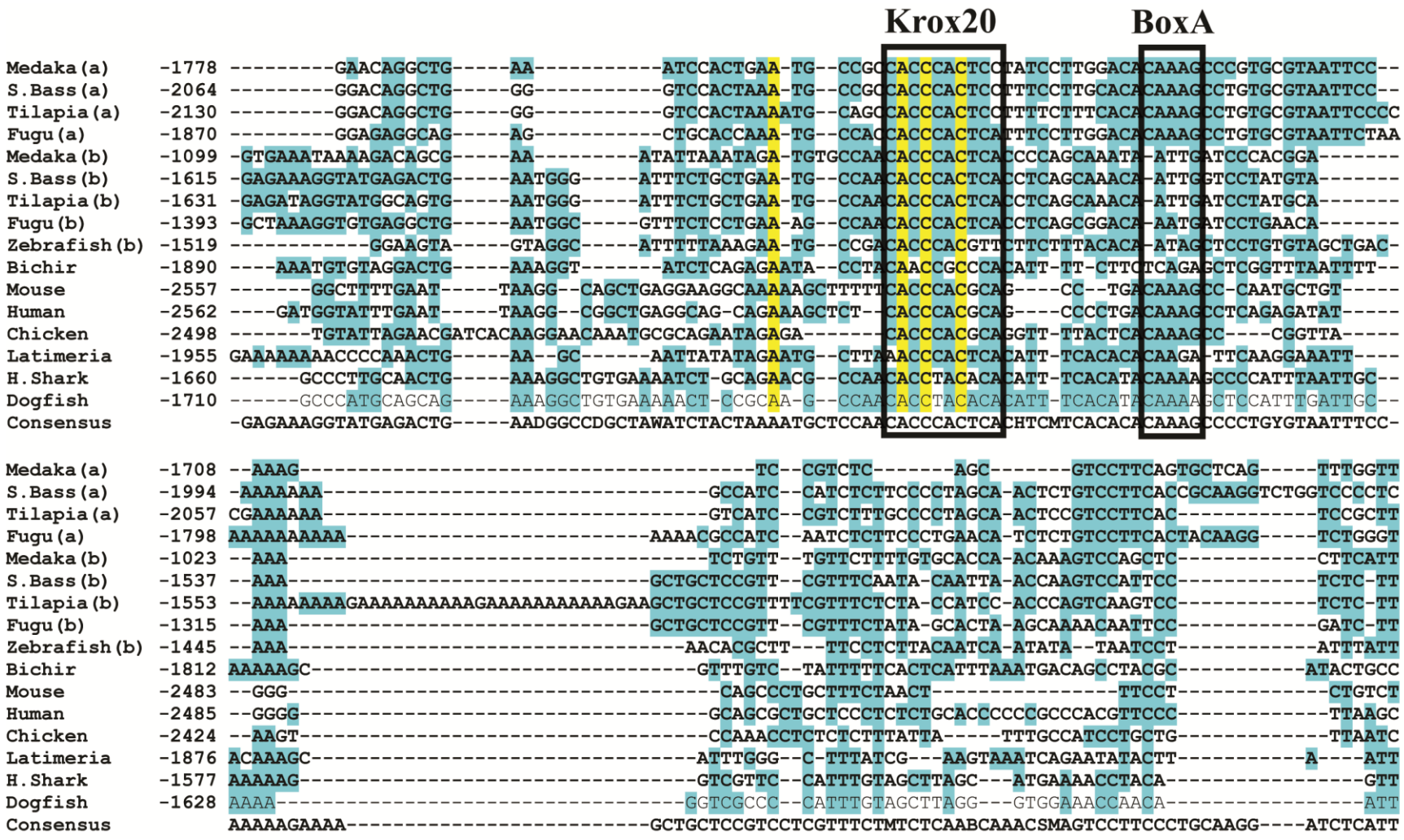
Prince, 2002; Le Pabic et al., 2010). Future studies involving photoresponsive MOs may help to facilitate spatiotemporal gene knockdown of medaka *Hox* PG2 genes in PA2 at the chondrogenic phase of development. These morpholinos remain inactive until they are exposed to a short dose of ultraviolet light (Shestopalov et al., 2007; Tomasini et al., 2009).

