

Article

Functional and Comparative Genomics of *Hoxa2* Gene *cis*-Regulatory Elements: Evidence for Evolutionary Modification of Ancestral Core Element Activity

Adam Davis ^{1,*}, Michael C. Reubens ² and Edmund J. Stellwag ³¹ Department of Biology and Physical Sciences, Gordon State College, Barnesville, GA 30204, USA² The Scripps Research Institute, 10550 N. Torrey Pines Road, MB3, La Jolla, CA 92037, USA; michaelreubens82@gmail.com³ Department of Biology, Howell Science Complex, East Carolina University, Greenville, NC 27858, USA; stellwage@ecu.edu

* Correspondence: davisa0314@gmail.com; Tel.: +1-484-678-9582

Academic Editor: Vincenzo Zappavigna

Received: 16 February 2016; Accepted: 17 March 2016; Published: 26 March 2016

Abstract: *Hoxa2* is an evolutionarily conserved developmental regulatory gene that functions to specify rhombomere (r) and pharyngeal arch (PA) identities throughout the Osteichthyes. Japanese medaka (*Oryzias latipes*) *hoxa2a*, like orthologous *Hoxa2* genes from other osteichthyans, is expressed during embryogenesis in r2–7 and PA2–7, whereas the paralogous medaka pseudogene, *ψhoxa2b*, is expressed in noncanonical *Hoxa2* domains, including the pectoral fin buds. To understand the evolution of *cis*-regulatory element (CRE) control of gene expression, we conducted eGFP reporter gene expression studies with extensive functional mapping of several conserved CREs upstream of medaka *hoxa2a* and *ψhoxa2b* in transient and stable-line transgenic medaka embryos. The CREs tested were previously shown to contribute to directing mouse *Hoxa2* gene expression in r3, r5, and PA2–4. Our results reveal the presence of sequence elements embedded in the medaka *hoxa2a* and *ψhoxa2b* upstream enhancer regions (UERs) that mediate expression in r4 and the PAs (*hoxa2a* r4/CNCC element) or in r3–7 and the PAs *ψhoxa2b* r3–7/CNCC element), respectively. Further, these elements were shown to be highly conserved among osteichthyans, which suggests that the r4 specifying element embedded in the UER of *Hoxa2* is a deeply rooted rhombomere specifying element and the activity of this element has been modified by the evolution of flanking sequences that redirect its activity to alternative developmental compartments.

Keywords: Japanese medaka; *Hox* PG2 gene expression; *hoxa2a*; *ψhoxa2b*; embryonic development; *cis*-regulatory elements; vertebrate evolution; gene regulation

1. 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 [1]. *Hox* clusters comprise up to 14 genes, which are expressed along the A-P axis during embryonic development collinear with their physical location within a cluster [2–4]. 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 [5–11]. Post-genome duplication independent gene loss has generated clustered paralog groups (PGs) that differ in gene number depending on the historical timing of gene losses relative to genome duplications [6,12–15].

In addition to variation in gene number within and among paralog groups, results from expression and functional genetic studies have shown that duplicate *Hox* genes can exhibit either similar or heterogeneous 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 [6,12–21]. 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, respectively [22,23]. *Cis*-regulatory elements (CREs) that direct mouse *Hoxa2* expression in r2, r4 and r3, r5 and the CNCCs are located in exon 2 of *Hoxa2*, the intron and exon 1 of *Hoxa2*, and the *Hoxa3-a2* intergenic region, respectively [21,24–34]. There has been extensive evolutionary conservation of the regulatory circuitry that directs the hindbrain and PA patterning activity of *Hoxa2*, as similar gene regulation was observed for *Hox2* of lamprey (*Petromyzon marinus*) [35]. By contrast, mouse *Hoxb2* is expressed in r3–8 of the hindbrain but not in the CNCCs [36]. The CREs that direct mouse *Hoxb2* in r3, r4 and r5 are located in the mouse *Hoxb3-b2* intergenic region [36–38]. The variation in mouse *Hoxa2* and *b2* regulation and expression 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 trigeminal (Vth) and facial (VIIth) cranial motor nerve axons out of r2 and r4, respectively, while *Hoxb2* controls the specification of the somatic motor component of the VIIth cranial nerve exiting r4 [39–44]. Further, *Hoxa2*, but not *Hoxb2*, is involved in patterning the craniofacial elements arising from PA2 and the posterior arches in tetrapods [22,23,45–49].

Phylogenetic reconstructions that include a whole genome duplication event at the incipient stage of teleost evolution, which occurred 350–220 million years ago [7,50–54], support a post-genome duplication ancestral *Hox* PG2 gene complement consisting of two *Hoxa2* genes, *hoxa2a* and *a2b* (Figure 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 [5]. Results based on cloning and expression analyses of several acanthopterygians, including striped bass (*Morone saxatilis*), Nile tilapia (*Oreochromis niloticus*), fugu (*Takifugu rubripes*) and Japanese medaka (*Oryzias latipes*), have shown that while striped bass, Nile tilapia and fugu have two functional *Hoxa2* genes, *hoxa2a* and *a2b*, medaka has just one, *hoxa2a* [12,14,19]. The paralog of medaka *hoxa2a* was shown to be a transcribed pseudogene (*ψhoxa2b*), wherein several premature stop codons were discovered in the region corresponding to exon 2 upstream of the putative homeodomain [12]. However, the amino terminal 138 amino acid sequence was shown to be intact and contained a conserved hexapeptide motif, indicative that transcripts arising from this sequence could be translated into a product with activities like hexapeptide motif-related Pbx binding [55]. Further, while the *hoxa2a* and *a2b* genes of most teleosts are expressed in a conserved manner in much of the hindbrain (r2–5), PA2 and the posterior arches similar to orthologous genes in tetrapods [6,12,14,19,28,30,45,49,56,57], the medaka *ψhoxa2b* was shown to be 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 [12]. These noncanonical expression patterns may be a reflection of sequence changes resulting from relaxed selection within the medaka *ψhoxa2b* CREs following the mutational inactivation of the homeodomain but not the hexapeptide region of the medaka *hoxa2b* coding sequence [12].

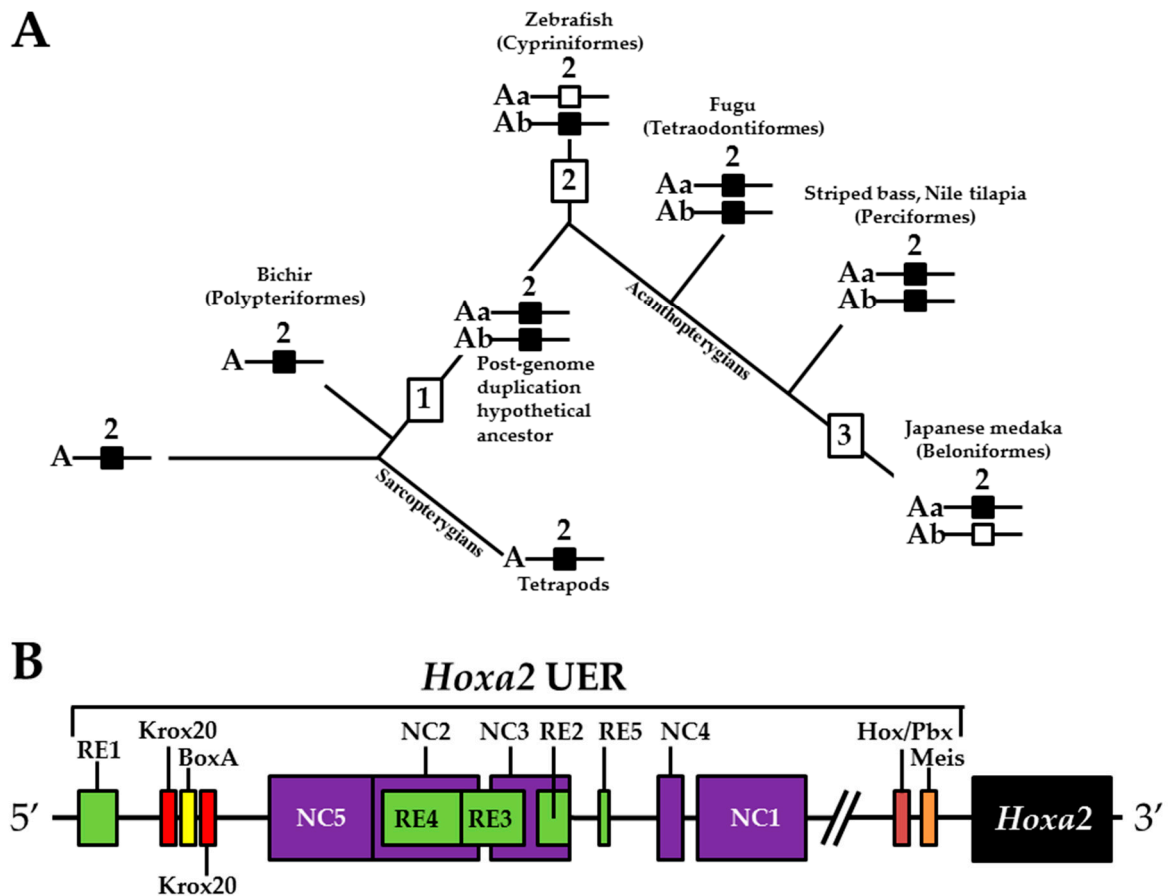


Figure 1. Evolution of *Hoxa2* gene complement and *Hoxa2* UER composition. (A) Phylogeny based on Steinke *et al.* (2006) [58]: (1) Genome duplication; (2) *hoxa2a* gene loss; and (3) *hoxa2b* gene loss. Filled in boxes represent functional *Hoxa2* genes. Open boxes represent loss of *Hoxa2* gene function. (B) Genomic map of the *Hoxa2* UER characterized in mouse. The genomic map was not drawn to scale. RE, rhombomeric element; NC, neural crest.

The *Hoxa3/Hoxa2* intergenic region is the most extensively studied functional genomic region that directs *Hoxa2* gene expression. Results from comparative and functional genomic analyses in mouse and chick embryos have identified a highly conserved upstream enhancer region (UER) that is responsible for directing *Hoxa2* in r3 and r5 of the hindbrain and the CNCCs in the PA2 and the posterior arches (Figure 1B) [21,24,25,27–32]. Identified rhombomeric CREs within this region include binding sites for Krox20, a protein shown to be integral for hindbrain development [21,30,31,34,59]. Other rhombomeric elements, termed RE1–5 and BoxA, have been shown to function in conjunction with Krox20 in potentiating *Hoxa2* expression in r3 and r5 across diverse taxa [21,25,28,30–32], which is indicative that these elements may represent a regulatory kernel subcircuit, *sensu* Davidson and Erwin (2006) [60]. Five functional CREs that direct *Hoxa2* expression in the CNCCs populating the PA2 and posterior arches that are referred to as neural crest (NC) 1–5 have been identified in the mouse [27,29]. One of these elements, NC4, encodes a binding site for AP-2, a protein that is crucial for CNCC survival and chondrocyte differentiation within the PAs [27,61]. A recent analysis provides evidence for Hox/Pbx and Meis protein binding elements within the proximal promoter region of mouse *Hoxa2* [24]. The proteins that bind these elements have been shown to be crucial for craniofacial skeleton morphogenesis, [22–24,27,45–49,56,62–66].

The present study was undertaken to examine the activity of the medaka *hoxa2a* and ψ *hoxa2b* UERs in their natural background, and to understand the relationship between sequence differences of the two medaka UERs relative to the striking differences in the expression of their cognate genes, one

of which, *ψhoxa2b*, is expressed in development well outside the compartments in which expression is observed for any other osteichthyan *Hox* PG2 gene [12]. Most of the CREs mentioned above have recently been shown using comparative and functional genomic analyses to be present within the *hoxa3a-a2a* and *hoxa9b-a2b* intergenic regions of teleost fishes, except for the NC4 and NC1 elements [29]. At the time that the present study was conducted for the CREs upstream of medaka *hoxa2a* and *ψhoxa2b*, only a subset of *Hoxa2* UER-specific CREs were identified within the teleost *hoxa3a-a2a* and *hoxa9b-a2b* intergenic regions, and included, 5' to 3', Krox20, BoxA, RE4, RE3, RE2 and RE5 [21], hereafter referred to as the UER(K20-RE5). This sequence also included NC5, NC2, and NC3 sequences, which overlap many of the RE sequences mentioned above. Further, when the medaka *hoxa2a* UER(K20-RE5) was tested for its function in chick embryos through electroporation, no reporter gene expression was observed in the hindbrain or PAs [21]. However, the paralogous UER(K20-RE5) of medaka *ψhoxa2b* yielded reporter gene staining in r3 and r5 when it was electroporated into chick embryos [21]. These results were inconsistent with *in situ* hybridization results for medaka *hoxa2a*, which showed that this gene is expressed in r2–7 of the hindbrain, PA2 and the posterior PAs, and *ψhoxa2b*, which was shown to be expressed within noncanonical *Hox* PG2 expression domains, including the pectoral fin buds [12]. Therefore, consistent with the results of *in situ* hybridization, it was hypothesized that the medaka *hoxa2a* UER(K20-RE5) region would direct reporter gene expression in r3, r5, PA2 and the posterior PAs when expressed in medaka. Similarly, it was hypothesized that the paralogous region cloned from medaka *ψhoxa2b* would direct expression of the reporter gene in the noncanonical domains mentioned above. Contrary to expectations, reporter gene expression analyses showed that the medaka *hoxa2a* and *ψhoxa2b* UER(K20-RE5)s directed reporter gene expression in r4 and r3–7, respectively, as well as in the migratory CNCCs of the hyoid and post-otic streams and the post-migratory CNCCs in PA2 and the posterior arches. These results suggest that the functional nature of the UER(K20-RE5) has diverged between medaka *hoxa2a* and *ψhoxa2b*. Deletion mapping also revealed a short 88/89 base pair (bp) region within the medaka *hoxa2a* and *ψhoxa2b* UER(K20-RE5)s that includes consensus Hox/Pbx and Prep/Meis sites that are known to direct expression in r4 for mouse *Hoxb1*, *a2*, and *b2* and the CNCCs in mouse *Hoxa2* [21,24,26,27,33,37,67,68] and likely represents a deeply rooted core CRE element. This core CRE element is present in the regulatory sequences common to gnathostome *Hox* A clusters and its effects on expression of cognate *Hoxa2* genes appears to be influenced by interactions with cis-elements outside and sequence variation within the core element itself.

