HOX CLUSTER INTERGENIC SEQUENCE EVOLUTION

By

JEREMY DON RAINCROW

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

and

The Graduate School of Biomedical Sciences

University of Medicine and Dentistry of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Cell and Developmental Biology

written under the direction of

Chi-hua Chiu

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October 2010

ABSTRACT OF THE DISSERTATION

HOX GENE CLUSTER INTERGENIC SEQUENCE EVOLUTION

by JEREMY DON RAINCROW

Dissertation Director:
Chi-hua Chiu

The *Hox* gene cluster system is highly conserved among jawed-vertebrates. Specifically, the coding region of *Hox* genes along with their spacing and occurrence is highly conserved throughout gnathostomes. The intergenic regions of these clusters however are more variable. During the construction of a comprehensive non-coding sequence database we discovered that the intergenic sequences appear to also be highly conserved among cartilaginous and lobe-finned fishes, but much more diverged and dynamic in the ray-finned fishes. Starting at the base of the Actinopterygii a turnover of otherwise highly conserved non-coding sequences begins. This turnover is extended well into the derived ray-finned fish clade, Teleostei. Evidence from our population genetic study suggests this turnover, which appears to be due mainly to loosened constraints at the macro-evolutionary level, is highlighted by evidence of strong positive selection acting at the micro-evolutionary level. During the construction of the non-coding sequence database we also discovered that along with evidence of both relaxed constraints and positive selection emerges a pattern of transposable elements found within the *Hox* gene cluster system. The highly conserved Chondrichthyes and Sacropterygii *Hox* gene clusters have an invasion of type I transposons whereas the Actinopterygii *Hox* gene clusters have an invasion of type II transposons. Specifically,

the Tc1 transposon is found throughout the ray-finned fishes *Hox* gene clusters and is highlighted by the presence of two intact Tc1 transposons in and adjacent to bichir's *Hox* gene clusters. Expression in human cell lines suggests that at least one of these Tc1 transposons are active. This combined with simulations ran in our lab point to transposons having a role in past and on-going restructuring of ray-finned fishes genomes.

These findings help shed light on the possible genomic changes that occurred and are occurring within the ray-finned fish clade that help shed light on their past and present species radiations.

Dedication

To

Jim and Patsy Brown

for their tireless support throughout my education

To

Wilder and Angelica Pajares

for giving me a home away from home

To

Maricruz Pajares

for keeping my spirits high

To

Kim Harris

for starting me on the right path

IV. Table of Contents

V. List of Tables

VI. List of Illustrations

Introduction- The Unrivaled Teleosts and the Model Meta-Gene

1.1 The Unrivaled Teleosts

1.1.1 Speciose Teleosts

Teleost fishes (division: **Teleostei**) are a highly successful clade both from the species and morphological perspective. Teleosts currently contain ~27,000 species that comprise nearly one-half of all jawed-vertebrates (superclass: **Gnathostomata**) and 99% of all ray-finned fishes (class: **Actinoptergyii**; Nelson 2006). These species are divided into 38 orders with 426 families and 4,084 genera (Nelson 2006).

Along with this taxonomic success is a stunning array of adaptations. Some of the more striking adaptations include fishes that have secondarily lost paired appendages (true eels; **Anguilliformes**), have the ability to inflate their bodies to twice the normal size (pufferfishes; **Tetraodontidae**), have the ability to create their own light (flashlight fish, *Photoblepharon palpebratus*) and even produce electrical fields strong enough to stun or kill (electric ray; **Torpedinidae**, electric eel; **Electrophoridae**, stargazer; **Uranoscopidae**). Teleosts range in size from a 7.9 mm cyprinid (*Paedocypris progenetica*, Kottelat et al. 2006) to a 2-ton ocean sunfish (*Mola mola*, McCann 1961).

The anterior dorsal fin is an example of the variety of adaptations that can be seen even in a single anatomical structure. This single fin has been modified into a lure (**Lophiidae**), camouflage (**Antennariidae**), suction disks (**Echeneidae**), and venom delivery (**Synanceja**, **Trachinidae**, **Thalassophryninae**). As a prime example of the oddity of adaptation seen in teleosts; a recent and very unique phenotype was discovered in the Monterey Bay off the coast of California. A close relative of salmonids, *Macropinna microstoma* is an opisthoproctid fish with a clear skull and rotating tubular

eyes that let it take in the dim light from above as well as perform its front-facing predatory role (Robison and Reisenbichler 2008).

Along with anatomical diversity, teleosts also maintain increased variation at the cellular level. One example is coloration. Teleost fishes have a highly complex pigmentation patterning system that involves many pigment cell types known as chromatophores which are all derived from the neural crest. They come in 6 varieties including melanophores (black/brown), xanthophores (yellow/orange), iridophores (iridescent), erythrophores (red), leucophores (white) and even cyanophores (blue; Goda and Fujii 1995). This is in stark contrast to the single pigment cell type in mammals and birds (melanocytes; black/brown/yellow). This increase in chromatophores in teleosts is thought to be partly due to the duplication of color gene pathways both from whole genome duplications as well as lineage specific gene duplications (Braasch et al. 2006, 2009).

1.1.2 Teleost Specific Morphology

Teleost fishes have undergone rapid speciation coincident with an array of adaptive morphologies (Alfaro et al. 2009). They are one of if not the most diverse and successful extant vertebrate lineages. What conditions in the evolutionary history of actinopterygians (ray-finned fishes) underlie both their speciosity and morphological variability?