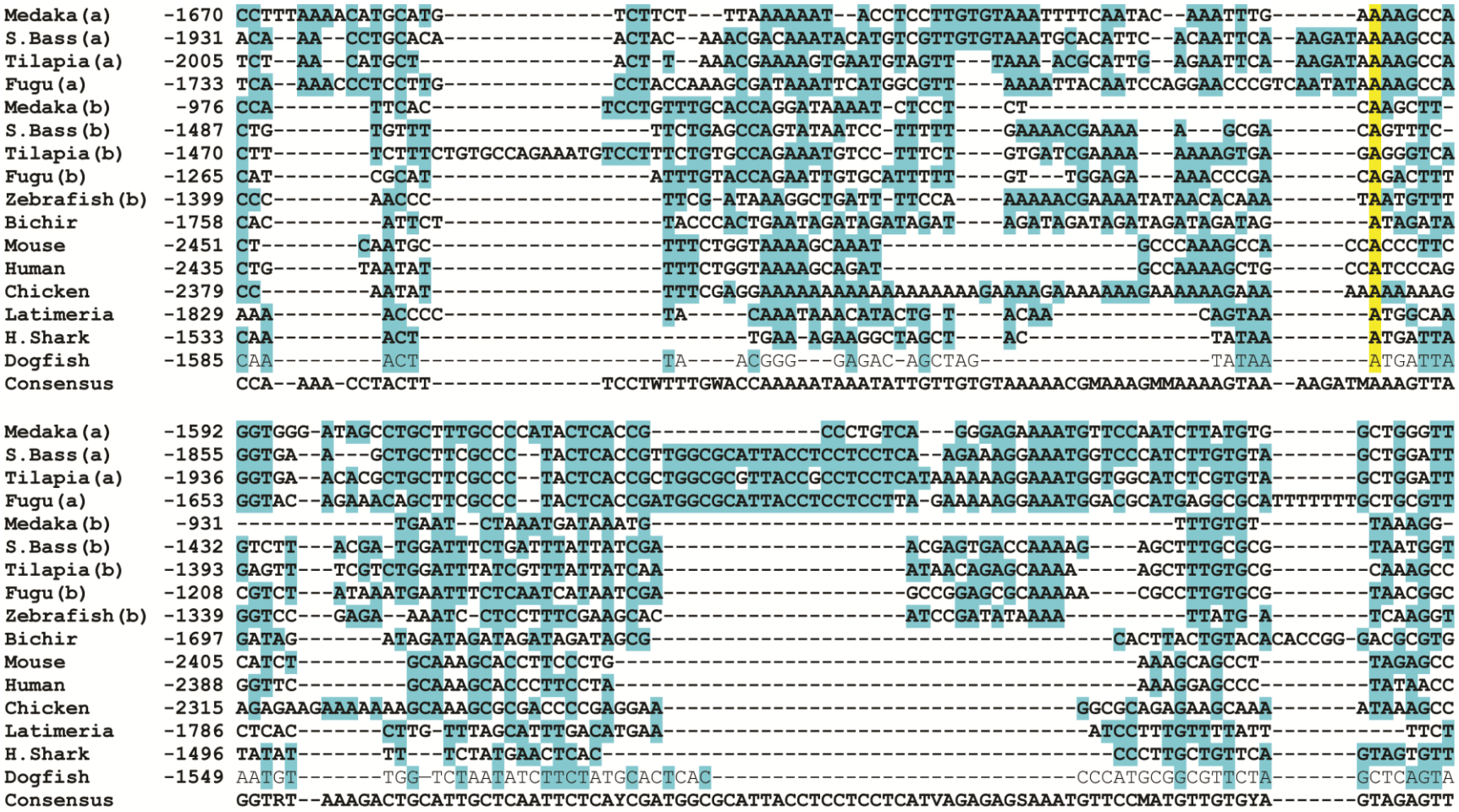
APPENDIX C: FREQUENCY RESULTS OF STABLE-LINE TRANSGENIC MEDAKA EMBRYOS

Table AC-1: Frequency results of stable-line transgenic medaka embryos.