2. Materials and Methods

2.1. Tol2 Plasmid Construction

Transient and stable-line transgenic analyses employed the pT2AL200L150G plasmid vector for transmitting transposon insertions into medaka embryos (generous gift from Koichi Kawakami) [69]. 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 pT2AL200R150G 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 [70]. The microinjection procedure used for medaka zygotes is described below.

In order to analyze the function of the CREs in control of the medaka *hoxa2a* and *ψhoxa2b* UER(K20-RE5)s in RNA expression, the full length *EFI-αS* promoter of the pT2AL200L150G was truncated to include only the region encompassing the TATA box so as to diminish the amount of transcription occurring from the *Xenopus* promoter and maximize transcription resulting from cloned sequences of putative *cis*-regulatory elements. The TATA box from the

full length *EFl α -S* promoter was amplified from pT2AL200R150G using the forward primer 5'-GATAGGGATCCAGTTCTCAGGATCGGTTCG-3' (containing the underlined BamHI sequence) and the reverse primer 5'-GGGGGCTCGAGTATATAAAGGGTGGTTAAGG-3' (containing the underlined XhoI sequence) using standard PCR procedures. To generate pTolTATA- β MCS the *EFl α -S* promoter and *β -globin* intron were removed from pT2AL200R150G using the flanking XhoI and BamHI sites, and the amplified TATA box containing sequence was then ligated into the pT2AL200R150G digested backbone using the compatible BamHI and XhoI termini. To facilitate the ligation of potential enhancer sequences upstream of the resulting *EFl α -S* TATA box, a polylinker containing restriction sites for SmaI, EcoRI, PmeI, and EcoRV was cloned into the 5' XhoI site. The pTolTATA- β MCS vector was also co-microinjected with and without transposon mRNA to test the efficacy of this vector system in the medaka model.

2.2. 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, Franklin Lakes, NJ, USA). 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-Aldrich, St. Louis, MO, USA) was added to the cell homogenate and the tube was incubated for 30 min at 37 °C. The homogenate was then treated with 100 μ L of protease K (10 mg/mL) (Life Technologies, Carlsbad, CA, USA) and incubated at 55 °C for 6 h. The RNase A and protease K-treated homogenate was transferred to 1.5 mL centrifuge tubes (Fisher Scientific, Pittsburgh, PA, USA) and centrifuged for 30 s at 14,000 rpm using an Eppendorf 5514 C centrifuge (Hauppauge, NY, USA). The resulting supernatant was then divided into two 500 μ L 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 μ L of PCI (Life Technologies, Carlsbad, CA, USA) was added to each 1.5 mL tube and mixed until a homogeneous emulsion formed. Microcentrifuge tubes containing the emulsion were centrifuged for 30 s 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, Brookfield, CT, USA). 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 \times g for 1 min. 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 μ L of Nuclease Free Water (Life Technologies, Carlsbad, CA, USA). 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 (Life Technologies, Carlsbad, CA, USA).

2.3. Amplification of Medaka *hoxa2a* and *ψ hoxa2b* UER(K20-RE5)s

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 *ψ hoxa2b* 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 *ψ hoxa2b* are listed in Table 1. All primers for all analyses were developed using published genomic sequences of the Japanese medaka [71]. The PCR products generated from long-range PCR were cloned into pCR II vectors (Life Technologies, Carlsbad, CA, USA) 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 UER(K20-RE5)s of medaka *hoxa2a* and *ψhoxa2b*, which were tested for enhancer activity in the pTolTATA-βMCS plasmid vector.

Table 1. Primers used for the amplification of the medaka *hoxa2a* and *ψhoxa2b* UER(K20-RE5)s and the exclusion of sequence elements for the functional testing of these regulatory regions. The 5'-start site for primers pertains to the nucleotide base position with respect to the ATG start site of the downstream gene (*hoxa2a*) or pseudogene (*ψhoxa2b*).

Primer	Sequence 5' to 3'	5' Start Site
Medaka <i>hoxa2a</i> Genomic Primers		
A2a For	TTATTCCCACAACCCTTTCATTTCG	−2691
A2a Rev	CACACTCAGCCACAATCTCTTCTTC	1846
Medaka <i>ψhoxa2b</i> Genomic Primers		
A2b For	ACACAGCAGGGGTCAACAATAGGTC	−3093
A2b Rev	ATAGGCAGAGCAGAAAACAAAATG	3193
Medaka <i>hoxa2a</i> UER(K20-RE5) Forward Primers		
AF1	GATCGATATCGAACAGGCTGAAATCCACTGAATGC	−1778
AF2	GATCGATATCGCTTCTAATCTGAGAAGCCAGTGTTTC	−1468
AF3	GATCGATATCATGTGTTGCGAGGGCACCGAGCTGTC	−1392
AF4	GATCGATATCGAGTAAGATTGATCGCGCACAGGCTTC	−1354
Medaka <i>hoxa2a</i> UER(K20-RE5) Reverse Primers		
AR1	GATCGAATTCGTTTGCTGTGGAACAGAGGAAAGAAG	−1247
AR2	GATCGAATTCCTATATAACCAAACAAAGAGTCCTGG	−1303
AR3	GATCGAATTCCTACTCGCCAAAAGGTCTGACAGCTC	−1348
Medaka <i>ψhoxa2b</i> UER(K20-RE5) Forward Primers		
BF1	GATCGATATCATGTGCCAACACCCACTCACCCAG	−1068
BF2	GATCGATATCCTTTCGCTCCGCACCGAGGGCATCCTC	−868
BF3	GATCGATATCATGTCTCTAAGGGCAAAGAGCTGTC	−803
BF4	GATCGATATCTGAAAAGATTGATCACACAGAATACC	−765
Medaka <i>ψhoxa2b</i> UER(K20-RE5) Reverse Primers		
BR1	GATCGAATTCAAAAAGCTGCAGGAAAAGGAGGGGATC	−671
BR2	GATCGAATTCCTCCGGGCTCTGAACAAAAGATTCTCTG	−715
BR3	GATCGAATTCCTTCCAGCCAAGAGCTCTGACAGCTC	−759

2.4. PCR-Mediated Deletion Mutagenesis of the Medaka *hoxa2a* and *ψhoxa2b* UER(K20-RE5)s

DNA sequences that contained the UER(K20-RE5)s of medaka *hoxa2a* and *ψhoxa2b* were amplified using PCR primers listed in Table 1. These primers were used to generate a product referred to as the UER(K20-RE5), which encompassed sequence elements corresponding to Krox20, BoxA, RE4, RE3, RE2, RE5, NC2, NC3 and NC5, and was similar to the region tested in chicken embryos [21], and a series of nested deletions that were derived from the UER(K20-RE5). Specific primer pairs and their amplified genomic DNA products are listed in Table 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 green fluorescent protein (GFP) sequences located downstream in the pTolTATA-βMCS vector. PCR products were purified using the QIAquick PCR purification kit (Qiagen, Valencia, CA, USA) 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, USA). The pTolTATA-βMCS plasmid vector was pre-digested with EcoRI and EcoRV, purified using the QIAquick PCR sequence kit (Qiagen, Valencia, CA, USA) and treated with shrimp alkaline phosphatase (New England Biolabs, Ipswich, MA, USA) 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 *Escherichia coli* (*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, St. Louis, MO, USA) according to the manufacturer’s instructions. Further confirmation of cloned PCR products in the pTotTATA-βMCS plasmid vector was performed using DNA sequencing of the inserts by dideoxyterminator sequencing chemistry (Big Dye v. 3.0, Applied Biosystems, Foster City, CA, USA) on a 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).

Table 2. Primer pairs used for the amplification of sequences of the medaka *hoxa2a* and *ψhoxa2b* UER(K20-RE5)s used in functional genomic analyses. Schematics of constructs, frequencies of transient transgenic embryos (F0s) showing reporter gene expression in the hindbrain and CNCCs, and numbers of stable-line transgenic adults (F1s) are shown. K, Krox20, B, BoxA, R, Rhombomeric element, *, embryos showing nearly undetectable *eGFP* expression in the hindbrain and pharyngeal arches.

Construct	Primer Pairs	Amplicon Length	Construct Schematic	Hindbrain Expression (F0)	CNCC Expression (F0)	F1s
Medaka <i>hoxa2a</i> UER(K20-RE5) construct design and transgenic analysis						
1	AF1/AR1	531 bp		42/48 (87.5%)	42/48 (87.5%)	3
2	AF2/AR1	221 bp		64/84 (76%)	56/84 (67%)	3
3	AF3/AR1	145 bp		39/49 (80%)	41/49 (84%)	0
4	AF4/AR1	107 bp		7/47 (15%) *	23/49 (49%) *	0
5	AF1/AR2	475 bp		42/50 (84%)	42/50 (84%)	4
6	AF1/AR3	430 bp		0/52 (0%)	0/52 (0%)	0
7	AF3/AR2	89 bp		52/62 (84%)	52/62 (84%)	4
Medaka <i>ψhoxa2b</i> UER(K20-RE5) construct design and transgenic analysis						
8	BF1/BR1	397 bp		46/51 (90%)	46/51 (90%)	2
9	BF2/BR1	197 bp		47/52 (90%)	47/52 (90%)	4
10	BF3/BR1	132 bp		33/38 (87%)	33/38 (87%)	3
11	BF4/BR1	94 bp		27/52 (52%) *	27/52 (52%)*	0
12	BF1/BR2	353 bp		33/41 (80%)	33/41 (80%)	3
13	BF1/BR3	309 bp		9/64 (14%) *	15/64 (23%) *	0

2.5. Microinjection of Medaka Embryos

Cultivation of medaka was performed as previously described [12]. Zygotes (stage 1) were collected from several females to ensure genetic variation per each injection experiment [72]. Zygotes were transferred to ice-cold 1× medaka embryo rearing medium (ERM) (17 mM· NaCl, 0.4 mM· KCl, 0.66 mM· MgSO₄· 7H₂O, 0.27 mM· CaCl₂· 2H₂O) between 30 min to 2 h to arrest development [73]. 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 × 1 mm deep × 4 mm long and were made using 60 mL of 1.5% agarose (Sigma-Aldrich, St. Louis, MO, USA). To reduce high internal egg pressures when microinjecting, the agarose corrals were overlaid with 10 mL of 15% Ficoll 400 (Sigma-Aldrich, St. Louis, MO, USA) 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, USA). Glass needles were pulled on a Sutter P-97 needle puller (Novato, CA, USA) 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, USA) housed in 100 mm petri dishes with dampened Kimwipes (Kimberly Clark®, Irving, TX, USA) 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, USA). 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, 1× Yamamoto Buffer (128 mM· NaCl, 2.7 mM· KCl, 1.8 mM· CaCl₂, 0.24 mM· NaHCO₃, 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 1× ERM. After transfer they were allowed to equilibrate to less than 1% Ficoll for roughly 2 h at room temperature (RT) without shaking. The medium housing the embryos was then replaced with fresh 1× ERM and the embryos were incubated at 28.5 °C to continue development.

2.6. Generation and Visualization of Transient and Stable-Line Transgenic Medaka Embryos

Microinjected embryos were reared in 1× ERM at 28.5 °C. After 24 h, 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 1× ERM containing 0.1 mM phenylthiourea (Sigma-Aldrich, St. Louis, MO, USA) to reduce pigmentation [74]. Embryos were visualized for *eGFP* expression using a Zeiss inverted compound microscope (Thornwood, NY, USA) at selected times during development, focusing primarily on developmental stage 29/30 (74–82 hpf [72]). 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. Images of *eGFP* expression were processed using Axiovision AC 4.4 software. Depending on the strength of the *eGFP* signal, camera exposure times ranged from 1 to 2 s 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 as a morphological landmark to determine whether the *eGFP* signal was occurring anteriorly or posteriorly in the hindbrain. Rhombomeres 3 and 4 develop dorsal to 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 reared 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 reared until hatching in 1× ERM at 28.5 °C. Hatched embryos were reared in 1× ERM in breeding nets in 5 gallon tanks until they were large and strong enough to swim against filter-generated currents. F1

medaka embryos were assayed for *eGFP* expression in the hindbrain and pharyngeal arches according to the procedure outlined above for transient transgenic embryos.

2.7. 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 developmental stages 22 (nine somites), 29/30 and 34 (121 hpf) [72]. We assayed embryos at stage 22 because we wanted to determine if the CREs of the medaka *hoxa2a* and *ψhoxa2b* UER(K20-RE5)s were directing reporter gene expression in the migratory CNCCs of the hyoid and post-otic streams, at stage 29/30 to visualize the exact rhombomeres and pharyngeal arches that were expressing *eGFP* and at stage 34 to determine if the CREs of the medaka *hoxa2a* and *ψhoxa2b* UER(K20-RE5)s were directing expression in post-migratory CNCCs at the chondrogenic stages of pharyngeal arch development.

Embryos from stable-line transgenic medaka fish were collected, reared, anesthetized, fixed in 4% paraformaldehyde (PFA) and dehydrated as previously described [12]. Medaka embryos were developmentally staged according to Iwamatsu (2004) [72]. Whole-mount *in situ* hybridization was performed according to Davis *et al.* (2008) [12]. All experiments used digoxigenin (DIG)-labeled sense and antisense riboprobes that were produced and purified according to Scemama *et al.* (2006) [19]. 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) [19]. 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 [13,16]. 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) [19].