One hypothesis is that the acquisition of a few shared, advantageous adaptations early in teleost evolution may be necessary to maintain an ecological foothold while facilitating unique adaptations to a variety of habitats resulting in morphological variability (Figure 3; Hulsey et al. 2006). Such a scenario could lead to the phenomenon

of adaptive morphological plasticity which we define as a species or group of species that appears to rapidly adapt to a variety of habitats. This plasticity is exemplified by the filling of several ecological niches by cichlid fishes in two separate occasions in Lake Malawi, Lake Tanganyika and Lake Victoria of Africa. All of these rapid speciations occurred independently and in different yet short evolutionary time frames (summary, Koblmüller et al. 2008). Do teleost fishes share such a set of advantageous adaptations?

Teleost fishes diverged from there closest living relative ~250 mya (Figure 2; Hurley et al. 2007). Since that time teleost fishes have evolved into fast, agile fishes able to exploit an almost limitless variety of food sources and habitats. The main adaptations that have contributed to their success have been a reduction in the number and thickness of bones, the amount of bony material in the scales, and fin position and composition allowing for greater flexibility, speed and maneuverability leading to a more dynamic predator/prey relationship (Janvier 1996). Perhaps as important as the increased locomotion is the modification of the feeding structures (Hulsey et al. 2006).

Teleosts have lost many bones in the skull and jaw allowing for the evolution of the extendable front/marginal/oral jaw. The extension of this oral jaw causes negative pressure within the mouth sucking in surrounding water and its contents allowing the fish to better capture prey (Schaeffer and Rosen 1961). However, the oral jaw reduces significantly the bite force and hence the ability of fishes to eat foodstuff containing hard materials such as the chitinous shell of the arthropods and the cellulose cell wall of most plants (Schaeffer and Rosen 1961; Liem 1973). Teleost fishes have developed a second pair of jaws known as pharyngeal jaws that are able to process these harder foodstuffs

(Schaeffer and Rosen 1961). Pharyngeal jaws evolved from a dentin surface on the pharynx and eventually developed musculature (Figure 1).

It is hypothesized that the development of the pharyngeal jaw and subsequent partitioning of duties between the oral and pharyngeal jaws has led to a decoupling of their evolution (Liem 1973). This is known as Liem's hypothesis and is thought to be a major reason for the successful species radiation of the teleost division. It is also possible that novel and useful adaptations such as the pharyngeal jaw may be necessary to allow species/groups, such as the teleosts, to maintain a morphological robustness while allowing unique adaptations to a variety of habitats resulting in adaptive morphological plasticity. This would cause a delay in the rapid speciation events following the key innovation as the new innovation allows the species to move into new habitats and increase in number and then subsequently adapt to those habitats resulting in rapid speciation. Alfaro et al. (2009) found this delay between a pharyngeal jaw innovation and rapid speciation in their labrid (Teleostei:Perciformes) fish.

The pharyngeal jaw and other unique features have allowed teleosts to exploit an almost endless variety of habitats (Liem 1973) including water temperatures of -2°C at the polar caps (DeVries and Wohlschlag 1969) to 41°C at the African hot springs (Johnston et al. 1994), from elevations of 3.8 km in the Andean Lake Titicaca (Arratia 1982) to depths of 7.7 km in an ocean trench off the shore of Japan (Jamieson et al. 2009) and various pH, saline, and dissolved oxygen concentrations.

1.1.3 Environmental Change

Another driving factor in teleosts success and uniqueness is their environment. Environmental pressures are known to expedite evolution through natural selection. Were there any environmental factors that may have led to the current state of teleosts?

Teleost fishes' fossil record dates back to the beginning of the Triassic period ~250 million years (my) (Hurley et al. 2007). Shortly after, continental drift was occurring that would lead to the break up of the super continent Pangea into Gondwana and Laurasia (~150 mya, Smith et al. 1994). Eventually (~100 mya) Gondwana and Laurasia would break up into the current continents that have maintained a relatively steady shape over the last 65 my. Plate tectonics occurring during the evolution of the teleost fishes were opening up 1,000s of miles of new shoreline for ocean life. This ocean life includes the evolution of modern corals that host nearly one-half of all extant teleost species and is postulated to have been an aide in the overall diversification of the teleost division (Bellwood and Wainwright 2002; Alfaro et al. 2007). Several mass extinctions also occurred during the evolution of the teleost fishes that would lead to less competition for habitat and resources for those surviving individuals. This may have allowed teleosts to reach great numbers that in turn created competition within the clade driving diversity from within while simultaneously suppressing the success/diversification of other clades.

Coincident with the dawning of the teleost fish division was the largest known mass extinction event in the history of Earth, the Permian-Triassic extinction. During this mass extinction event ~95% of all marine life went extinct. Another major extinction occurred at the end of the Triassic (~200 mya) killing off all marine reptiles except for

Ichthyosaurus. A third mass extinction (Cretaceous-Palaeogene K-P mass extinction) occurred at the end of the Cretaceous period (~65 mya) again resulting in the extinction of all marine reptiles except turtles. Friedman et al. (2010) was able to show that this mass extinction may have led to the Acanthomorph rapid species radiation by allowing the Acanthomorphs to fill the morphological space left void by the extinction of non-teleost vertebrates.

It is unclear if teleosts survived these mass extinctions in greater numbers than other lineages or if they were able to take advantage with the few species that were left. Crow et al. (2005) suggest that the fish specific genome duplication (see section 1.1.4) may have acted as a buffer against extinction for those surviving teleosts. However, Santini et al. (2009) estimated that despite an increase in overall diversification rates in teleosts, teleosts have a 6 times greater extinction rate than non-teleosts gnathostomes. This leads to a one and one-half greater death:birth ratio (of species) in teleosts as opposed to non-teleost gnathostomes. Regardless, teleosts are currently the most successful extant vertebrate lineage. There were also other ray-finned fish lineages that existed prior to the MRCA of living teleosts that were species rich as well that did not survive (Donoghue and Purnell 2005).