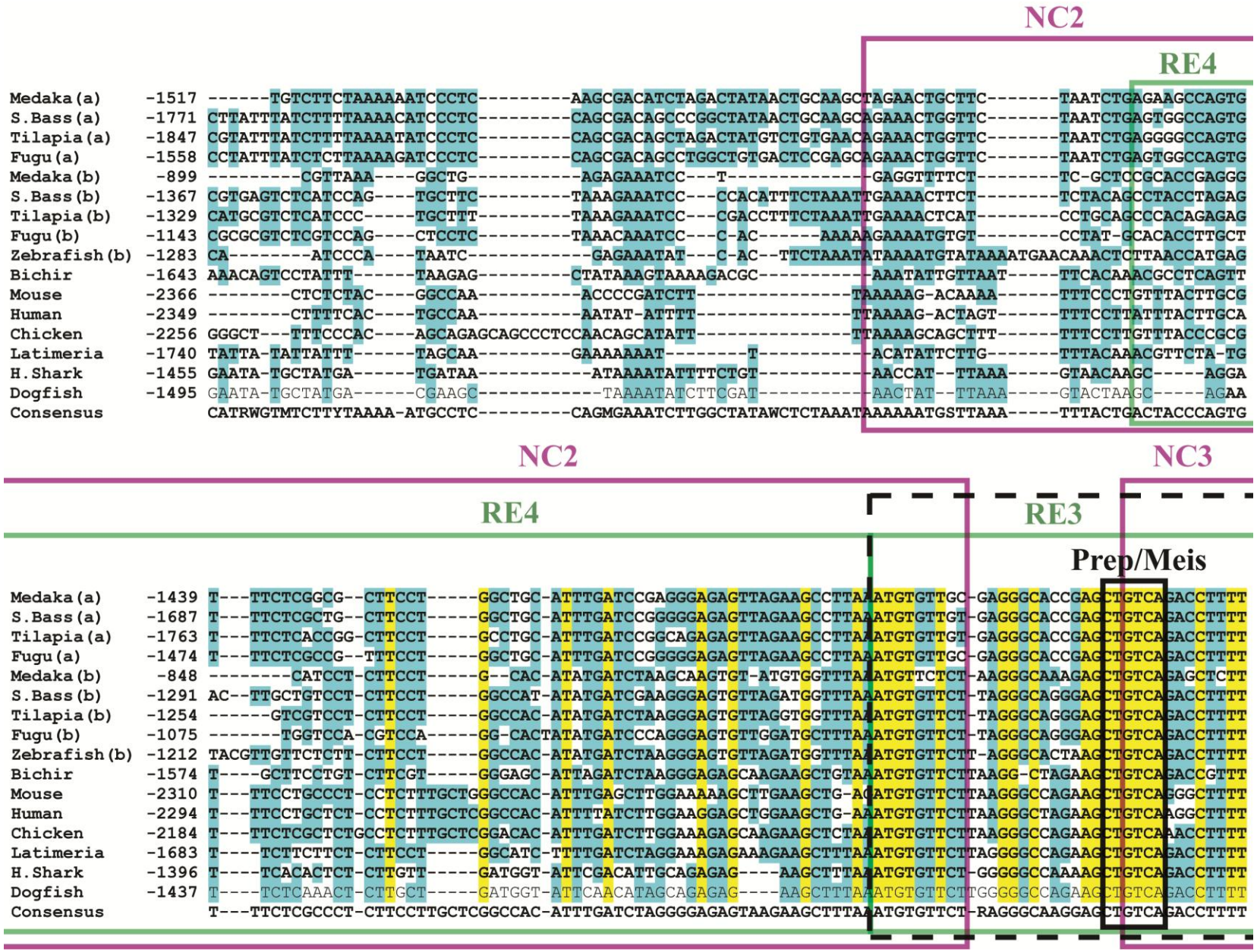
Construct	Construct Schematic	Frequency of Positive eGFP expressing embryos	Number of Female Survivors to adulthood	Number of eGFP-positive adult Females	Frequency of eGFP-expressing F2 embryos
		Medaka *hoxa2a* r3/5 ER Constructs			
1		42/48 (87.5%)	6	3	22/60 (37%)
2		64/84 (76%)	5	3	34/54 (53%)
3		41/49 (84%)	3	0	N/A
4		7/47 (15%)*	0	N/A	N/A
5		42/50 (84%)	5	4	30/82 (37%)
6		0/52 (0%)	N/A	N/A	N/A
7		52/62 (84%)	8	4	31/70 (44%)
		Medaka *ψhoxa2b* r3/5 ER Constructs			
8		46/51 (90%)	3	2	24/72 (33%)
9		47/52 (90%)	6	4	33/60 (55%)
10		33/38 (87%)	7	3	24/80 (30%)
11		27/52 (52%)*	0	N/A	N/A
12		33/41 (80%)	3	3	32/77 (42%)
13		15/64 (23%)*	0	N/A	N/A

APPENDIX D: SEQUENCE ALIGNMENT OF THE VERTEBRATE *HOXA2* R3/5 ENHANCER REGION




Medaka(a) -1670 CCTTTAAAACATGCATG----------TCTTCT---TTAAAAAAT--ACCTCCTTGTGTAAATTTTCAATAC--AAATTTG-------AAAAGCCA
S.Bass(a) -1931 ACA--AA--CCTGCACA----------ACTAC--AAACGACAAATACATGTCGTTGTGTAAATGCACATTC--ACAATTCA--AAGATAAAAGCCA
Tilapia(a) -2005 TCT--AA--CATGCT----------ACT-T--AAACGAAAAGTGAATGTAGTT---TAAA-ACGCATTG--AGAATTCA--AAGATAAAAGCCA
Fugu(a) -1733 TCA--AAACCCTCCTTG----------CCTACCAAAGCGATAAATTCATGGCGTT----AAAATTACAATCCAGGAACCCGTCAATATAAAAGCCA
Medaka(b) -976 CCA--------TTCAC----------TCCTGTTTGCACCAGGATAAAAT-CTCCT----CT----------CAAGCTT-
S.Bass(b) -1487 CTG--------TGTTT----------TTCTGAGCCAGTATAATCC-TTTTT----GAAAACGAAAA---A---GCGA-------CAGTTTC-
Tilapia(b) -1470 CTT--------TCTTTCTGTGCCAGAAATGTCCTTTCTGTGCCAGAAATGTCC-TTTCT----GTGATCGAAAA---AAAAGTGA-------GAGGGTCA