2.8. Comparative Genomic Sequence Analysis

Genomic DNA sequences corresponding to the UER(K20-RE5)s of medaka *hoxa2a* and *ψhoxa2b* that were shown to be required for directing reporter gene expression in the rhombomeric and PA embryonic domains using functional genomic analyses outlined above were examined for putative transcription factor binding sites using JASPAR [75]. We also performed a comparative genomic sequence analysis of the UER(K20-RE5)s between medaka *hoxa2a* and *ψhoxa2b* and the UER(K20-RE5)s of other vertebrate *Hoxa2* genes using the software programs Dialign-TX and CLUSTALX in order to determine if the putative transcription factor binding sites were conserved in other vertebrates [76,77]. The *Hoxa2* UER(K20-RE5)s from genomic sequences of medaka (accession numbers AB232918 and AB232919), fugu (accession numbers DQ481663 and DQ481664), zebrafish (accession number AL645795, direct submission), Nile tilapia (AF533976 and GCA_000188235.1), bichir, mouse (accession number: NC000072), human (accession number: NG012078, direct submission), chicken (accession number: AC163712, direct submission), ceolacanth (accession number: FJ497005), horn shark (accession number: AF224262), dogfish (accession number: FQ032658), and western clawed frog (accession number: NW004668239) were used for sequence comparison [71,78–85]. All default and recommended parameters were used in the Dialign-TX and CLUSTALX programs for conducting genomic DNA sequence comparisons.

3. Results

3.1. 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 pT2AL200L150G vector. This plasmid vector contains the *Xenopus laevis* eFI- α S promoter upstream of *eGFP*. We expected that positive control embryos, which were co-microinjected with pT2AL200L150G vector and *Tol2* transposon mRNA, would show green fluorescence during development. However, we were unsure if negative control embryos, which were microinjected solely with pT2AL200L150G vector, would not exhibit green fluorescence, since the *Tol2* element is present within the medaka genome [70]. All positive control embryos showed strong reporter gene expression throughout the body of the medaka embryos (68/68; 100%) (Figure 2A,C) while negative control embryos failed to show any reporter gene expression (0/52; 0%) (Figure 2B,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- β MCS vector alone (0/49, 0%) or co-microinjected with vector and transposon mRNA (0/64, 0%) lacked detectable fluorescence (0/64, 0%) (Figure 2B,D). These results showed that the *Tol2* transposon system and the pTolTATA- β MCS vector can be used in medaka embryos for studying *cis*-regulatory element control of gene expression.

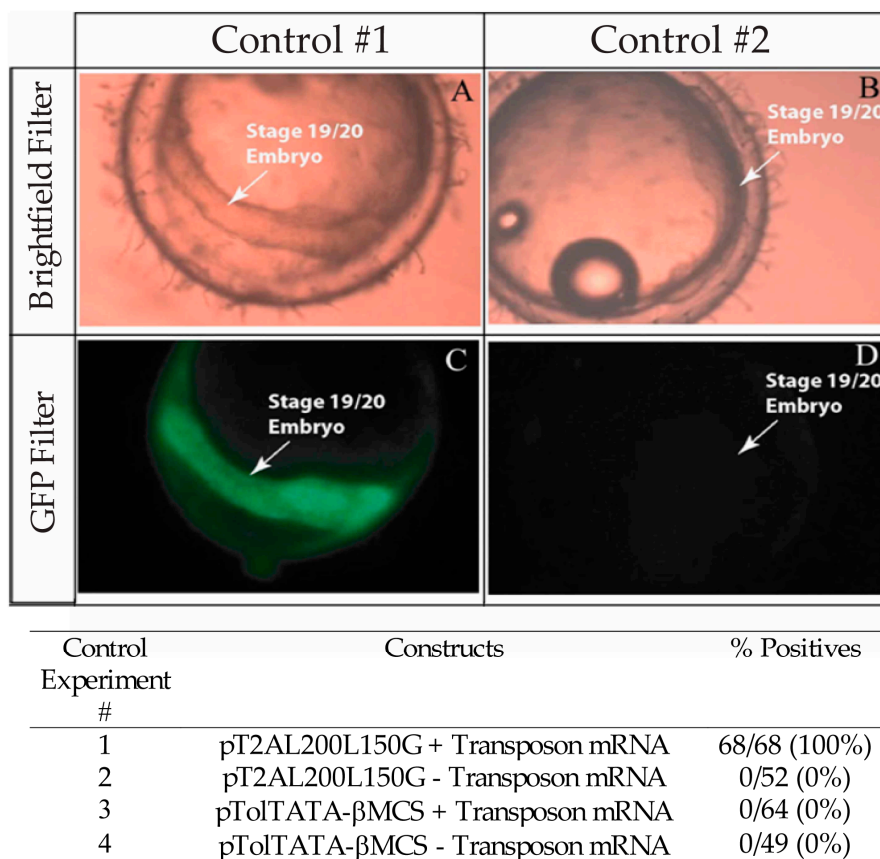


Figure 2. Control experiments used for the validation of the *Tol2* transposon system for medaka embryos. (A) Brightfield image of representative positive control embryo (pT2AL200L150G + Transposon mRNA); (C) Green fluorescent protein (GFP) filter image of the same embryo shown in A; (B) brightfield image of representative negative control embryo (pT2AL200L150G – Transposon mRNA); and (D) GFP filter image of the same embryo shown in B. No embryos in any of the other negative control experiments showed eGFP expression (pTolTATA- β MCS + Transposon mRNA or pTolTATA- β MCS-Transposon mRNA).

3.2. Functional Genomic Analysis of the Medaka *Hoxa2a* UER(K20-RE5)

Medaka *hoxa2a* is expressed in a relatively conserved manner in the hindbrain and pharyngeal arches similar 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 [6,12,14,19,23,24,31,45,49,56,57]. Based on this extensively conserved pattern of expression, we hypothesized that the medaka *hoxa2a* UER(K20-RE5) would direct 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 construct containing the medaka *hoxa2a* UER(K20-RE5) (Construct #1) directed reporter gene expression in r4 of the hindbrain and in PA2 and the posterior PAs (Figure 3A–G). Even though we observed *eGFP* expression in the hindbrain (87.5%) and pharyngeal arches (87.5%) in a high percentage of transgenic embryos (Table 2), both direct fluorescence and whole-mount *in situ* hybridization with riboprobes directed against *eGFP* expressed in stable-line transgenic embryos showed that the *eGFP* expression in the hindbrain was restricted to r4 (Figure 3D–G). This was unexpected given that the UERs of chicken and mouse *Hoxa2* or zebrafish *hoxa2b* directed reporter gene expression in r3 and r5 [21,25,27–32]. This restricted and unusual pattern of expression prompted us to examine the validity of rhombomere assignments. To authenticate the rhombomere assignments, additional whole-mount *in situ* hybridization experiments were conducted using medaka *hoxb1a* and *hoxd3a*, which are expressed in r4, and r6–8 of the hindbrain, respectively, but not in the pharyngeal arches. [13,16]. 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 and to overlap the region that hybridizes to the *hoxa2a* UER(K20-RE5) reporter construct #1. (Figure 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 (Figure 3I). These results are consistent with expression of *eGFP* in r4 with a gap in expression corresponding to r5 and *hoxd3a* expression in r6–8 (Figure 3I), and show that the medaka *hoxa2a* UER(K20-RE5) reporter expression was restricted to just r4. Further, *eGFP* riboprobes were observed to hybridize to their mRNA targets 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 (Figure 4A–C). These results confirm that the medaka *hoxa2a* UER(K20-RE5) construct #1 directs reporter gene expression in the CNCCs. Moreover, the presence of reporter gene expression within the CNCCs in the PAs of medaka is consistent with results from previous studies for orthologous UER sequences from mouse *Hoxa2*, fugu *hoxa2a* and zebrafish *hoxa2b* [24,27,29]. Interestingly, we observed more robust staining in *in situ* hybridization analyses using *eGFP* riboprobes in the posterior PAs than in PA2, which suggests that the sequence elements in the medaka *hoxa2a* UER(K20-RE5) construct #1 are expressed more strongly in the posterior PAs than for PA2, which is the converse of that observed for the cognate gene transcripts.

To pinpoint the CREs responsible for directing the r4 and PA-specific reporter gene expression in medaka embryos, 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* UER(K20-RE5) was performed (Constructs #2-7) (Table 2). These deletions eliminated regions corresponding to previously mapped enhancer elements including Krox20, BoxA, RE4, RE3, RE2 and RE5 [21,25,28–32]. Similar to transient transgenic embryos generated with Construct #1, microinjection of constructs that included 5'-deletions of Krox20 and BoxA binding sites (Construct #2) or Krox20, BoxA and RE4 (Construct #3) yielded high percentages of transient transgenic embryos with *eGFP* expression in the hindbrain (76%, Construct #2; 80%, Construct #3) and PAs (67%, Construct #2; 84%, Construct #3) (Table 2). Further, stable-line transgenic embryos generated with Construct #2 showed similar *eGFP* expression patterns to transgenic embryos generated with constructs containing the entire medaka *hoxa2a* UER(K20-RE5) (Construct #1). Specifically, we observed *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 PAs (Figure 4D–F). Like the full-length UER(K20-RE5) (construct #1) we observed stronger *eGFP* expression in the posterior PAs than in PA2. Unfortunately, we did not have success in obtaining stable-line transgenic embryos using Construct #3 (data not shown). Further, the extreme mosaic nature of the expression of the transient transgenic embryos treated with Construct #3 precluded us from drawing conclusions about the expression patterns. A much lower percentage of transient transgenic embryos generated with Construct #4, which included deletions of *Krox20*, *BoxA*, RE4 and RE3 sequences and only retained RE2 and RE5 sequences (Table 2), showed *eGFP* expression in the hindbrain (15%) and PAs (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 PAs and did not allow us to draw any conclusions of the expression patterns (data not shown). Unfortunately, we did not obtain any stable-line transgenic embryos treated with Construct #4 (data not shown). These results suggested that sequence elements downstream of RE4 within the medaka *hoxa2a* UER(K20-RE5), specifically, RE3, RE2 and RE5, are necessary for directing medaka *hoxa2a* expression in r4 and the CNCCs of PA2 and the posterior arches.

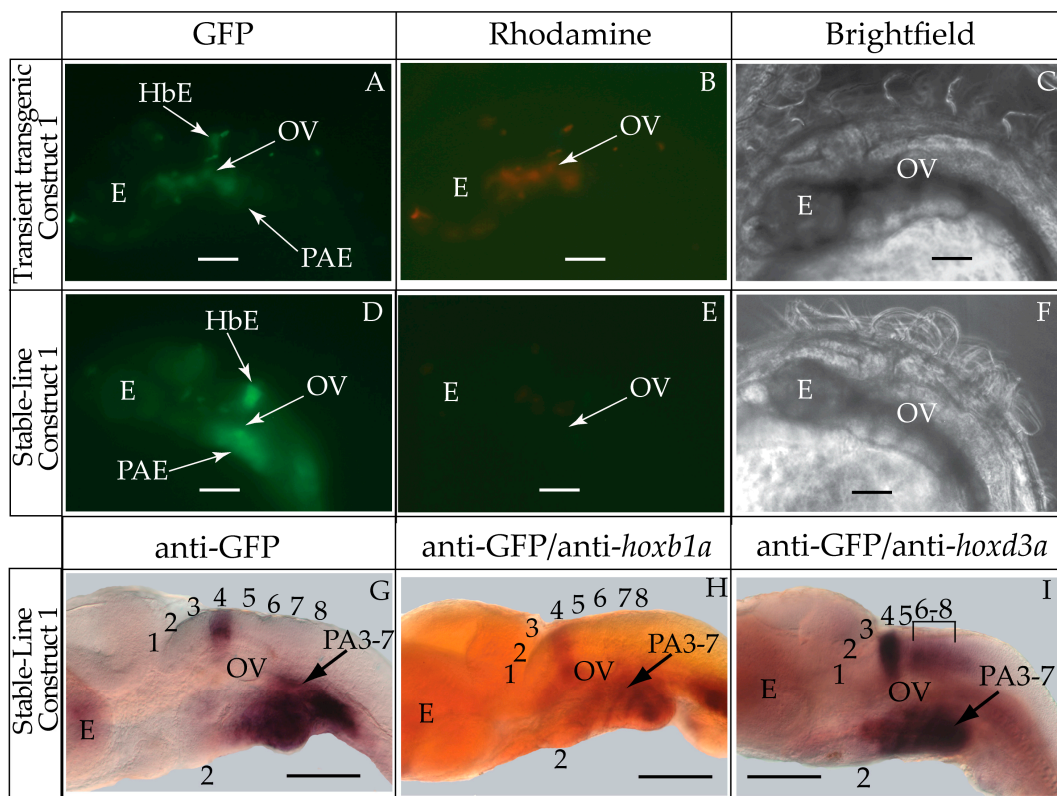


Figure 3. Transient (A–C) and stable-line (D–I) transgenic data from the medaka *hoxa2a* UER(K20-RE5) (Construct #1). (A–F) Images of transient (A–C) and stable-line (D–F) transgenic embryos at stage 29/30 (72–84 hpf) were taken using: GFP (A,D); rhodamine (B,E); and brightfield (C,F) filters. All embryos 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. Medaka *hoxd3a* expressing rhombomeres (r6–8) are indicated by the bracket in I. E, eye; HbE, hindbrain expression; OV, otic vesicle; PA, pharyngeal arch; PAE, pharyngeal arch expression. Scale bars equal 0.1 mm.