Unique and advantageous adaptations, mass extinctions of competitors and vast amounts of new habitat available for niche-filling have helped shape the explosion of teleost fishes' numbers and diversity of species. Are environmental factors and well-timed advantageous adaptations enough or do teleosts fishes have an inherent molecular structure that allows for their success?

1.1.4 Fish Specific Genome Duplication

This species and morphological diversity in teleost fishes has been aided by or at least evolved along side not only environmental factors but also changes in genomic architecture and accelerated molecular evolution. A major genomic change that occurred prior to the divergence of extant teleost fishes is a whole genome duplication dubbed the fish specific genome duplication (FSGD; Meyer et al. 1999, 2005). The best evidence that the FSGD is a whole genome duplication comes from synteny maps from the completed genome of the spotted green pufferfish (*Tetraodon nigroviridis)* showing this taxon has two syntenic regions (paralogons) that correspond to single regions in the human genome (Jaillon et al. 2004). Comparative mapping shows that paralogons of spotted green pufferfish and zebrafish (*Danio rerio*) are homologous, supporting that the FSGD occurred prior to the divergence of the major teleost clades, **Acanthopterygii** and **Ostariophysi** (Woods et al. 2005).

Our labs recent work on *Hox* genes in the goldeye (*Hiodon alosoides*) suggests that **Osteoglossomorpha**, the most basal teleost subdivision, also shares the FSGD with higher teleost fishes (Chambers et al. 2009). Because of the importance of genome duplication in vertebrate evolution (Ohno, 1970; Sidow, 1996) many authors have speculated that the FSGD is directly responsible for the biological diversification (i.e. speciosity) of teleost fishes (Hoegg et al. 2004; Postlethwait et al. 2004; Meyer and Van de Peer 2005).

There are however several lines of evidence that the FSGD is not even coincidental with the teleost species radiation much less a causal agent. Recent

examination of the ray-finned fishes fossil record, shows that there are 11 extinct clades between teleosts and their closest living relatives (Donoghue and Purnell 2005). The authors conclude that the character acquisitions normally attributed as synapomorphies of derived teleost fishes arose gradually in ray-finned fish phylogeny with many acquisitions predating the FSGD. Additionally, many of these extinct clades that have been shown to pre-date the FSGD were species rich themselves. Hence fossil evidence suggests that the FSGD is uncoupled to species richness.

Molecular clock analysis between human (*Homo sapiens*) and Japanese pufferfish (*Takifugu rubripes)* estimates the FSGD occurred ~350 mya (Christoffels et al. 2004). Evidence from *Hox* (Crow et al. 2006), Para*Hox* (Mulley et al. 2006), and three nuclear, non-*Hox* genes (Hoegg et al. 2004) supports that the FSGD is coincident with the origin of teleosts in the osteoglossomorphs, after divergence of the bowfin (*Amia calva*). Bowfin fossils suggest this node is ~250 million years old (Hurley et al 2007). Importantly, our work with an Osteoglossomorph has shown that possession of a duplicated genome is not sufficient for species richness (Chambers et al. 2009). It is also important to note that the Actinoptyerygian super-orders (**Ostariophysi** and **Acanthopterygii**) that contain 93% of the present-day teleost species diverged ~135 mya (Benton 2005), much later in ray-finned fish phylogeny than when the FSGD occurred (250-350 mya). Furthermore it has been shown that these species rich super-orders haven't retained many genes as duplicates from the FSGD.

In fact, it is estimated that only ~10% and ~15% of paralogous genes in zebrafish (**Ostariophysi**) and spotted green pufferfish (**Acanthopterygii**) genomes, respectively are a result of the FSGD (Blomme et al. 2006). Interestingly, for both species the

majority of paralogous genes result from lineage specific duplications (~40% of paralogs in zebrafish and ~17% in green-spotted pufferfish (Blomme et. al. 2006). Hence these data demonstrate that the FSGD likely did not play a large role in the diversification of teleost fishes but does suggest gene duplication in general may be important.

1.1.5 Accelerated Evolution

If not the FSGD what was the major molecular contributor to the diversification of the teleost fishes? It is possible that accelerated evolution due to increased mutation rate and increased genetic recombination are factors.

Both coding and non-coding sequences have been shown to be under accelerated evolution in the teleost division of ray-finned fishes. Robison-Rechavi and Laudet (2001) set out to test the Ohno model that one of two duplicate genes should be less constrained by selection and thus be able to acquire a new function (and presumably an accelerated evolution). What they found was that 16/19 genes in fish regardless of their duplication status showed an overall accelerated evolution when compared to mammals. Growth hormone genes have also been found to be under accelerated evolution (5X more substitutions) in teleosts compared to mammals and this acceleration has been shown to be a product of both relaxation of purifying selection and positive selection (Ryynanen and Primmer 2006).

On a genome-wide scale Jaillon et al. (2004) showed that the higher mutation rate held true for 5,082 protein-coding genes when comparing spotted green pufferfish and Japanese pufferfish divergence to the mammalian proxy of human (*Homo sapiens*) and the house mouse (*Mus musculus).* For the 15-25% of genes retained as pairs there is often an asymmetric acceleration of nucleotide substitution (synonymous and

nonsynonymous) rate in one of the duplicated paralogs (Wagner et al. 2005; Braasch et al. 2006; Brunet et al. 2006; Crow et al. 2006; Steinke et al. 2006). This pattern is also lineage specific and, for *Hox* genes, also cluster specific (i.e. the *Hox*Aa versus *Hox*Ab clusters of teleosts, Chiu et al. 2004; Wagner et al. 2005).