Fugu(b) -1265 CAT--------CGCAT----------ATTTGTACCAGAATTGTGCATTTTT----GT---TGGAGA---AAACCCGA-------CAGACTTT
Zebrafish(b) -1399 CCC--------AACCC----------TTCG-ATAAAGGCTGATT-TTCCA----AAAAACGAAAATATAACACAAA-------TAATGTTT
Bichir -1758 CAC---------ATTCT----------TACCCACTGAATAGATAGATAGAT----AGATAGATAGATAGATAGATAG--------ATAGATA
Mouse -2451 CT--------CAATGC----------TTTCTGGTAAAAGCAAAT----------GCCCAAAGCCA------CCACCCTTC
Human -2435 CTG-------TAATAT----------TTTCTGGTAAAAGCAGAT----------GCCAAAAGCTG------CCATCCCAG
Chicken -2379 CC---------AATAT----------TTTCGAGGAAAAAAAAAAAAAAAAAGAAAAGAAAAAAAGAAAAAAGAAA------AAAAAAAAG
Latimeria -1829 AAA---------ACCCC----------TA-----CAAATAAACATACTG-T----ACAA----------CAGTAA--------ATGGCAA
H.Shark -1533 CAA---------ACT----------TGAA-AGAAGGCTAGCT----AC----------TATAA--------ATGATTA
Dogfish -1585 CAA---------ACT----------TA----ACGGG---GAGAC-AGCTAG----------TATAA--------ATGATTA
Consensus CCA--AAA-CCTACTT----------TCCTWTTTGWACCAAAAATAAATATTGTTGTGTAAAAACGMAAAGMMAAAAGTAA--AAGATMAAAGTTA

Medaka(a) -1592 GGTGGG-ATAGCCTGCTTTGCCCCATACTCACCG----------CCCTGTCA---GGGAGAAAATGTTCCAATCTTATGTG-------GCTGGGTT
S.Bass(a) -1855 GGTGA--A---GCTGCTTCGCCC--TACTCACCGTTGGCGCATTACCTCCTCCTCA--AGAAAGGAAATGGTCCCATCTTGTGTA-------GCTGGATT
Tilapia(a) -1936 GGTGA--ACACGCTGCTTCGCCC--TACTCACCGCTGGCGCGTTACCGCCTCCTCATAAAAAAGGAAATGGTGGCATCTCGTGTA-------GCTGGATT
Fugu(a) -1653 GGTAC--AGAAACAGCTTCGCCC--TACTCACCGATGGCGCATTACCTCCTCCTTA-GAAAAAGGAAATGGACGCATGAGGCGCATTTTTTTGCTGCGTT