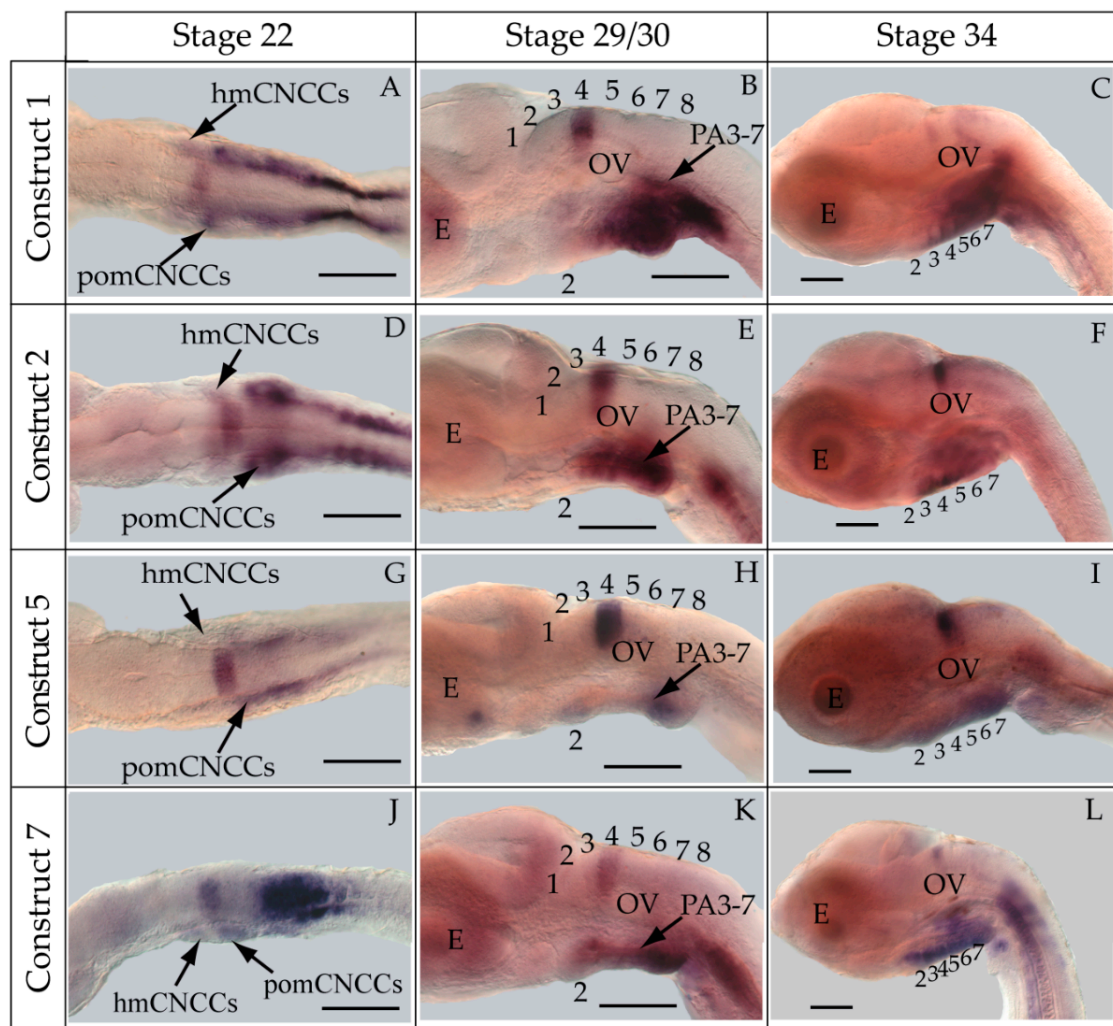


Figure 4. Whole-mount *in situ* hybridization of *eGFP* in stable-line *hoxa2a* UER(K20-RE5) transgenic medaka embryos generated with Construct #1 (A–C); Construct #2 (D–F); Construct #5 (G–I); and Construct #7 (J–L) at stages 22 (9 s) (A,D,G,J); 29/30 (72–84 hpf) (B,E,H,K); and 34 (121 hpf) (C,F,I,L). (A,D,G,J) Embryos were mounted with their anterior sides facing left and their dorsal sides facing the reader. (B,C,E,F,H,I,K,L) Embryos were mounted with their anterior sides facing left and lateral sides facing the reader. (B,E,H,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.

The microinjection of constructs with the 3'-specific deletion of RE5 from the medaka *hoxa2a* UER(K20-RE5) but retention of Krox20, BoxA, RE4, RE3 and RE2 (Construct #5) yielded similar results to that of Construct #1, namely a high percentage of transient transgenic embryos with *eGFP* expression in the hindbrain (84%) and PAs (84%) (Table 2). Further, stable-line transgenic embryos generated with Construct #5 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 in PA2 and posterior arches (Figure 4G–I). However, embryos injected with Construct #5 showed reduced *eGFP* expression levels in the posterior PAs when compared to embryos treated with Constructs #1 and #2 (Figure 4B,E,H). Interestingly, embryos injected with Construct #6, in which both RE2 and RE5 were deleted but Krox20, BoxA, RE4, and RE3 were retained, did not show any *eGFP* expression in the hindbrain (0%) or in

the pharyngeal arches (0%) (Table 2). Overall, these results showed that the sequence containing the elements corresponding to RE3 and RE2 of the medaka *hoxa2a* UER(K20-RE5) is required for r4 expression but the surrounding elements corresponding to Krox20, BoxA, RE4 and RE5 are not. Similar results were also observed for the reporter gene expression in the PAs, such that the loss of either RE3 or RE2 sequences yielded a loss in CNCC expression (Constructs #4 and 6) in transient transgenic embryos.

The aforementioned *eGFP* expression patterns obtained from the nested deletion constructs of the medaka *hoxa2a* UER(K20-RE5) prompted us to develop a construct that included the RE3 and RE2 elements only (Construct #7; Table 2). This construct was 89 bp in length and spanned from genomic bp positions –1392 to –1303 (Table 2). As expected based on the results from earlier tested constructs containing both RE3 and RE2 (Constructs #1, 2, 3 and 5), a high percentage of transient transgenic medaka embryos generated with Construct #7 showed *eGFP* expression in the hindbrain (84%) and PAs (84%) (Table 2). Stable-line transgenic medaka embryos confirmed that the sequence elements within this region directed *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 arches (Figure 4J–L). Further, in comparison to stable-line transgenic embryos generated with constructs containing RE3, RE2 and other surrounding elements (Constructs #1, 2 and 5), embryos generated with Construct #7 showed much more robust *eGFP* expression in PA2. Altogether, these results indicate that this 89 bp sequence contains elements necessary and sufficient to direct *Hoxa2* expression in r4, PA2 and the posterior PAs in the absence of other elements.

To identify putative transcription factor binding sites in the 89 bp region of the medaka *hoxa2a* UER(K20-RE5), hereafter referred to as the r4/CNCC-specifying element or core element, we performed a comparative genomic sequence analysis of this DNA sequence fragment with orthologous sequences from *Hoxa2* UERs of divergent gnathostomes, including horn shark, dogfish, coelacanth, frog, chicken, mouse, human, bichir, zebrafish, tilapia, fugu and medaka. Sequence alignments revealed several regions that were highly conserved among gnathostome *Hoxa2* UER genomic sequences but restricted to the region corresponding to the r4/CNCC-specifying element from medaka (Figure 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* UER(K20-RE5) and a Hox/Pbx binding site (5'-AGATTGATCG-3') beginning at bp position –1349 (Figure 5). Interestingly, the Prep/Meis and Hox/Pbx sites of the medaka *hoxa2a* UER(K20-RE5) 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 [37], which were shown to direct mouse *Hoxb2* expression in r4 and the migratory CNCCs [37]. To further support the concept of the r4/CNCC element as a core element, it is important to note that Hox/Pbx and Meis binding sites located upstream of mouse *Hoxa2*, but downstream of the sequence orthologous to the medaka r4/CNCC element, were shown to direct *Hoxa2* expression in the CNCCs of PA2 and the posterior arches [24]. Likewise, 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 [21,26,33]. Therefore, it is possible that flanking sequences not analyzed in this study that are orthologous to sequences of the *Hoxa2* UERs in other gnathostomes function to redirect the intrinsic r4 activity of the r4/CNCC-specifying element to r3 and r5.

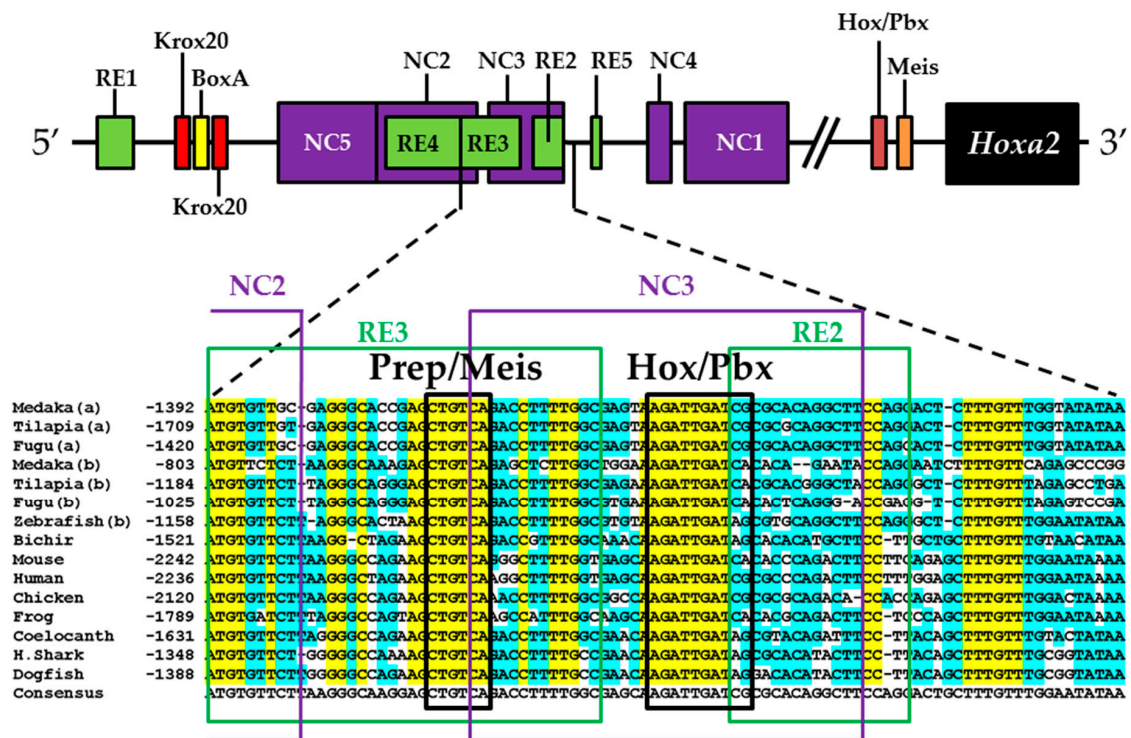


Figure 5. Comparative genomic sequence analysis of the 89 bp DNA fragment of the medaka *hoxa2a* UER(K20-RE5) (top sequence) that directs expression in r4 and the CNCCs. This sequence was compared to orthologous DNA sequences upstream of *hoxa2a* (denoted by a in parentheses after species name) and *hoxa2b* (denoted by b in parentheses after species name) of teleosts, *Hoxa2* of tetrapods and *Hoxa2* of coelocanth, 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 mouse *Hoxa2* UER 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 *Hoxa2* UER [27,28]. Black boxed regions correspond to transcription factor binding sites identified in this study. A consensus sequence was derived from the aligned sequences. NC, neural crest; RE, rhombomeric element.

3.3. Functional Genomic Analysis of the Medaka ψ *Hoxa2b* UER(K20-RE5)

Unlike the conserved hindbrain and pharyngeal arch expression patterns common to many teleost *hoxa2a* and *a2b* and tetrapod *Hoxa2* genes, medaka ψ *hoxa2b* 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 [12]. Based on the divergence in expression observed in comparisons of medaka with other teleost and tetrapod *Hoxa2* genes, we hypothesized that the UER(K20-RE5) of medaka ψ *hoxa2b* would direct expression in the noncanonical *Hox* PG2 domains instead of 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 ψ *hoxa2b* UER(K20-RE5) (Construct #8, Table 2) directed reporter gene expression in the hindbrain and PAs (Figure 6A–I, Table 2) instead of the expected noncanonical *Hox* PG2 expression domains, such as the pectoral fins. We observed a high percentage of transient transgenic embryos showing *eGFP* expression in the hindbrain (90%) and pharyngeal arches (90%) (Table 2). Whole-mount *in situ* hybridization analyses of stable-line transgenic embryos

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 arches and the chondrogenic CNCCs in PA2 and the posterior arches (Figure 7A–C). Interestingly, in transgenic reporter gene assays conducted in chicken embryos, a very similar construct containing a region corresponding to the the medaka *ψhoxa2b* UER(K20-RE5) used in this study, and also lacking the RE1 sequence element, directed reporter gene expression in r3 and r5 but not in the CNCCs, the pharyngeal arches or the noncanonical expression domains observed for *ψhoxa2b* expression in medaka [12,21].

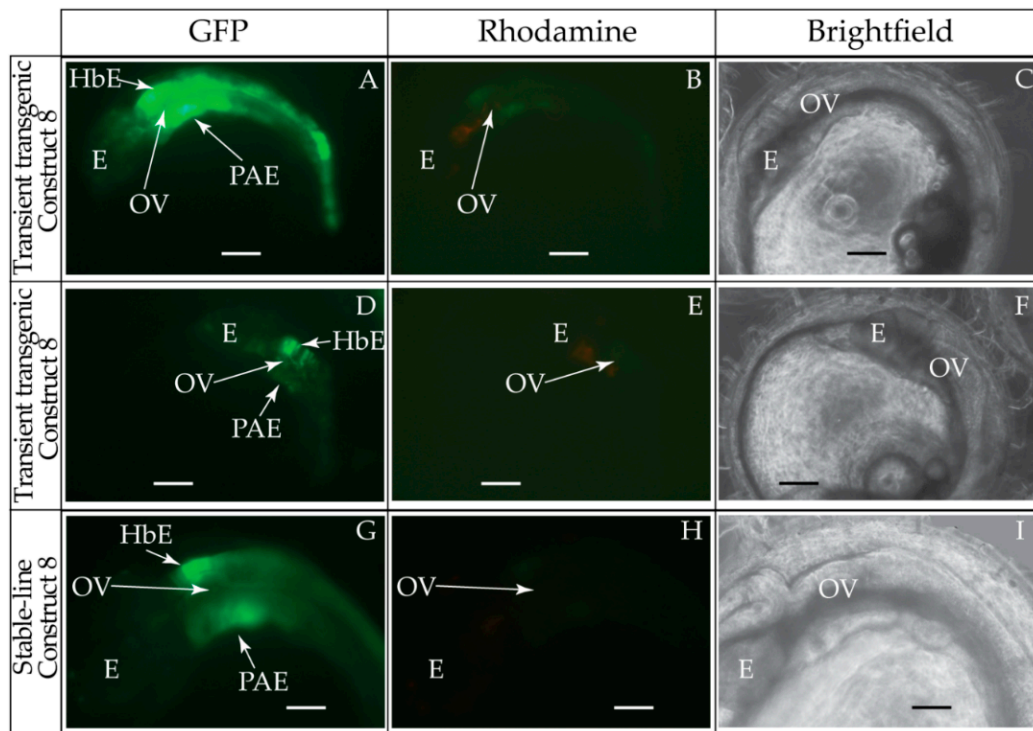


Figure 6. Transient (A–F) and stable-line (G–I) transgenic data from the medaka *ψhoxa2b* UER(K20-RE5)-(Construct #8). (A–I) Images of transient transgenic embryos at stage 29/30 (72–84 hpf) were taken using: GFP (A,D,G); rhodamine (B,E,H); and brightfield (C,F,I) filters. Transient transgenic analyses 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. E, eye; HbE, hindbrain expression; OV, otic vesicle; PAE, pharyngeal arch expression. Scale bars equal 0.1 mm.