One signature of molecular adaptation is Ka/Ks >1 (Nei and Gojobori 1986). This means that there is a greater frequency of non-synonynomous changes than synonynomous changes which implies that non-synonomous changes are preferentially fixed in the population. Investigations of duplicated paralogs in teleost fishes have found one paralog to show evidence of positive Darwinian selection; this asymmetry is species specific (Steinke et al. 2006; Brunet et al. 2006). Interestingly, Brunet et al. (2006) suggest that this asymmetric pattern of Darwinian evolution observed is not a result of changes in selective patterns after the FSGD (Vogel 1998; Meyer and Schartl 1999; Volff 2005). Instead, they propose that genes already under strong selective pressure were preferentially maintained as duplicates. Implying that duplication of genes that are advantageous to adaptation leads to greater adaptive success. This is the first insight into possible molecular mechanisms for the success of the teleosts, the accelerated evolution and possibly positive selection on groups of genes in ray-finned fishes prior to the divergence of teleosts that were also advantageous after their divergence.

This accelerated evolution in the teleost fishes has also been observed in conserved non-coding sequences (CNS). Work done by Chiu et al. (2002) showed that conserved non-coding sequences found between human and horn shark (*Heterodontus francisci*) *Hox*A clusters were more often than not absent in the teleost fishes, zebrafish (*Danio rerio*) *Hox*Aa and *Hox*Ab clusters and striped bass (*Morone saxatilis*) *Hox*Aa

cluster. Venkatesh et al. (2007) showed that this loss of CNSs occurred at the genome wide scale as well. They were able to show that between human and elephant shark (*Callorhinchus milii*) there were 4,782 CNSs whereas between human and zebrafish there were only 2,838 and between human and Japanese pufferfish 2,107. This turnover of CNSs in the teleost fish division can be postulated to result in differing expression patterns since CNSs are predicted *cis*-regulatory elements. However conservation of putative or known *cis*-regulatory elements does not always coincide with conserved function. Fisher et al. (2006) show human and zebrafish RET-1 *cis*-regulatory elements have equivalent function in cross-species expression studies, despite no sequence similarity. More research in investigating the effect of *cis*-regulatory element evolution on gene expression changes in ray-finned fishes is urgently needed.

Although the exact mechanisms of the teleosts' success are unknown their radiation is unmatched among extant vertebrates. It is true that fossil records show other lineages that may have been as successful as teleosts (Donoghue and Purnell 2005), however understanding mechanisms that led to the radiation and quite possibly continued radiation of the extant teleosts will give insight into the environmental, phenotypic, and genetic conditions necessary for successful species radiations.

1.2 The Model Meta-Gene

1.2.1 Discovery of *Hox* Genes

In order to gain more insight into possible molecular mechanisms that have contributed to the teleosts rapid species radiation I focus my study on the *Hox* genes. First, I will give an introduction into *Hox* gene discovery and evolution as well as their clustered nature.

Homeotic mutations are those that transpose whole body parts or segments. Homeotic mutations were first officially described in 1915 in the fruit fly (*Drosophila melanogaster;* Bridges and Morgan 1923). In an effort to isolate the genes responsible for these mutations E.B. Lewis devoted a career to studying them in *D. melanogaster*. In the 1950s, 1960s and 1970s Lewis was able to isolate several clustered homeotic genes that were termed the bithorax complex (BX-C). In a review in 1978 Lewis eloquently laid out a proposal for the rules governing these genes and their expression where each gene had an anterior but not posterior boundary and the expression of each new gene caused the next thoracic or abdominal segment to take on a unique characteristic.

Lewis was incorrect about the number of protein coding genes that were contained within this complex but future insight discovered that micro RNA genes also lie within this complex and help regulate the anterior-posterior patterning in *Drosophila* and thus implementing them as major contributors to the insect body plan (review, Chopra and Mishra 2006). The BX-C contains 3 homeotic genes (Sanchez-Herrero et al. 1985) or as they are known in *Drosophila* Hom-C genes (Akam 1989). The rest of the Hom-C genes were described in a cluster by Kaufman et al. (1980) and were coined the Antennapedia complex or ANT-C. Together the BX-C and ANT-C contain 8 Hom-C

genes that are responsible for patterning the anterior-posterior axis of *Drosophila melanogaster*.

The homeotic genes are now defined by a conserved sequence known as the homeobox. McGinnis et al. (1984a) first described the homeobox as a repetitive DNA sequence that was conserved among genes in the BX-C and ANT-C. This 180bp homeobox codes for the 60 amino acid homeodomain. The homeodomain was considered evidence that homeotic genes were transcription factors due to the homology with the *MAT-a1* and *MAT-α2* genes of yeast that are known transcription factors that regulate genes responsible for mate type switching (Shepherd et al. 1984). Evidence of homeotic mutations in vertebrates and the discovery of the conserved homeobox sequence led to rapid isolation of homeotic genes in other animals (chicken; *Gallus gallus*, mouse and human McGinnis et al. 1984b; McGinnis et al. 1984c).

That same year an entire homeobox gene was cloned from the African clawed frog (*Xenopus laevis*) and it was found to have 55 of 60 amino acids conserved in the homeodomain with *D. melanogaster's Antennapedia* (Carrasco et al. 1984) and a homeobox gene from mouse was cloned in the same lab (McGinnis et al. 1984c). Because of the diversity of animal types *Hox* genes were found in, they were looked for in more and more animals to help establish their universality. This was important due to their possible implication as major contributors to the embryological and evolutionary patterning of all animal bauplans.