Medaka(b) -931 ----------TGAAT--CTAAATGATAAATG----------TTTGTGT---------TAAAGG-
S.Bass(b) -1432 GTCTT---ACGA-TGGATTTCTGATTTATTATCGA----------ACGAGTGACCAAAAG----AGCTTTGCGCG--------TAATGGT
Tilapia(b) -1393 GAGTT---TCGTCTGGATTTATCGTTTATTATCAA----------ATAACAGAGCAAAA-----AGCTTTGTGCG--------CAAAGCC
Fugu(b) -1208 CGTCT---ATAAATGAATTTCTCAATCATAATCGA----------GCCGGAGCGCAAAAA----CGCCTTGTGCG--------TAACGGC
Zebrafish(b) -1339 GGTCC---GAGA--AAATC-CTCCTTTCGAAGCAC----------ATCCGATATAAAA----------TTATG-A--------TCAAGGT
Bichir -1697 GATAG-------ATAGATAGATAGATAGATAGCG----------CACTTACTGTACACACCGG-GACGCGTG
Mouse -2405 CATCT---------GCAAAGCACCTTCCCTG----------AAAGCAGCCT---------TAGAGCC
Human -2388 GGTTC---------GCAAAGCACCCTTCCTA----------AAAGGAGCCC---------TATAACC
Chicken -2315 AGAGAAGAAAAAAAGCAAAGCGCGACCCCGAGGAA----------GGCGCAGAGAAGCAAA-------ATAAAGCC
Latimeria -1786 CTCAC-------CTTG-TTTAGCATTTGACATGAA----------ATCCTTTGTTTTATT----------TTCT
H.Shark -1496 TATAT-------TT---TCTATGAACTCAC----------CCCTTGCTGTTCA-------GTAGTGTT
Dogfish -1549 AATGT-------TGG-TCTAATATCTTCTATGCACTCAC----------CCCATGCGGCGTTCTA-------GCTCAGTA
Consensus GGTRT--AAAGACTGCATTGCTCAATTCTCAYCGATGGCGCATTACCTCCTCCTCATVAGAGAGSAAATGTTCCMATGTTGTGYA-------GTAGAGTT


NC2
RE4
NC2
NC3
RE4
RE3
Prep/Meis

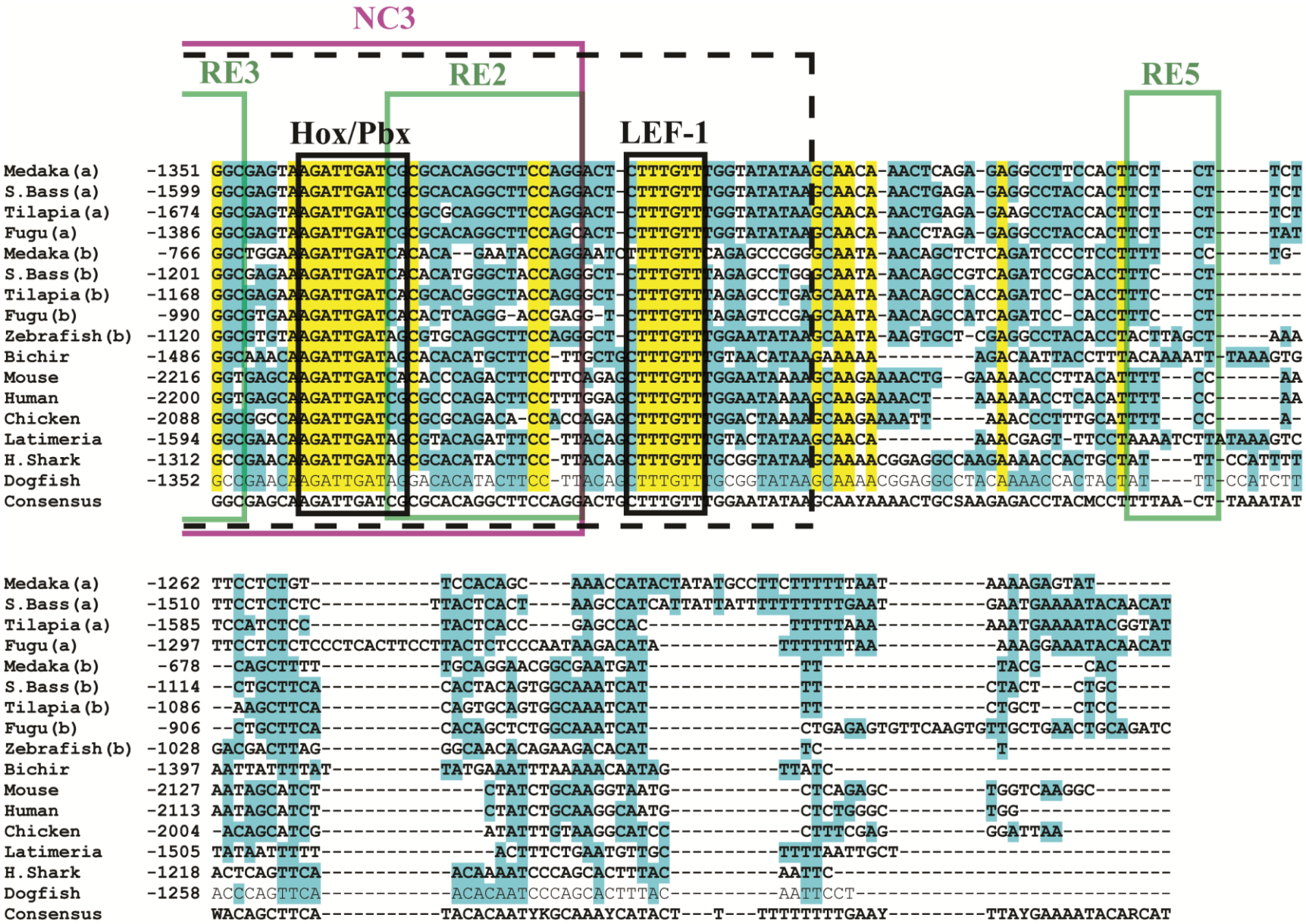

NC3
RE3
RE2
RE5
Hox/Pbx
LEF-1
Medaka(a) -1351 GGCGAGTAAGATTGATCGCGCACAGGCTTCCAGGACT-CTTTGTTTGGTATATAAGCAACA-AACTCAGA-GAGGCCTTCCACTTCT---CT-----TCT