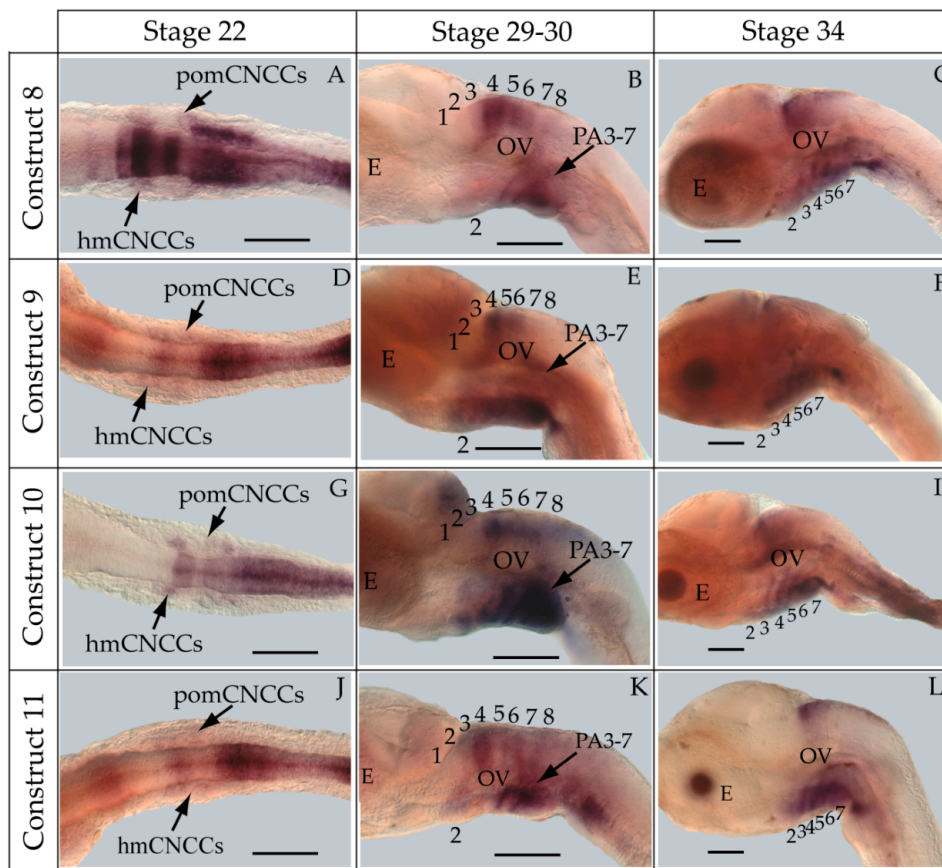


Figure 7. Whole-mount *in situ* hybridization of *eGFP* in stable-line *phiHOXA2b* UER(K20-RE5) transgenic medaka embryos generated with: Construct #8 (A–C); Construct #9 (D–F); Construct #10 (G, H and I); and construct #12 (J–L) at stages 22 (9 s) (A,D,G,J); 29/30 (72–84 hpf); (B,E,H,K); and 34 (121 hpf) (C,F,I,L). (A,D,G,J) Embryos were mounted with their anterior sides facing left and their dorsal sides facing the reader. (B,C,E,F,H,I,K,L) Embryos were mounted with their anterior sides facing left and lateral sides facing the reader. (B,E,H,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.

To precisely map the functional activity of the medaka *phiHOXA2b* UER(K20-RE5) on control of expression, a series of nested deletion constructs from the medaka *phiHOXA2b* UER(K20-RE5) that selectively eliminated enhancer elements and transcription factor binding sites similar to those that were deleted in the *HOXA2a* UER(K20-RE5) series of nested deletions was performed (Constructs #9–13) (see Table 2). Similar to embryos microinjected with Construct #8, microinjection of constructs that included 5'-deletions of Krox20 and BoxA binding sites (Construct #9) or Krox20, BoxA and RE4 (Construct #10) yielded high percentages of transient transgenic embryos with *eGFP* expression in the hindbrain (90%, Construct #9; 97%, Construct #10) and PAs (90%, Construct #9; 87%, Construct #10) (Table 2). Stable-line transgenic embryos generated with Constructs #9 and #10 showed similar *eGFP* expression pattern to those treated 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 arches and the chondrogenic CNCCs of PA2 and the posterior arches (Figure 7D–I). Similar to the medaka *HOXA2a* UER(K20-RE5) analysis with Construct #4 in which the Krox20, BoxA, RE4 and RE3 sequences were eliminated, a lower percentage of transient transgenic medaka embryos generated with Construct #11, which included the same deletions, showed *eGFP*

expression in the hindbrain (52%) and pharyngeal arches (52%) when compared to embryos generated with Constructs #8, 9 or 10 (Table 2). Further, *eGFP* expression was almost undetectable 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 (Data not shown). These results, like those for the medaka *hoxa2a* UER(K20-RE5), suggest that sequence elements downstream of the RE4 element function to direct medaka *phoxa2b* in r3–7, PA2 and the posterior arches. The microinjection of constructs with the 3'-specific deletion of RE5 from the medaka *phoxa2b* UER(K20-RE5) but retention of all sequence elements 5' to this site, including Krox20, BoxA, RE4, RE3 and RE2 (Construct #12), yielded similar results to Construct #8 with a high percentage of transient transgenic embryos showing *eGFP* expression in the hindbrain (80%) and the PAs (80%) (Table 2) and stable-line transgenic embryos generated with Construct #12, showed a similar *eGFP* expression pattern relative to stable-line transgenic embryos generated with Constructs #8, 9 and 10 (Figure 7J–L). Interestingly, similar to the medaka *hoxa2a* UER(K20-RE5) analysis with Construct #6 in which the RE2 and RE5 sequences were eliminated a low percentage of transient transgenic embryos generated with Construct #13, which contained the same deletions, showed *eGFP* expression in the hindbrain (14%) and pharyngeal arches (23%) (Table 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 (Data not shown). Overall, these results showed that the sequence containing the elements corresponding to RE3 and RE2 of the medaka *phoxa2b* UER(K20-RE5) is required for r3–7 and CNCC expression in PA2 and the posterior PAs but the surrounding elements corresponding to Krox20, BoxA, RE4 and RE5 are not.

A comparative genomic sequence analysis revealed that the 88 bp DNA sequence fragment of the medaka *phoxa2b* UER(K20-RE5), hereafter referred to as the r3–7/CNCC-specifying element, like the paralogous *hoxa2a* r4/CNCC-specifying element and orthologous *Hoxa2* UER(K20-RE5) genomic sequences of other gnathostomes contains conserved Hox/Pbx, and Prep/Meis transcription factor binding sites (see Figure 5). While these sites may be involved in directing *eGFP* expression in r4 and the CNCCs, their activity may also be redirected to a much broader domain of expression, including r3, r5, r6 and r7. We observed 33 bp differences between the paralogous core sequences upstream of medaka *hoxa2a* and *phoxa2b* (Figure 8), and it is possible that the sequence differences between them have allowed the *phoxa2b* sequence, but not the *hoxa2a* sequence, to be responsive to transcription factors that mediate expression 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 *phoxa2b* fragment but not for *hoxa2a*. Sox proteins have been shown to be integral factors in directing *Hox* gene expression among different rhombomeres of the hindbrain [59], indicative that the sequence differences leading to the evolution of additional Sox binding elements in the UER(K20-RE5) of *phoxa2b* fragment relative to the *hoxa2a* UER(K20-RE5) may be responsible for the broadened rhombomeric expression observed for the *phoxa2b* UER(K20-RE5) relative to the *hoxa2a* UER(K20-RE5).

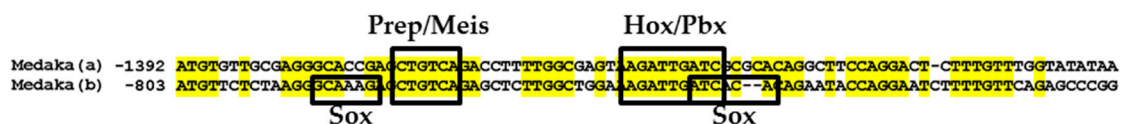


Figure 8. Sequence alignment of the medaka *hoxa2a* r4/CNCC and the *phoxa2b* r3–7/CNCC specifying elements. The sequence corresponding to the medaka *hoxa2a* r4/CNCC specifying element is denoted by “a” in parentheses. The sequence corresponding to the medaka *phoxa2b* r3–7/CNCC specifying element is denoted by “b” in parentheses. Numbers correspond to genomic base pair positions relative to the ATG start site of *hoxa2a* and *phoxa2b*. 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.

4. Discussion

4.1. 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 [70]. 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- α S 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- α S promoter and transposon mRNA showed strong reporter gene expression throughout the body whereas embryos that were microinjected solely with constructs containing the *Xenopus* eFI- α S promoter were not able to direct reporter gene expression.

Our reporter gene expression results showed that the UER(K20-RE5)s from medaka *hoxa2a* and *phoxa2b* 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* UER(K20-RE5)-directed reporter gene expression in r4 of the hindbrain and the CNCCs of PA2-7 while the medaka *phoxa2b* UER(K20-RE5) directed reporter gene expression in r3-7 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. In support, neither the medaka *hoxa2a* nor *phoxa2b* UER(K20-RE5)s yielded similar reporter gene expression results when tested in chicken embryos [21]. Specifically, while the *hoxa2a* UER(K20-RE5) did not direct any reporter gene expression in the chicken embryonic head, the *phoxa2b* UER(K20-RE5) only directed expression in r3 and r5 but not the CNCCs [21]. The inconsistent results for the medaka *hoxa2a* and *phoxa2b* UER(K20-RE5)s within the medaka and chicken embryos suggest 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, previous reporter gene experiments showed that the mouse *Hoxa2* UER was able to generate reporter gene expression in r3, r5 and the CNCCs in both mouse and zebrafish embryos but unable to direct any reporter gene expression in the hindbrain or CNCCs of chick embryos [29,32]. Similar results were also observed for the fugu *hoxa2a* UER, such that it was shown to be functional when it was tested in both mouse and zebrafish embryos, but not in the chicken [29,32]. Conversely, the UER of chick *Hoxa2* was unable to direct reporter gene expression in either mouse or zebrafish embryos [29]. These results along with our reporter gene expression results of the medaka *hoxa2a* and *phoxa2b* UER(K20-RE5)s in medaka embryos suggest that either the functional nature of the sequence elements of the UER, the transcription factors of the genetic regulatory networks that interact with these sequences or a combination of both have diverged greatly among chicken, and possibly other archosaurs, from evolutionarily divergent osteichthyans. Future analyses of the medaka *hoxa2a* and *phoxa2b* UER(K20-RE5)s should incorporate the use of zebrafish and/or mouse embryos to assay for species-specific differences in regulatory requirements among these species. Further, orthologous sequences from the mouse *Hoxa2* and zebrafish *hoxa2b* UERs should be tested within medaka embryos.

4.2. Medaka *Hoxa2a*-Directed Gene Expression in the Hindbrain

In this study we showed, using reporter gene assays and whole-mount *in situ* hybridization, that the medaka *hoxa2a* UER(K20-RE5) does not direct expression in r3 and r5 of the hindbrain as expected based on previous whole-mount *in situ* hybridization results of medaka *hoxa2a* [12], but instead directs expression in r4. We found these results interesting since this region of genomic DNA was shown to be conserved structurally with orthologous genomic regions in evolutionarily divergent osteichthyans [21,29,32,59,86]. Further, the *Hoxa2* UERs of zebrafish, mouse and chicken have been shown to direct reporter gene expression in r3 and r5 when tested in their respective homologous

host systems [21,25,28–32,59]. An explanation for the lack of reporter gene expression in r3 and r5 in this study is that the genomic DNA sequence tested requires the cooperation of other surrounding sequences within the medaka *hoxa3a-a2a* intergenic region. Our reporter gene analysis of the medaka *hoxa2a* UER(K20-RE5) utilized a 531 bp genomic DNA sequence and contained all of the *cis*-regulatory elements that were tested in the chick embryonic model system [21], including Krox20 and BoxA binding sequences and sequence elements that pertain to RE4, RE3, RE2 and RE5. Functional genomic analyses of the mouse *Hoxa2* UER included regulatory elements orthologous to those mentioned above as well as a RE1 sequence element located upstream of Krox20 [28]. The deletion of the sequences upstream of the Krox20 elements, including the RE1 element, from the mouse *Hoxa2* UER resulted in the loss of reporter gene expression in r3 and r5 of stable-line transgenic mouse embryos [28]. Recent comparative and functional genomic sequence analyses utilizing the UERs of fugu *hoxa2a* and *a2b* and zebrafish *hoxa2b* have shown the presence of a RE1 sequence that is upstream of the Krox20 site in teleosts [29]. Further, these sequences were shown to be much farther upstream from Krox20 in teleosts in comparison to tetrapods [29]. For instance, in comparison to the mouse *Hoxa2* UER, in which the RE1 sequence begins 88 bp upstream of the Krox20 binding site, the orthologous RE1 sequence of medaka *hoxa2a* begins 290 bp upstream of Krox20 [29]. Our analysis of the medaka *hoxa2a* UER(K20-RE5) utilized a sequence that began at genomic bp position –1778, and it excluded the recently identified teleost RE1 sequence that begins at genomic bp position –2041 [29]. Future reporter gene analyses in medaka embryos should be performed that utilize constructs containing medaka *hoxa3a-a2a* sequences that are orthologous to the “full-length” UERs tested in other gnathostomes to determine if the flanking sequences, including the RE1 element, function in conjunction with the *hoxa2a* UER(Krox20-RE5) sequence to direct reporter gene expression in r3 and r5.