1.2.2 Evolution of *Hox* Genes

Hox genes are part of a class of helix-turn-helix proteins that are known to bind to DNA and help regulate transcription. The helix-turn-helix motif is one of the most

conserved (oldest) motifs known in living organisms. Giraldo and Diaz-Orejas (2001) found that RepA, which is responsible for DNA replication initiation in *Pseudomonas* (bacteria) and contains a helix-turn-helix motif, was homologous to Orc4p in *Saccharomyces cerevisiae* (eukaryote) and Cdc6p in *Pyrobaculum aerophilum* (archaea) in both sequence and structure making them likely orthologs. This places the origin of the helix-turn-helix motif at or before the most recent common ancestor of all known living organisms.

As mentioned previously, the *Hox*-related family is determined specifically by the presence of the homeodomain, a 60 amino-acid motif. Proteins containing this domain can be found as deep in animal phylogeny as **Cnidaria** (Murtha et al. 1991). Although, to be a true *Hox* family member the gene must be part of a cluster as they are in the BX-C and ANT-C in *Drosophila*. So far *Hox* clusters have been discovered in various states of clustering in both protostomes and deuterostomes, which would date the ancestral *Hox* cluster as far back as the origin of the **Bilateria** (Figure 2).

The **Radiata** (**Cnidaria** and **Ctenophora**) contain *Hox*-related genes but there is no decisive proof that they are clustered in the genome (review; Aboobaker and Blaxter 2003). Among **Bilateria** the oldest known members to contain *Hox* genes would be the phylum **Acoelomorpha**. *Hox* gene orthologs have been found in both classes of **Acoelomorpha**, **Acoela** (Cook et al. 2004) and **Nemertodermatida** (Jimenez-Guri et al. 2006). *Hox* genes have also been found in the other basal Bilaterian phylum, **Chaetognatha** (Papillon et al. 2003). Again it is important to point out that there is no indication that these *Hox* genes are clustered in these sister groups to the superphyla of **Protostomia** and **Deuterstomia**.

In protostomes *Hox* genes have been found in the phyla **Nematoda**, **Onychophora** (Grenier et al. 1997), **Arthropoda**, **Nemertea** (Kmita-Cunisse et al. 1998), **Platyhelminthes** (overview, Badets and Verneau 2009), **Bryozoa** (Passamaneck and Halanych 2004), **Annelida** (Bleidorn et al. 2009), and **Mollusca** (Degnan and Morse 1993). The nematode *Caenorhabditis elegans* has *Hox* genes but they are not clustered in its genome (review; Aboobaker and Blaxter 2003). This would indicate that the *Hox* cluster was secondarily dispersed in this lineage. The model arthropods *Drosophila* have *Hox* genes in clusters both complete and broken. For deuterostomes *Hox* genes have been found in the phyla **Echinodermata** (Dolecki et al. 1988), **Hemichordata** (Peterson 2004) and **Chordata**. For chordates the first *Hox* genes were discovered in the model organisms chicken, mouse and human (McGinnis et al. 1984b). They have also been found in more basal taxa including **Urochordata** and **Cephalochordata** (Bell et al. 1993; Holland et al. 1992).

Amphioxus (*Branchiostoma floridae*) is thought to contain the ancestral *Hox* gene cluster for the **Gnathostomata** (jawed vertebrates; Garcia-Fernandez and Holland 1994). **Gnathostomata** can be divided into 3 major classes: **Chondrichthyes** (cartilaginous fishes), **Sarcopterygii** (lobe-finned fishes) and **Actinopterygii** (ray-finned fishes). These classes all share the first instance of multiple (4) paralogous *Hox* gene clusters (Kim et al., 2000; Chiu et al., 2002; 2004; Prohaska et al., 2004; Raincrow et al. submitted). These are thought to have arisen via two rounds of genome duplication (Holland et al. 1994) but the cluster count of the intermediate species (lampreys and hagfish) is undetermined. The *Hox* genes in chordates are highly conserved and paralogs are easily discernible. This conservation has proven as useful for phylogenetics in chordates as it

has in determining membership in the protostome superphyla of **Lophotrochozoa**, **Ecdysozoa** and **Platyzoa**. Central to the conservation of *Hox* genes in chordates is the conservation of their clustered formation. Clustering is yet one unique aspect of *Hox* genes.

1.2.3 Evolution of the *Hox* Gene Cluster

As mentioned previously, the first *Hox* gene clusters were discovered in *Drosophila* by E.B. Lewis (1978). Lewis first discovered what is now known as the BX-C or bithorax cluster. The BX-C contains *Hox* genes ultrabithorax, abdominal-A and abdominal-B. The other *Hox* gene cluster found in *Drosophila* was known as the ANT-C or antennapedia cluster. The ANT-C was discovered by Kaufman et al. (1980) and contains *Hox* genes labial (Lab), proboscipedia (Pb), deformed (Dfd), sex combs reduced (Scr) and antennapedia (Ant). The first cluster of *Hox* genes in a vertebrate were found in mouse on chromsome 11 (Hart et al. 1985).

The most primitive animal with an intact *Hox* gene cluster found to date is the nemertean ribbonworm (*Lineus sanguineus*). This ribbonworm has a cluster containing 6 *Hox* genes (Kmita-Cunisse et al. 1998). Other more derived protostomes such as arthropods have evidence of both an intact (red flour beetle; *Tribolium castaneum*, Brown et al. 2002; Shippy et al. 2008, mosquito; *Anopheles gambiae*, Powers et al. 2000) as well as divided (*D. melanogaster*, silk moth; *Bombyx mori*, Yasukochi et al. 2004) *Hox* gene cluster, whereas more primitive protostomes such as nematodes (Burglin and Ruvkun 1993) and platyhelminths (Pierce et al. 2005) have dispersed *Hox* gene clustering.