S.Bass(a) -1599 GGCGAGTAAGATTGATCGCGCACAGGCTTCCAGGACT-CTTTGTTTGGTATATAAGCAACA-AACTGAGA-GAGGCCTACCACTTCT---CT-----TCT
Tilapia(a) -1674 GGCGAGTAAGATTGATCGCGCGCAGGCTTCCAGGACT-CTTTGTTTGGTATATAAGCAACA-AACTGAGA-GAAGCCTACCACTTCT---CT-----TCT
Fugu(a) -1386 GGCGAGTAAGATTGATCGCGCACAGGCTTCCAGCACT-CTTTGTTTGGTATATAAGCAACA-AACCTAGA-GAGGCCTACCACTTCT---CT-----TAT
Medaka(b) -766 GGCTGGAAAGATTGATCACACA--GAATACCAGGAATCTTTTGTTCAGAGCCCGGGCAATA-AACAGCTCTCAGATCCCCTCCTTTT---CC-----TG-
S.Bass(b) -1201 GGCGAGAAAGATTGATCACACATGGGCTACCAGGGCT-CTTTGTTTAGAGCCTGGGCAATA-AACAGCCGTCAGATCCGCACCTTTC---CT--------
Tilapia(b) -1168 GGCGAGAAAGATTGATCACGCACGGGCTACCAGGGCT-CTTTGTTTAGAGCCTGAGCAATA-AACAGCCACCAGATCC-CACCTTTC---CT--------
Fugu(b) -990 GGCGTGAAAGATTGATCACACTCAGGG-ACCGAGG-T-CTTTGTTTAGAGTCCGAGCAATA-AACAGCCATCAGATCC-CACCTTTC---CT--------
Zebrafish(b) -1120 GGCGTGTAAGATTGATAGCGTGCAGGCTTCCAGGGCT-CTTTGTTTGGAATATAAGCAATA-AAGTGCT-CGAGGCCTACACCTACTTAGCT-----AAA
Bichir -1486 GGCAAACAAGATTGATAGCACACATGCTTCC-TTGCTGCTTTGTTTGTAACATAAGAAAAA---------AGACAATTACCTTTACAAAATT-TAAAGTG
Mouse -2216 GGTGAGCAAGATTGATCACACCCAGACTTCCTTCAGAGCTTTGTTTGGAATAAAAGCAAGAAAACTG--GAAAAACCCTTACATTTT---CC------AA
Human -2200 GGTGAGCAAGATTGATCGCGCCCAGACTTCCTTTGGAGCTTTGTTTGGAATAAAAGCAAGAAAACT----AAAAAACCTCACATTTT---CC------AA
Chicken -2088 GGCGGCCAAGATTGATCGCGCGCAGACA-CCACCAGAGCTTTGTTTGGACTAAAAGCAAGAAAATT-----AAACCCTTTGCATTTT---CC------A-
Latimeria -1594 GGCGAACAAGATTGATAGCGTACAGATTTCC-TTACAGCTTTGTTTGTACTATAAGCAACA---------AAACGAGT-TTCCTAAAATCTTATAAAGTC