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 riboprobes of specific rhombomeric molecular markers. These molecular markers included medaka *hoxb1a*, which is expressed exclusively in r4 of the hindbrain [16], and medaka *hoxd3a*, which is expressed in r6, r7 and r8 [13]. Microinjection of a series of nested deletion constructs derived from the 531 bp fragment encompassing the medaka *hoxa2a* UER(K20-RE5) showed that the r4-specifying sequence lies within a 89 bp fragment that spans from genomic bp positions –1392 to –1303. Comparative genomic sequence analyses showed the presence of putative transcription factor binding sites, Prep/Meis and Hox/Pbx, located within this fragment. Further, these sites were shown to be highly conserved in *Hoxa2* genomic sequences across several evolutionarily divergent gnathostome vertebrates, including shark, bichir, coelacanth, teleosts and tetrapods (see Figure 5). Interestingly, Hox/Pbx and Prep/Meis binding sites have been shown in multiple studies to be crucial sequences in directing *Hox* expression in r4 [21,33,37]. 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 this fragment of the medaka *hoxa2a* UER(K20-RE5) 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 vertebrate *Hoxa3-a2* intergenic sequences direct reporter gene expression in r4.

A potential explanation for the presence of the r4-directed expression is that the surrounding genomic sequences that were not included in this analysis, such as the RE1 sequence, may function to restrict the r4/CNCC-specifying element from potentiating r4-specific expression or redirect it to direct expression in r3 and r5. In support, recent analyses of CREs responsible for directing fugu *pax2* paralogs, *pax2.1* and *pax2.2*, have shown that the functional activity of the paralogous CREs cognate to these two genes are influenced by interaction among clustered *cis*-regulatory elements in a manner very similar to those from the medaka *hoxa2a* UER(K20-RE5) and their immediately flanking sequences [87]. Specifically, the conserved noncoding element (CNE) pair 1, along with 50 bp extending both 5' and 3' of this element for *pax2.1* was shown to direct reporter gene expression in the notochord, muscle

and fins [87]. However, when the flanking sequences were removed, the deletion construct containing the more tightly defined CNE pair 1 of fugu *pax2.1* directed an expanded expression pattern that included the otic vesicle, blood, heart, skin and pronephric region [87]. By contrast, the removal of flanking sequences from the paralogous element of *pax2.2* resulted in the gain of directed reporter gene expression in the fins but loss of expression from the eye and otic vesicle [87]. By comparison, we argue that the inclusion of flanking sequences to the medaka *hoxa2a* UER(K20-RE5), including the RE1 sequence, may change the functional activity of the medaka UER(K20-RE5) region to direct reporter gene expression in r3 and r5 but negate expression in r4.

4.3. *Hoxa2a*-Directed Gene Expression in the Cranial Neural Crest Cells

Beyond the hindbrain expression directed activity of the medaka *hoxa2a* UER(K20-RE5), this genomic DNA region was also shown to possess functional activity that directs *hoxa2a* expression in the CNCCs. Our whole-mount *in situ* hybridization results from stable-line transgenic medaka embryos have shown that the medaka *hoxa2a* UER(K20-RE5) directs 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 craniofacial skeleton 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 craniofacial skeleton [45,56,62,64,88]. Our analyses suggest that the genomic sequences of the medaka UER(K20-RE5) are utilized *in vivo* to maintain medaka *hoxa2a* expression up until chondrogenesis in PA2 and the posterior arches to allow for proper craniofacial skeleton development.

Although we observed *eGFP* expression patterns in the post-migratory CNCCs that were directed by the medaka *hoxa2a* UER(K20-RE5), they did not fully phenocopy the post-migratory CNCC expression patterns shown by medaka *hoxa2a* whole-mount *in situ* hybridization experiments [12]. 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 Figure 4). However, we observed strong *eGFP* expression in PA2 and the posterior arches in stable-line transgenic embryos generated with Construct #7 (see Figure 4). 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* UER(K20-RE5) and flanking genomic sequences will be required to fully phenocopy the medaka *hoxa2a* expression pattern in the post-migratory CNCCs.

Our functional and comparative genomic analyses showed that the 89-bp region (Construct #7) responsible for directing reporter gene expression in the CNCCs of PA2 and the posterior arches contains *Hox/Pbx* and *Prep/Meis* binding sites (see Figure 5). These results are suggestive of auto- and/or cross-regulation of medaka *Hox* PG2 genes in the PAs. Recent functional genomic analyses in mouse showed that *Hox/Pbx* and *Meis* binding sequences located within the proximal promoter region upstream of mouse *Hoxa2* are necessary for *Hoxa2* expression and function in PA2 and posterior arches [24]. Further, this study showed that mouse *Hoxa2*, *Pbx* and *Meis* utilize these binding sites to function in a positive feedback loop that maintains and amplifies *Hoxa2* expression in the PAs [24]. Specifically, *Hoxa2* was shown to positively regulate *Meis* within the PAs, and *Meis* along with *Hoxa2*, were shown to work in concert in amplifying *Hoxa2* expression within these domains [24]. Unlike tetrapods, many species of ray-finned fishes contain multiple *Hox* PG2 genes that are expressed in PA2 and the posterior arches up until chondrogenesis, including *hoxa2a*, *a2b* and *b2a* [12,14,19,56,89]. Further, functional and comparative genomic sequence analyses have shown that *Hox/Pbx* and *Prep/Meis* sites are located within the *hoxb3a-b2a* intergenic region of ray-finned fishes and *Hoxb3-b2* of lobe-finned fishes [37,89]. These results suggest that the *Hox/Pbx* and *Prep/Meis* sites are used by teleost *Hox* PG2 genes in auto- and cross-regulatory mechanisms of gene expression in the PAs. In

support, the independent knockdown of *hoxa2a* and *b2a* in tilapia resulted in altered expression levels of tilapia *Hox* PG2 genes in the PAs [64]. Specifically, the knockdown of tilapia *hoxa2a* resulted in a reduction of *hoxa2a* expression from PA2 and the posterior arches and *hoxb2a* expression from PA2, and the knockdown of *hoxb2a* resulted in the reduction of *hoxa2a* and *a2b* expression from PA2 [64]. These altered expression levels corresponded to the varying severity of loss-of-function phenotypes in tilapia, wherein the loss of *hoxa2a* resulted in a full homeotic transformation the PA2-derived skeletal elements but the loss of *hoxb2a* resulted in altered bony morphology of PA2 elements without a full homeotic transformation [64]. By contrast, in zebrafish the knockdown of both *hoxa2b* and *b2a* were required to obtain a full homeotic transformation of PA2-derived skeletal elements [56], which suggests that both zebrafish *Hox* PG2 genes function redundantly in utilizing the Hox/Pbx and Prep/Meis binding sites to auto- and cross-regulate each other's expression in PA2. Future functional genetic analyses must be performed to understand if medaka *hoxa2a* and *b2a* utilize the Hox/Pbx and Prep/Meis binding sites similar to *hoxa2b* and *b2a* of zebrafish or *hoxa2a* and *b2a* of tilapia.

4.4. Functional Nature of the Medaka ψ *Hoxa2b* 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 medaka ψ *hoxa2b* UER(K20-RE5) is functional and is able to direct robust reporter gene expression in r3–7 of the hindbrain, PA2 and the posterior PAs. These results were contrary to our expectations based on whole-mount *in situ* hybridization analyses of medaka ψ *hoxa2b*, 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 [12]. These contradictory results suggest that regions of genomic sequences surrounding the medaka ψ *hoxa2b* UER(K20-RE5) function to redirect the activity of this region to mediate expression in the noncanonical *Hox* PG2 embryonic compartments. In addition to the unexpected functional nature of the medaka ψ *hoxa2b* UER(K20-RE5), this region was observed to be divergent in function from the UER(K20-RE5) of medaka *hoxa2a*, which directs gene expression in r4, PA2 and the posterior arches. Further, deletion mutagenesis of the medaka ψ *hoxa2b* UER(K20-RE5) showed that a relatively short sequence that is paralogous to the r4/CNCC-specifying element of the medaka *hoxa2a* UER(K20-RE5) was responsible for directing reporter gene expression in r3–7, PA2 and the posterior PAs. Comparative genomic analysis of the medaka ψ *hoxa2b* r3–7/CNCC-specifying element revealed the presence of conserved Hox/Pbx and Prep/Meis binding sites (see Figure 5). However, this sequence was shown to be divergent from the medaka *hoxa2a* r4/CNCC-specifying element by 33 bp (see Figure 8). It is possible that these substitutions between the medaka *hoxa2a* and ψ *hoxa2b* genomic sequences have allowed the ψ *hoxa2b*-specific sequence, but not the *hoxa2a*, to be receptive to transcription factors expressed within r3, r5, r6 and r7 of the hindbrain. In support, analyses of the CREs that direct fugu *pax2.1* and *pax2.2* gene expression showed that while there is high sequence similarity between several paralogous CREs, these elements direct divergent expression patterns [87]. For instance, although the pair 1 CNEs of *pax2.1* and *pax2.2* show roughly 90% similarity in sequence, the pair 1 CNE of fugu *pax2.1* directs gene expression in the spinal cord, muscle, notochord and fins while the paralogous sequence of *pax2.2* directs expression in the forebrain, midbrain, hindbrain, spinal cord, eye, otic vesicle, blood, heart, skin, thyroid region and pronephric region [87]. Interestingly, our analyses of the medaka *Hox cis*-regulatory sequences using the JASPAR software program revealed several Sox binding elements within the ψ *hoxa2b* r3–7/CNCC-specifying element but not in the *hoxa2a* r4/CNCC-specifying element (see Figure 8). Sox proteins have been shown to be involved in directing *Hox* gene expression in several rhombomeres of the hindbrain [59], so it is possible that these elements were involved in directing reporter gene expression in r3, r5, r6 and r7. The base pair mutations between the r3–7/CNCC-specifying element of medaka ψ *hoxa2b* and the paralogous r4/CNCC-specifying element may have occurred due to relaxation of selective pressures on the ψ *hoxa2b* genomic sequence after the inactivation of the *hoxa2b* gene in the lineage leading to medaka [12].

Our reporter gene expression results of the medaka *ψhoxa2b* UER(K20-RE5) coupled with previous expression pattern results of medaka *ψhoxa2b* during embryonic development may provide an excellent example of how evolutionary changes in noncoding sequences contribute to morphological evolution. An increasing body of evidence in the field of evolutionary and developmental biology has shown that sequence 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 [90]). Assuming that the medaka *ψhoxa2b* transcript gives rise to a functional product that can affect developmental specification, the heterotopic shift in expression that lead to medaka *ψhoxa2b* co-option into noncanonical *Hox* PG2 developmental compartments may have contributed to unique morphological and functional traits. Based on *in silico* translation of the likely mRNA product derived from *ψhoxa2b* cDNA clones, it is clear that this pseudogene cannot generate a fully functional Hox protein product [12]. However, it can generate a truncated protein that possesses a hexapeptide that is conserved with orthologous hexapeptides of *hoxa2b* genes from evolutionarily divergent teleosts [12]. 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 transcriptional activity in developmental compartments in which it is expressed. In the case of medaka, which is a member 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 developing pectoral fin buds [12]. Several functional genetic studies have shown that *Meis*, *Pbx*, and posterior *Hox* genes within the A and D clusters function to pattern the proximal-distal axis of the developing limbs [91–96]. Likewise, *Pbx* has been shown to be involved in the upregulation of posterior *HoxA* and *D* genes, and that the loss of function of either *Pbx* or posterior *HoxA* and *D* cluster genes results in truncated limbs [92,93]. Given the location of expression of the pseudogene and its potential to interact with *Hox* gene 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, a remarkable adaptation that is restricted to this taxonomic clade. In order to test these hypotheses, functional genomic analyses of the *hoxa2b* UER(K20-RE5) 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 members of this taxonomic clade.

5. Conclusions

In conclusion, our reporter gene analyses showed that the medaka *hoxa2a* and *ψhoxa2b* UER(K20-RE5)s are functionally divergent from one another. While both UER(K20-RE5)s were shown to direct reporter gene expression in the CNCCs of PA2 and the posterior PAs, the medaka *hoxa2a* UER(K20-RE5) directed expression in r4 of the hindbrain while the medaka *ψhoxa2b* directed expression in r3–7. These results are different from heterologous reporter gene expression results of the medaka *hoxa2a* and *ψhoxa2b* UER(K20-RE5)s when they were tested in chicken embryos. In chicken, the medaka *hoxa2a* UER(K20-RE5) did not direct expression in the hindbrain or pharyngeal arches while the *ψhoxa2b* UER(K20-RE5) directed reporter gene expression in just r3 and r5 of the hindbrain. Overall, our analyses show that when tested independent of their natural genomic surroundings, the UER(K20-RE5)s of medaka *hoxa2a* and *ψhoxa2b* are able to direct expression within r4 and *Hox* PG2 canonical domains, respectively.

Acknowledgments: We acknowledge Koichi Kawakami for the generous gift of the pT2AL200L150G plasmid vector. We acknowledge Jean-Luc Scemama for providing guidance with this project. We acknowledge Micheal Odom for aiding in the development of the medaka microinjection method.