The same variation in *Hox* gene clustering can be seen in deuterostomes but not to the same extent. The sea urchin (*Strongylocentrotus purpuratus*) has an intact *Hox* gene

cluster but the genes are not in the canonical order (Cameron et al. 2006). Amphioxus, a cephalochordate, contains an intact *Hox* gene cluster that contains 14 *Hox* genes that are all transcribed from the same strand (Garcia-Fernandez and Holland 1994; Ferrier et al. 2000). This is thought to be the archetypal state of *Hox* gene clusters for the chordates. After the divergence of Amphioxus and prior to the divergence of known gnathostomes it is postulated that two rounds of genome duplication occurred giving rise to 4 *Hox* gene clusters (Amores et al. 2004). For accounting purpose Scott (1993) developed a nomenclature system that assigns each of the four paralogous clusters a letter (A, B, C, D) and orthologs to each of the 14 *Hox* genes found in Amphioxus 1-14. For example the gene orthologous to the 9th gene on Amphioxus's *Hox* gene cluster will be known as *Hox*A9 on the A cluster, *Hox*B9 on the B cluster and so on. Each of these clusters is ordered the same as the Amphioxus *Hox* gene cluster but none of them contain the entire complement of 14 genes.

To date intact *Hox* gene clusters homologous to the 4 *Hox* gene clusters have been found in the cartilaginous fish horn shark (*Heterodontus francisci*, Kim et al. 2000), the lobe-finned fishes coelacanth (*Latimeria menadoensis*, Koh et al. 2003; Danke et al. 2004), Western clawed frog (*Xenopus tropicalis*), the domesticated chicken (Richardson et al. 2007), human (Acampora et al. 1989), mouse (Duboule and Dolle 1989), the domesticated dog (*Canis familiaris*), the ray-finned fish, Senegal bichir (*Polypterus senegalus*) and other species with whole genome sequences.

Several gnathostomes lineages have had further genome duplications or tetraploidization events but none of these occurred during a time that allows them to be both shared by many species and still be easily detectable. Over 27,000 species share the

FSGD and all species sequenced to date appear to have between 7-8 intact *Hox* gene clusters (Figure 3). For accounting purposes these *Hox* gene clusters follow the Scott (1993) nomenclature with 2 paralogous clusters per original gnathostome cluster (A > Aa, Ab; B > Ba, Bb; C > Ca, Cb; D > Da, Db). To date only fishes of the division **Teleostei** have been found to share the FSGD and it has been proposed that the FSGD is coincident with the base of the teleosts group although there is much evidence to the contrary. There are two main groups of teleosts based on numbers of species: superorder **Ostariophysi** (6,502 species) and superorder **Acanthopterygii** (13,421 species). Only a single fish from **Ostariophysi** has a complete genome and/or complete *Hox* gene cluster complement sequenced; zebrafish. Several species of **Acanthopterygii** have been sequenced including medaka (*Oryzias latipes;*Kurosawa et al. 1999; Naruse et al. 2004), 3-spine stickleback (*Gasterosteus aculeatus;* Hoegg et al. 2007), spotted green pufferfish (Jaillon et al. 2004), Japanese pufferfish (Kurosawa et al. 2006), Nile tilapia (*Oreochromis niloticus;* Santini and Bernardi 2005) and African cichlid (*Astatotilapia burtoni;* Hoegg et al. 2007). So far the two separate superorders can be differentiated based on their *Hox* gene cluster content as Ostariophysi has Aa, Ab, Ba, Bb, Ca, **Cb** and Da clusters and **Acanthopterygii** has Aa, Ab, Ba, Bb, Ca, Da and **Db**. This 8 *Hox* gene cluster architecture is considered the ultimate in *Hox* gene cluster evolution although the number of genes is only slightly higher than states prior (i.e. mouse 39; *Tetraodon nigroviridis* 48). Despite the slight increase in *Hox* gene number they are more tightly clustered and more modular in nature (Duboule 2007), which can lead to a greater flexibility of expression and co-option.

Known *Hox* gene cluster states throughout animal phylogeny leads to the question of what the ancestral state of the *Hox* gene cluster was and why the *Hox* gene cluster exists. There are 2 major theories on the ancestral state of the metazoan *Hox* gene cluster, but they both result in the following groupings. First, the *Hox* gene cluster is divided into 4 major categories: anterior, group 3, central and posterior genes. Anterior genes include group 1 and 2 of deuterostomes and Lb and Pb of protostomes. Group 3 genes appear to have an ancient origin and have been secondarily lost in protostomes, central genes include group 4-8 in deuterostomes and Dfd, Scr, Antp, Ubx and Abd-A in invertebrates and posterior genes include group 9-14 in deuterostomes and Abd-B in protostomes. The origin of each category is thought to lie in a single *Hox* gene. The origin of these 4 *Hox* genes is in question. There are competing theories that there were 2, 3 or 4 original genes. For the 2 and 3 gene models the remaining genes were a result of tandem duplication of a more ancient *Hox* gene. A review of these models can be found in Garcia-Fernandez (2005).