H.Shark -1312 GCCGAACAAGATTGATAGCGCACATACTTCC-TTACAGCTTTGTTTGCGGTATAAGCAAAACGGAGGCCAAGAAAACCACTGCTAT----TT-CCATTTT
Dogfish -1352 GCCGAACAAGATTGATAGGACACATACTTCC-TTACAGCTTTGTTTGCGGTATAAGCAAAACGGAGGCCTACAAAACCACTACTAT----TT-CCATCTT
Consensus GGCGAGCAAGATTGATCGCGCACAGGCTTCCAGGACTGCTTTGTTTGGAATATAAGCAAYAAAACTGCSAAGAGACCTACMCCTTTTAA-CT-TAAATAT
Medaka(a) -1262 TTCCTCTGT----------TCCACAGC----AAACCATACTATATGCCTTCTTTTTTAAT---------AAAAGAGTAT-------
S.Bass(a) -1510 TTCCTCTCTC---------TTACTCACT----AAGCCATCATTATTATTTTTTTTTGAAT---------GAATGAAAATACAACAT
Tilapia(a) -1585 TCCATCTCC----------TACTCACC----GAGCCAC----------TTTTTAAA---------AAATGAAAATACGGTAT
Fugu(a) -1297 TTCCTCTCTCCCTCACTTCCTTACTCTCCCAATAAGACATA----------TTTTTTTAA----------AAAGGAAATACAACAT
Medaka(b) -678 --CAGCTTTT----------TGCAGGAACGGCGAATGAT----------TT--------------TACG----CAC-----
S.Bass(b) -1114 --CTGCTTCA----------CACTACAGTGGCAAATCAT----------TT--------------CTACT---CTGC-----
Tilapia(b) -1086 --AAGCTTCA----------CAGTGCAGTGGCAAATCAT----------TT--------------CTGCT---CTCC-----
Fugu(b) -906 --CTGCTTCA----------CACAGCTCTGGCAAATCAT----------CTGAGAGTGTTCAAGTGTTGCTGAACTGCAGATC
Zebrafish(b) -1028 GACGACTTAG----------GGCAACACAGAAGACACAT----------TC--------------T----------
Bichir -1397 AATTATTTTAT---------TATGAAATTTAAAAACAATAG---------TTATC----------
Mouse -2127 AATAGCATCT----------CTATCTGCAAGGTAATG----------CTCAGAGC---------TGGTCAAGGC-------
Human -2113 AATAGCATCT----------CTATCTGCAAGGCAATG----------CTCTGGGC---------TGG----------