Author Contributions: A.D. designed and developed all constructs, collected all transgenic medaka data for his Ph.D. dissertation, and wrote much of this manuscript under the guidance of E.J.S. E.J.S. advised with the design

of this project and aided in the writing of this manuscript. M.C.R. designed the pToTATA- β MCS vector system for this M.S. thesis under the guidance of E.J.S and aided in the writing of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

A-P	anterior-posterior
bp	base pair
CNCC	cranial neural crest cell
CNE	conserved noncoding element
CRE	<i>cis</i> -regulatory element
eGFP	enhanced green fluorescent protein
ERM	embryo rearing medium
<i>E. coli</i>	<i>Escherichia coli</i>
GFP	green fluorescent protein
NC	neural crest
PA	pharyngeal arch
PG	paralog group
r	rhombomere
RE	rhombomeric element
UER	upstream enhancer region
UER(K20-RE5)	upstream enhancer region spanning 5' from Krox20 binding element to the 3' RE5 element

References

1. McGinnis, W.; Krumlauf, R. Homeobox genes and axial patterning. *Cell* **1992**, *68*, 283–302. [[CrossRef](#)]
2. Ferrier, D.E.; Minguillon, C.; Holland, P.W.; Garcia-Fernandez, J. The amphioxus Hox cluster: Deuterostome posterior flexibility and Hox14. *Evolut. Dev.* **2000**, *2*, 284–293. [[CrossRef](#)]
3. Holland, P.W.; Garcia-Fernandez, J. Hox genes and chordate evolution. *Dev. Biol.* **1996**, *173*, 382–395. [[CrossRef](#)] [[PubMed](#)]
4. Powers, T.P.; Amemiya, C.T. Evidence for a Hox14 paralog group in vertebrates. *Curr. Biol.* **2004**, *14*, R183–R184. [[CrossRef](#)] [[PubMed](#)]
5. Amores, A.; Force, A.; Yan, Y.L.; Joly, L.; Amemiya, C.; Fritz, A.; Ho, R.K.; Langeland, J.; Prince, V.; Wang, Y.L.; *et al.* Zebrafish *hox* clusters and vertebrate genome evolution. *Science* **1998**, *282*, 1711–1714. [[CrossRef](#)] [[PubMed](#)]
6. Amores, A.; Suzuki, T.; Yan, Y.L.; Pomeroy, J.; Singer, A.; Amemiya, C.; Postlethwait, J.H. Developmental roles of pufferfish *Hox* clusters and genome evolution in ray-fin fish. *Genome Res.* **2004**, *14*, 1–10. [[CrossRef](#)] [[PubMed](#)]
7. Hoegg, S.; Boore, J.L.; Kuehl, J.V.; Meyer, A. Comparative phylogenomic analyses of teleost fish Hox gene clusters: Lessons from the cichlid fish *Astatotilapia burtoni*. *BMC Genom.* **2007**, *8*, 317. [[CrossRef](#)] [[PubMed](#)]
8. Moghadam, H.; Ferguson, M.; Danzmann, R. Evolution of Hox Clusters in Salmonidae: A Comparative Analysis Between Atlantic Salmon (*Salmo salar*) and Rainbow Trout (*Oncorhynchus mykiss*). *J. Mol. Evol.* **2005**, *61*, 636–649. [[CrossRef](#)] [[PubMed](#)]
9. Mungpakdee, S.; Seo, H.; Angotzi, A.R.; Dong, X.; Akalin, A.; Chourrout, D. Differential evolution of the 13 Atlantic salmon *hox* clusters. *Mol. Biol. Evol.* **2008**, *25*, 1333–1343. [[CrossRef](#)] [[PubMed](#)]
10. Prince, V. The *Hox* paradox: More complex(es) than imagined. *Dev. Biol.* **2002**, *249*, 1–15. [[CrossRef](#)] [[PubMed](#)]
11. Stellwag, E.J. Hox gene duplication in fish. *Semin. Cell Dev. Biol.* **1999**, *10*, 531–540. [[CrossRef](#)] [[PubMed](#)]

12. Davis, A.; Scemama, J.L.; Stellwag, E.J. Japanese medaka *Hox* paralog group 2: Insights into the evolution of *Hox* PG2 gene composition and expression in the Osteichthyes. *J. Exp. Zool. B Mol. Dev. Evol.* **2008**, *310*, 623–641. [[CrossRef](#)] [[PubMed](#)]
13. Davis, A.; Stellwag, E.J. Spatio-temporal patterns of *Hox* paralog group 3–6 gene expression during Japanese medaka (*Oryzias latipes*) embryonic development. *Gene Expr. Patterns* **2010**, *10*, 244–250. [[CrossRef](#)] [[PubMed](#)]
14. Le Pabic, P.; Stellwag, E.J.; Brothers, S.N.; Scemama, J.L. Comparative analysis of *Hox* paralog group 2 gene expression during Nile tilapia (*Oreochromis niloticus*) embryonic development. *Dev. Genes Evol.* **2007**, *217*, 749–758. [[CrossRef](#)] [[PubMed](#)]
15. Lyon, R.S.; Davis, A.; Scemama, J.L. Spatio-temporal expression patterns of anterior *Hox* genes during Nile tilapia (*Oreochromis niloticus*) embryonic development. *Gene Expr. Patterns* **2013**, *13*, 104–108. [[CrossRef](#)] [[PubMed](#)]
16. Hurley, I.A.; Scemama, J.; Prince, V.E. Consequences of *Hoxb1* duplication in teleost fish. *Evol. Dev.* **2007**, *9*, 540–554. [[CrossRef](#)] [[PubMed](#)]
17. Le Pabic, P.; Stellwag, E.J.; Scemama, J. Embryonic development and skeletogenesis of the pharyngeal jaw apparatus in the cichlid Nile tilapia (*Oreochromis niloticus*). *Anat. Rec.* **2009**, *292*, 1780–1800. [[CrossRef](#)] [[PubMed](#)]
18. Mungpakdee, S.; Seo, H.; Chourrout, D. Spatio-temporal expression patterns of anterior *Hox* genes in Atlantic salmon (*Salmo salar*). *Gene Expr. Patterns* **2008**, *8*, 508–514. [[CrossRef](#)] [[PubMed](#)]
19. Scemama, J.L.; Vernon, J.L.; Stellwag, E. Differential expression of *Hoxa2a* and *Hoxa2b* genes during striped bass embryonic development. *Gene Expr. Patterns* **2006**, *6*, 843–848. [[CrossRef](#)] [[PubMed](#)]
20. Soshnikova, N.; Dewaele, R.; Janvier, P.; Krumlauf, R.; Duboule, D. Duplications of *hox* gene clusters and the emergence of vertebrates. *Dev. Biol.* **2013**, *378*, 194–199. [[CrossRef](#)] [[PubMed](#)]
21. Tümpel, S.; Cambroner, F.; Wiedemann, L.M.; Krumlauf, R. Evolution of cis elements in the differential expression of two *Hoxa2* coparalogous genes in pufferfish (*Takifugu rubripes*). *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 5419–5424. [[CrossRef](#)] [[PubMed](#)]
22. Gendron-Maguire, M.; Mallo, M.; Zhang, M.; Gridley, T. *Hoxa-2* mutant mice exhibit homeotic transformation of skeletal elements derived from cranial neural crest. *Cell* **1993**, *75*, 1317–1331. [[CrossRef](#)]
23. Rijli, F.M.; Mark, M.; Lakkaraju, S.; Dierich, A.; Dolle, P.; Chambon, P. 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* **1993**, *75*, 1333–1349. [[CrossRef](#)]
24. Amin, S.; Donaldson, I.J.; Zannino, D.A.; Hensman, J.; Rattray, M.; Losa, M.; Spitz, F.; Ladam, F.; Sagestrom, C.; Bobola, N. *Hoxa2* selectively enhances Meis binding to change a branchial arch ground state. *Dev. Cell* **2015**, *32*, 265–277. [[CrossRef](#)] [[PubMed](#)]
25. Frasch, M.; Chen, X.; Lufkin, T. Evolutionary-conserved enhancers direct region-specific expression of the murine *Hoxa-1* and *Hoxa-2* loci in both mice and *Drosophila*. *Development* **1995**, *121*, 957–974. [[PubMed](#)]
26. Lampe, X.; Samad, O.A.; Guiguen, A.; Matis, C.; Remacle, S.; Picard, J.J.; Rijli, F.M.; Rezsöházy, R. An ultraconserved *Hox*-*Pbx* responsive element resides in the coding sequence of *Hoxa2* and is active in rhombomere 4. *Nucleic Acids Res.* **2008**, *36*, 3214–3225. [[CrossRef](#)] [[PubMed](#)]
27. Maconochie, M.K.; Krishnamurthy, R.; Nonchev, S.; Meier, P.; Manzanares, M.; Mitchell, P.J.; Krumlauf, R. Regulation of *Hoxa2* in cranial neural crest cells involves members of the *AP-2* family. *Development* **1999**, *126*, 1483–1494. [[PubMed](#)]
28. Maconochie, M.K.; Nonchev, S.; Manzanares, M.; Marshall, H.; Krumlauf, R. Differences in *Krox20*-dependent regulation of *Hoxa2* and *Hoxb2* during hindbrain development. *Dev. Biol.* **2001**, *233*, 468–481. [[CrossRef](#)] [[PubMed](#)]
29. McEllin, J.A.; Alexander, T.B.; Tümpel, S.; Wiedemann, L.M.; Krumlauf, R. Analyses of *fugu hoxa2* genes provide evidence for subfunctionalization of neural crest cell and rhombomere cis-regulatory modules during vertebrate evolution. *Dev. Biol.* **2016**, *409*, 530–542. [[CrossRef](#)] [[PubMed](#)]
30. Nonchev, S.; Maconochie, M.; Vesque, C.; Aparicio, S.; Ariza-McNaughton, L.; Manzanares, M.; Maruthinar, K.; Kuroiwa, A.; Brenner, S.; Charnay, P.; et al. The conserved role of *Krox-20* in directing *Hox* gene expression during vertebrate hindbrain segmentation. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 9339–9345. [[CrossRef](#)] [[PubMed](#)]

31. Nonchev, S.; Vesque, C.; Maconochie, M.; Seitanidou, T.; Ariza-McNaughton, L.; Frain, M.; Marshall, H.; Sham, M.H.; Krumlauf, R.; Charnay, P. Segmental expression of *Hoxa-2* in the hindbrain is directly regulated by *Krox-20*. *Development* **1996**, *122*, 543–554. [[PubMed](#)]
32. Tümpel, S.; Maconochie, M.; Wiedemann, L.M.; Krumlauf, R. 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.* **2002**, *246*, 45–56. [[CrossRef](#)] [[PubMed](#)]
33. Tümpel, S.; Cambroner, F.; Ferretti, E.; Blasi, F.; Wiedemann, L.M.; Krumlauf, R. Expression of *Hoxa2* in rhombomere 4 is regulated by a conserved cross-regulatory mechanism dependent upon *Hoxb1*. *Dev. Biol.* **2007**, *302*, 646–660. [[CrossRef](#)] [[PubMed](#)]
34. Tümpel, S.; Cambroner, F.; Sims, C.; Krumlauf, R.; Wiedemann, L.M. A regulatory module embedded in the coding region of *Hoxa2* controls expression in rhombomere 2. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 20077–20082. [[CrossRef](#)] [[PubMed](#)]
35. Parker, H.A.; Bronner, M.E.; Krumlauf, R. A *Hox* regulatory network of hindbrain segmentation is conserved to the base of vertebrates. *Nature* **2014**, *514*, 490–493. [[CrossRef](#)] [[PubMed](#)]
36. Sham, M.H.; Vesque, C.; Nonchev, S.; Marshall, H.; Frain, M.; Gupta, R.D.; Whiting, J.; Wilkinson, D.; Charnay, P.; Krumlauf, R. The zinc finger gene *Krox20* regulates *HoxB2* (*Hox2.8*) during hindbrain segmentation. *Cell* **1993**, *72*, 183–196. [[CrossRef](#)]
37. Ferretti, E.; Marshall, H.; Popperl, H.; Maconochie, M.; Krumlauf, R.; Blasi, F. Segmental expression of *Hoxb2* in r4 requires two separate sites that integrate cooperative interactions between *Prep1*, *Pbx* and *Hox* proteins. *Development* **2000**, *127*, 155–166. [[PubMed](#)]
38. Vesque, C.; Maconochie, M.; Nonchev, S.; Ariza-McNaughton, L.; Kuroiwa, A.; Charnay, P.; Krumlauf, R. *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 J.* **1996**, *15*, 5383–5396. [[PubMed](#)]
39. Barrow, J.R.; Capecchi, M.R. Targeted disruption of the *Hoxb-2* locus in mice interferes with expression of *Hoxb-1* and *Hoxb-4*. *Development* **1996**, *122*, 3817–3828. [[PubMed](#)]
40. Barrow, J.R.; Stadler, H.S.; Capecchi, M.R. Roles of *Hoxa1* and *Hoxa2* in patterning the early hindbrain of the mouse. *Development* **2000**, *127*, 933–944. [[PubMed](#)]
41. Davenne, M.; Maconochie, M.K.; Neun, R.; Pattyn, A.; Chambon, P.; Krumlauf, R.; Rijli, F.M. *Hoxa2* and *Hoxb2* control dorsoventral patterns of neuronal development in the rostral hindbrain. *Neuron* **1999**, *22*, 677–691. [[CrossRef](#)]
42. Gavalas, A.; Davenne, M.; Lumsden, A.; Chambon, P.; Rijli, F.M. Role of *Hoxa-2* in axon pathfinding and rostral hindbrain patterning. *Development* **1997**, *124*, 3693–3702. [[PubMed](#)]
43. Gavalas, A.; Ruhrberg, C.; Livet, J.; Henderson, C.E.; Krumlauf, R. Neuronal defects in the hindbrain of *Hoxa1*, *Hoxb1* and *Hoxb2* mutants reflect regulatory interactions among these *Hox* genes. *Development* **2003**, *130*, 5663–5679. [[CrossRef](#)] [[PubMed](#)]
44. Oury, F.; Murakami, Y.; Renaud, J.S.; Pasqualetti, M.; Charnay, P.; Ren, S.Y.; Rijli, F.M. *Hoxa2*- and rhombomere-dependent development of the mouse facial somatosensory map. *Science* **2006**, *313*, 1408–1413. [[CrossRef](#)] [[PubMed](#)]
45. Baltzinger, M.; Ori, M.; Pasqualetti, M.; Nardi, I.; Rijli, F.M. *Hoxa2* knockdown in *Xenopus* results in hyoid to mandibular homeosis. *Dev. Dyn.* **2005**, *234*, 858–867. [[CrossRef](#)] [[PubMed](#)]
46. Grammatopoulos, G.A.; Bell, E.; Toole, L.; Lumsden, A.; Tucker, A.S. Homeotic transformation of branchial arch identity after *Hoxa2* overexpression. *Development* **2000**, *127*, 2355–5365.
47. Kitazawa, T.; Fujisawa, K.; Narboux-Nême, N.; Arima, Y.; Kawamura, Y.; Inoue, T.; Wada, Y.; Kohro, T.; Aburatani, H.; Kodama, T.; et al. Distinct effects of *Hoxa2* overexpression in cranial neural crest populations reveal that the mammalian hyomandibular-ceratohyal boundary maps within the styloid process. *Dev. Biol.* **2015**, *402*, 162–174. [[CrossRef](#)] [[PubMed](#)]
48. Minoux, M.; Antonarakis, G.S.; Kmita, M.; Duboule, D.; Rijli, F.M. Rostral and caudal pharyngeal arches share a common neural crest ground pattern. *Development* **2009**, *136*, 637–645. [[CrossRef](#)] [[PubMed](#)]
49. Pasqualetti, M.; Ori, M.; Nardi, I.; Rijli, F.M. Ectopic *Hoxa2* induction after neural crest migration results in homeosis of jaw elements in *Xenopus*. *Development* **2000**, *127*, 5367–5378. [[PubMed](#)]
50. Hurley, I.A.; Mueller, R.L.; Dunn, K.A.; Schmidt, K.J.; Friedman, M.; Ho, R.K.; Prince, V.E.; Yang, Z.; Thomas, M.G.; Coates, M.I. A new time-scale for ray-finned fish evolution. *Proc. R. Soc. B* **2007**, *274*, 489–498. [[CrossRef](#)] [[PubMed](#)]