The function of the *Hox* gene cluster has been a topic of debate. It is known that for species with an intact cluster that they show spatiotemporal colinearity with their expression along the anterior-posterior axis of the developing embryo (McGinnis and Krumlauf, 1992; Lufkin, 1996). For this reason it was postulated that the positioning of the *Hox* genes along the cluster and their tight cluster formation, which excludes any other genes, was necessary (although with an unknown mechanism) for proper spatiotemporal patterning. This is not true however as found for animals that do not have intact *Hox* gene clusters yet still maintain spatial patterning such as the fruit fly, the nematode and a tunicate (Seo et al. 2004; review in Ferrier and Minguillon 2003). As

described by Monteiro and Ferrier (2006), these species all share rapid development that does not lend itself to the temporal aspect of gene induction. This still allows clustering to be necessary for temporal patterning however. Chambeyron and Bickmore (2004) hypothesized that there could be a structural component to the cluster where the orientation and proximity were needed for proper expression pattern. They were able to show that the anterior portion of the *Hox*B cluster in mice (*Hox*B1-9) underwent chromatin decondensation in response to a known *Hox* gene inducer, retinoic acid. Then in sync with their expression pattern each gene "looped out" from the chromatin structure towards the center of the nucleus. It is possible that the pressures necessary to maintain the cluster formation are different in different lineages and we may never know what the original pressure was. The loss of the clustering formation in some species indicates that it is not essential for bilateral life but its widespread occurrence at least indicates that it is extremely advantageous.

The widespread occurrence of *Hox* genes, their clustered nature and role in the formation of the body plan make them an ideal candidate to study differences and similarities between wide-ranging species and, as we will show in chapter 4, between closely related species. This is especially true in teleosts since many *Hox* genes have been maintained as duplicates and since accelerated and asymmetric rates of nucleotide evolution are thought to be one molecular contribution to the success of this clade.

Chapter 2- Closing the Gap between Teleosts and Non-Actinopterygians

2.1 Introduction

2.1.1 Ray-finned Fish Phylogeny

Comparative genetics has been widely used to help decipher the function of the vast amounts of non-coding DNA in the genome (Lee et al. 2006; Siegel et al. 2007; Sato et al. 2009). In gnathostomes, the predilection towards humans and 'model' organisms has resulted in conducting the majority of long-range comparative genomics between human, mouse, chick (Sarcopterygians), zebrafish, pufferfishes (Actinopterygians) and sometimes sharks (Chondricthyians). While studies of this nature are useful at finding patterns of conservation/loss, it is also important to understand how these patterns arise. To do this a wider taxonomic sample of each order is necessary. In Sarcopterygians the inclusion of reptiles and frogs and possibly even lungfishes and coelacanths will be useful. In Actinopterygians however the knowledge of more basal taxa is limited and therefore we will give a short review.

Actinopterygii (the ray-finned fishes) is subdivided into 4 major clades **Polypteriformes** (bichir and reedfish), **Chondrostei** (sturgeons and paddlefish), **Holostei** (gar and bowfin) and the crown group **Teleostei** (all remaining ray-finned fishes) (Nelson 1969). Compared to cartilaginous and lobe-finned fishes, the ray-finned fishes have undergone a remarkable number of whole genome duplications, which makes genome comparisons slightly more complicated. The teleosts are by far the most successful extant ray-finned fish clade with 38 orders, 426 families, 4,064 genera and >27,000 species (Nelson 2006). All teleosts share a whole genome duplication dubbed the FSGD (Jaillon et al. 2004). One hallmark of the teleost clade is additional rounds of whole

genome duplications in a lineage-specific manner, e.g. salmonids (Ohno 1970; Allendorf and Thorgaard 1984) which have tetraploid genomes and as many as sixteen *Hox* clusters (Moghadam et al. 2005a; Moghadam et al. 2005b). Interestingly, evidence suggests the evolution of basal ray-finned fishes like gar (Ohno 1969) and paddlefish (Dingerkus and Howell 1976) may also be associated with lineage-specific duplication events.

As shown in Figure 2, **Amiiformes**, represented by a single extant species called the bowfin (*Amia calva*) is the closest sister group to the crown group of teleosts (Nelson 1994; Bemis et al. 1997). The bowfin lives in eastern North America and there is no indication of polyploidy in this species.

The next closest order is **Semionotiformes** (Nelson 1994; Bemis et al. 1997) (Figure 2). **Semionotiformes** are comprised of two extant genera with seven species known commonly as gars. Gars can be found in eastern North America and Central America. It has been postulated that gars may have undergone their own genome duplication (Ohno et al. 1969; Dingerkus and Howell 1976).

Order:**Acipenseriformes** (Figure 2) contains two families (**Acipenseridae**, **Polypdontidae**; Nelson 1994) comprised of six genera and twenty-six species known commonly as sturgeons and paddlefish. The twenty-four living species of sturgeons are found throughout the Northern Hemisphere; the two living species of paddlefish are found in the United States and China. Sturgeons have undergone at least three whole-genome duplications that result in phylogeny that can be deduced via chromosome number (Fontana and Colombo 1974; Ludwig et al. 2001; Fontana et al. 2008). Paddlefish appear to have undergone at least 2 whole-genome duplications (Dingerkus and Howell 1976).

The most basal ray-finned fish clade is that of the order **Polypteriformes**. Commonly known as bichirs and reedfish, **Polypteriformes** contains two genera and ten species. Evidence suggests that bichir has not undergone its own genome duplication (Chiu et al. 2004; Mulley et al. 2006; this study). Strikingly, the stem ray-finned fishes comprise ~ forty-four species or less than 1% of all ray-finned fishes. Despite a handful of useful candidate species among the basal ray-finned fishes (such as bichir and bowfin), most authors focus only on comparisons between outgroups to the Actinopterygians (Sarcopterygians; Chondrichthyians) and teleosts. We will present evidence that suggest many patterns and possible hypotheses are missed due to this simplistic approach.