Chicken -2004 -ACAGCATCG----------ATATTTGTAAGGCATCC----------CTTTCGAG---------GGATTAA---------
Latimeria -1505 TATAATTTTT----------ACTTTCTGAATGTTGC---------TTTTAATTGCT----------
H.Shark -1218 ACTCAGTTCA----------ACAAAATCCCAGCACTTTAC---------AATTC----------
Dogfish -1258 ACCCAGTTCA----------ACACAATCCCAGCACTTTAC---------AATTCCT----------
Consensus WACAGCTTCA----------TACACAATYKGCAAAYCATACT---T---TTTTTTTTGAAY---------TTAYGAAAATACARCAT

Fig. AD-1: Comparative genomic sequence analysis of the vertebrate *Hoxa2* r3/5 enhancer region. Teleost *hoxa2a* sequences are denoted by a in parentheses. Teleost *hoxa2b* sequences are denoted by b in parentheses. Numbers correspond to genomic base pair positions relative to the ATG start site of the *Hoxa2* genes. Base pairs colored in yellow correspond to complete conservation at particular sites across all sequences examined. Base pairs colored in blue represent the majority of the sequences containing specific base pairs at specific sites. A consensus sequence was derived from the aligned sequences. Boxed regions labeled as Krox20, BoxA, NC2, NC3, RE4, RE3, RE2 and RE5 correspond to functionally tested sites in Maconochie et al. (1999 and 2001) and Tümpel et al. (2006). Dashed boxed region corresponds to the medaka *hoxa2a* r4/CNCC and *ψhoxa2b* r3-7/CNCC elements found in this study. Boxed regions labeled as Prep/Meis, Hox/Pbx and LEF-1 correspond to transcription factor binding sites identified in this study.