51. Meyer, A.; van de Peer, Y. From 2R to 3R: Evidence for a fish-specific genome duplication (FSGD). *Bioessays* **2005**, *27*, 937–945. [[CrossRef](#)] [[PubMed](#)]
52. Postlethwait, J.; Amores, A.; Cresko, W.; Singer, A.; Yan, Y.L. Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet.* **2004**, *20*, 481–490. [[CrossRef](#)] [[PubMed](#)]
53. Santini, F.; Harmon, L.J.; Carnevale, G.; Alfaro, M.E. Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. *BMC Evol. Biol.* **2009**, *9*, 194. [[CrossRef](#)] [[PubMed](#)]
54. Taylor, J.S.; van de Peer, Y.; Meyer, A. Revisiting recent challenges to the ancient fish-specific genome duplication hypothesis. *Curr. Biol.* **2001**, *11*, R1005–R1008. [[CrossRef](#)]
55. Moens, C.B.; Selleri, L. Hox cofactors in vertebrate development. *Dev. Biol.* **2006**, *291*, 193–206. [[CrossRef](#)] [[PubMed](#)]
56. Hunter, M.; Prince, V.E. Zebrafish Hox paralogue group 2 genes function redundantly as selector genes to pattern the second pharyngeal arch. *Dev. Biol.* **2002**, *247*, 367–389. [[CrossRef](#)] [[PubMed](#)]
57. Prince, V.; Lumsden, A. *Hoxa-2* expression in normal and transposed rhombomeres: Independent regulation in the neural tube and neural crest. *Development* **1994**, *120*, 911–923. [[PubMed](#)]
58. Steinke, D.; Salzburger, W.; Meyer, A. Novel relationships among ten fish model species based on a phylogenomic analysis using ESTs. *J. Mol. Evol.* **2006**, *62*, 772–784. [[CrossRef](#)] [[PubMed](#)]
59. Tümpel, S.; Wiedemann, L.M.; Krumlauf, R. Hox genes and segmentation of the vertebrate hindbrain. *Curr. Top. Dev. Biol.* **2009**, *88*, 103–137. [[PubMed](#)]
60. Davidson, E.H.; Erwin, D.H. Gene regulatory networks and the evolution of animal body plans. *Science* **2006**, *311*, 796–800. [[CrossRef](#)] [[PubMed](#)]
61. Barralho-Gimeno, A.; Holzschuh, J.; Driever, W.; Knapik, E.W. Neural crest survival and differentiation in zebrafish depends on mont blanc/ tfap2a gene function. *Development* **2004**, *131*, 1463–1477. [[CrossRef](#)] [[PubMed](#)]
62. Crump, J.G.; Swartz, M.E.; Eberhart, J.K.; Kimmel, C.B. Moz-dependent Hox expression controls segment-specific fate maps of skeletal precursors in the face. *Development* **2006**, *133*, 2661–2669. [[CrossRef](#)] [[PubMed](#)]
63. Deflorian, G.; Tiso, N.; Ferretti, E.; Meyer, D.; Blasi, F.; Bortolussi, M.; Argenton, F. Prep1.1 has essential genetic functions in hindbrain development and cranial neural crest cell differentiation. *Development* **2004**, *131*, 613–627. [[CrossRef](#)] [[PubMed](#)]
64. Le Pabic, P.; Scemama, J.; Stellwag, E.J. Role of *Hox* PG2 genes in Nile tilapia pharyngeal arch specification: Implications for gnathostome pharyngeal arch evolution. *Evol. Dev.* **2010**, *12*, 45–60. [[CrossRef](#)] [[PubMed](#)]
65. Melvin, V.S.; Feng, W.; Hernandez-Lagunas, L.; Artinger, K.B.; Williams, T. A morpholino-based screen to identify novel genes involved in craniofacial morphogenesis. *Dev. Dyn.* **2013**, *242*, 817–831. [[CrossRef](#)] [[PubMed](#)]
66. Selleri, L.; Depew, M.J.; Jacobs, Y.; Chanda, S.K.; Tsang, K.Y.; Cheah, K.S.; Rubenstein, J.L.; O’Gorman, S.; Cleary, M.L. Requirement for Pbx1 in skeletal patterning and programming chondrocyte proliferation and differentiation. *Development* **2001**, *128*, 3543–3557. [[PubMed](#)]
67. Ferretti, E.; Cambronero, F.; Tümpel, S.; Longobardi, E.; Wiedemann, L.M.; Blasi, F.; Krumlauf, R. The *Hoxb1* enhancer and control of rhombomere 4 expression: Complex interplay between PREP1-PBX1-HOXB1 binding sites. *Mol. Cell. Biol.* **2005**, *25*, 8541–8552. [[CrossRef](#)] [[PubMed](#)]
68. Pöpperl, H.; Bienz, M.; Studer, M.; Chan, S.; Aparicio, S.; Brenner, S.; Mann, R.; Krumlauf, R. Segmental expression of *Hoxb1* is controlled by a highly conserved autoregulatory loop dependent on *exd/Pbx*. *Cell* **1995**, *81*, 1031–1042. [[CrossRef](#)]
69. Urasaki, A.; Morvan, G.; Kawakami, K. 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* **2006**, *174*, 639–649. [[CrossRef](#)] [[PubMed](#)]
70. Kawakami, K. *Tol2*: A versatile gene transfer vector in vertebrates. *Genome Biol.* **2007**, *8*, S7. [[CrossRef](#)] [[PubMed](#)]
71. Kurosawa, G.; Takamatsu, N.; Takahashi, M.; Sumitomo, M.; Sanaka, E.; Yamada, K.; Nishii, K.; Matsuda, M.; Asakawa, S.; Ishiguro, H.; *et al.* Organization and structure of hox gene loci in medaka genome and comparison with those of pufferfish and zebrafish genomes. *Gene* **2006**, *370*, 75–82. [[CrossRef](#)] [[PubMed](#)]
72. Iwamatsu, T. Stages of normal development in the medaka *Oryzias latipes*. *Mech. Dev.* **2004**, *12*, 605–618. [[CrossRef](#)] [[PubMed](#)]

73. Oxendine, S.L.; Cowden, J.; Hinton, D.E.; Padilla, S. Adapting the medaka embryo assay to a high-throughput approach for developmental toxicity testing. *Neuro Toxicol.* **2006**, *27*, 840–845. [[CrossRef](#)] [[PubMed](#)]
74. Karlsson, J.; Hofsten, J.V.; Olsson, P.E. Generating transparent zebrafish: A refined method to improve detection of gene expression during embryonic development. *Mar. Biotechnol.* **2001**, *3*, 522–527. [[CrossRef](#)] [[PubMed](#)]
75. Sandelin, A.; Alkema, W.; Engstrom, P.; Wasserman, W.W.; Lenhard, B. JASPAR: An open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* **2004**, *32*, D91–D94. [[CrossRef](#)] [[PubMed](#)]
76. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [[CrossRef](#)] [[PubMed](#)]
77. Subramanian, A.; Kaufmann, M.; Morgenstern, B. DIALIGN-TX: Greedy and Progressive Approaches for Segment-Based Multiple Sequence Alignment. *Algorithms Mol Biol.* **2008**, *3*, 6. [[CrossRef](#)] [[PubMed](#)]
78. Amemiya, C.T.; Powers, T.P.; Prohaska, S.J.; Grimwood, J.; Schmutz, J.; Dickson, M.; Miyake, T.; Schoenborn, M.A.; Myers, R.M.; Ruddle, F.H.; *et al.* Complete *Hox* cluster characterization of the coelacanth provides further evidence for slow evolution of its genome. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 3622–3627. [[CrossRef](#)] [[PubMed](#)]
79. Chiu, C.; Dewar, K.; Wagner, G.P.; Takahashi, K.; Ruddle, F.; Ledje, C.; Bartsch, P.; Scemama, J.; Stellwag, E.J.; Fried, C.; *et al.* Bichir *HoxA* cluster sequence reveals surprising trends in ray-finned fish genomic evolution. *Genome Res.* **2004**, *14*, 11–17. [[CrossRef](#)] [[PubMed](#)]
80. Church, D.M.; Goodstadt, L.; Hillier, L.W.; Zody, M.C.; Goldstein, S.; She, X.; Bult, C.J.; Agarwala, R.; Cherry, J.L.; DiCuccio, M.; *et al.* Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biol.* **2009**, *7*, e1000112. [[CrossRef](#)] [[PubMed](#)]
81. Hellsten, U.; Harland, R.M.; Gilchrist, M.J.; Hendrix, D.; Jurka, J.; Kapitonov, V.; Ovcharenko, I.; Putnam, N.H.; Shu, S.; Taher, L.; *et al.* The genome of the western clawed frog *Xenopus tropicalis*. *Science* **2010**, *328*, 633–636. [[CrossRef](#)] [[PubMed](#)]
82. Kim, C.B.; Amemiya, C.; Bailey, W.; Kawasaki, K.; Mezey, J.; Miller, W.; Minoshima, S.; Shimizu, N.; Wagner, G.; Ruddle, F. *Hox* cluster genomics in the horn shark, *Heterodontus francisci*. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 1655–1660. [[CrossRef](#)] [[PubMed](#)]
83. Lee, A.P.; Koh, E.G.L.; Tay, A.; Brenner, S.; Venkatesh, B. Highly conserved syntenic blocks at the vertebrate *Hox* loci and conserved regulatory elements within and outside *Hox* gene clusters. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 6994–6999. [[CrossRef](#)] [[PubMed](#)]
84. Malaga-Trillo, E.; Meyer, A. Genome duplications and accelerated evolution of *Hox* genes and cluster architecture in teleost fishes. *Am. Zool.* **2001**, *41*, 676–686.
85. 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. Evolution of *Hox* gene clusters in gnathostomes: Insights from a survey of a shark (*Scyliorhinus canicula*) transcriptome. *Mol. Biol. Evol.* **2010**, *27*, 2829–2838. [[CrossRef](#)] [[PubMed](#)]
86. Raincrow, J.D. *Hox* Cluster Intergenic Sequence Evolution. Ph.D. Thesis, Rutgers University, NJ, USA, 2010.
87. Goode, D.K.; Callaway, H.A.; Cerda, G.A.; Lewis, K.E.; Elgar, G. Minor change, major divergence: Divergent functions of highly conserved *cis*-regulatory elements subsequent to whole genome duplication events. *Development* **2011**, *138*, 879–884. [[CrossRef](#)] [[PubMed](#)]
88. Santagati, F.; Minoux, M.; Ren, S.Y.; Rijli, F.M. Temporal requirement of *Hoxa2* in cranial neural crest skeletal morphogenesis. *Development* **2005**, *132*, 4927–4936. [[CrossRef](#)] [[PubMed](#)]
89. Scemama, J.L.; Hunter, M.; McCallum, J.; Prince, V.; Stellwag, E. Evolutionary divergence of vertebrate *Hoxb2* expression patterns and transcriptional regulatory loci. *J. Exp. Zool. B Mol. Evol. Dev.* **2002**, *294*, 285–299. [[CrossRef](#)] [[PubMed](#)]
90. Carroll, S.B. Evo-Devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* **2008**, *134*, 25–36. [[CrossRef](#)] [[PubMed](#)]
91. Capellini, T.D.; di Giacomo, G.; Salsi, V.; Brendolan, A.; Ferretti, E.; Srivastava, D.; Zappavigna, V.; Selleri, L. Pbx1/Pbx2 requirement for distal limb patterning is mediated by the hierarchical control of *Hox* gene spatial distribution and *Shh* expression. *Development* **2006**, *133*, 2263–2273. [[CrossRef](#)] [[PubMed](#)]
92. Kmita, M.; Tarchini, B.; Zakany, J.; Logan, M.; Tabin, C.J.; Duboule, D. Early developmental arrest of mammalian limbs lacking *HoxA/HoxD* gene function. *Nature* **2005**, *435*, 1113–1116. [[CrossRef](#)] [[PubMed](#)]

93. Mercader, N.; Leonardo, E.; Azpiazu, N.; Serrano, A.; Morata, G.; Martinez, C.; Torres, M. Conserved regulation of proximodistal limb axis development by Meis1/Hth. *Nature* **1999**, *402*, 425–429. [[PubMed](#)]
94. Tamura, K.; Yokouchi, Y.; Kuroiwa, A.; Ide, H. Retinoic acid changes the proximodistal developmental competence and affinity of distal cells in the developing chick limb bud. *Dev. Biol.* **1997**, *188*, 224–234. [[CrossRef](#)] [[PubMed](#)]
95. Tamura, K.; Yonei-Tamura, S.; Yano, T.; Yokoyama, H.; Ide, H. The autopod: Its formation during limb development. *Dev. Growth Differ.* **2008**, *50*, S177–S187. [[CrossRef](#)] [[PubMed](#)]
96. Zeller, R. The temporal dynamics of vertebrate limb development, teratogenesis and evolution. *Curr. Opin. Genet. Dev.* **2010**, *20*, 384–390. [[CrossRef](#)] [[PubMed](#)]



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).