2.1.2 Teleost Bias

The majority of gen(om)e comparison studies that have been conducted comparing Sarcopterygians and/or Chondrichthyians to Actinopterygians have focused on members of the teleost clade (Figure 2). Siegel et al. 2007 compared the Para*Hox* clusters of several teleost fishes to those of human and mouse. In this study they find that teleosts do share some conserved gene order and conserved non-coding sequences with mammals. They also found several non-coding DNAs that are conserved only among teleost fishes. But without more basal **Actinopterygii** in the analysis, there is no way of knowing if these CNS are novel to teleosts or are the result of *cis*-regulatory sequences that evolved in ray-finned fish stem lineages. The lack of phylogenetic depth in this study skews their conclusions about how the Para*Hox* clusters have evolved in the teleosts. This same type of problem arises in phylogenetic comparative work done by other authors as well (Lee et al. 2006).

Sato et al. 2009 conducted loss/retention analysis of 130 genes of taste and olfactory transduction and tricarboxylic acid cycle pathways. Again they compared several teleosts genomes to that of human and determined that one-half of the genes were present prior to the FSGD and 40% of those were maintained as duplicates. Many more genes may have been present prior to the FSGD and subsequently lost after the FSGD although this is unknown given that there is ~150 my of evolution missing in the comparison. One may conclude that this missing information is not relevant given that the authors want to compare teleosts to any state prior to their genome duplication. However, without information of how ray-finned fishes evolved prior to the FSGD one has no way of knowing how or why genes are retained or lost after the FSGD and how this has contributed to the evolution of the teleosts. Although the data can stand on its own as far as lost/retention rates, the conclusions drawn from this data are lacking a certain understanding of immediately prior states. This point is argued in Johnston et al. (2007) where the authors find that genes are more likely to be maintained as duplicates if they were undergoing accelerated evolution prior to the duplication.

Another assumption that is made using this type of data is that teleosts have lost several DNA regions including genes and CNSs that are shared between lobe-finned and cartilaginous fishes (Chiu et al. 2002; Venkatesh et al. 2007; Yu et al. 2008). This loss is often contributed to an accelerated mutation rate and/or the FSGD (which may or may not be mutually exclusive events). However, when stem-ray finned fish data is included in these types of studies one finds that this turnover started prior to the teleost radiation (Chiu et al. 2004; this study). Ravi and Venkatesh (2008) addresses this issue as well but still fails to use any stem-ray finned fishes in their study.

The lack of data from pre-FSGD stem ray-finned fishes is glaring in phylogenetic studies. For example, a recent study attempted to use EST data to help solve gnathostome phylogenomics for the major clades (**Chondrichthyes**, **Sacropterygii**, **Actinopterygii** and **Tetrapoda**; Hallström and Janke 2008). They contributed lungfish (*Protopterus aethiopicus*) DNA sequence to the study in order to expand the depth of Sarcopterygian sampling, but despite having 4 teleosts, included no stem ray-finned fishes. It is impossible to confidently arrange these major clades without data from pre-FSGD ray-finned fishes. Similar work was done by Martin (2001).

If pre-FSGD stem ray-finned fishes are needed for genome comparison, then which are ideal candidates? Chiu et al. (2004) used bichir, the most basal, extant ray-finned fish. We propose that this is an ideal pre-FSGD ray-finned fish given its place in phylogeny as well as its apparent lack of a lineage specific genome duplication. Comparisons using bichir will add data from a genome that should be similar to Chondricthians and Sarcopterygians structurally, but undergone selective pressures of an Actinopterygian.

2.1.3 Bichir Phylogeny and Anatomy

At the base of the Actinopterygian tree is the genus *Polypterus* (bichir). Despite this phylogenetic position, bichir is not frequently used in studies. Perhaps this is because only recently has bichir held that phylogenetic position.

Polypterus is an enigmatic taxon with many shared and unique features that have given taxonomists constant headaches (Figure 4). Among the shared features are: 1. Lobe-like fins (**Sarcopterygii**) 2. Ganoid scales (**Sarcopterygii**) 3. 2 gular plates (**Coelacanthimorpha:Sarcopterygii** and **Holostei**) 4. Spiracles (**Chondrostei**) 5.

Feathery external gills and double ventral lungs as juveniles (**Dipnoi:Sarcopterygi**) and 6. Heterocercal tail (**Chondrostei**) (Janvier 1996).

Among the unique features are: 1. Dorsal fin rays that project horizontally from a vertical spine (Helfman et al. 1997) 2. Recoil aspiration (Brainerd et al. 1989) 3. Heterocercal tail, external gills and lobe-like pectoral fins are due to homoplasy and are anatomically different (Janvier 1996). Early systematists had difficulty classifying bichir using anatomical features and they were placed in at least four separate clades including the **Dipnoi** (lungfishes), **Ganoidei**, three separate groups of **Actinopterygii** (**Teleostei**, **Holostei**, and **Cladistia**) and an entirely separate subclass known as **Brachiopterygii** which was held in equal weight with the **Actinopterygii** and **Sarcopterygii** (Helfman et al. 1997). A consensus was reached in 1982 when Patterson placed **Polypteriformes** as the basal group to the **Actinopterygii**. After the advent of molecular phylogeny much work has been done and today it is accepted that bichir is indeed the most basal ray-finned fish (Inoue et al. 2003; Kikugawa et al. 2004; Figure 2).

Due to the diligent work put into phylogenetically positioning bichir among other fishes we can now use bichir in studies requiring a basal ray-finned fish. But without knowing the ploidy level of bichir we may still incur the issues that arise with other stem ray-finned fishes like the sturgeons, paddlefish and gars. For this reason our lab has put in great effort to help determine whether or not bichir has undergone a lineage specific genome duplication or at the very least chromosomal duplication resulting in greater than 4 *Hox* gene clusters.

2.2 Bichir *Hox* Gene Discovery

2.2.1 PCR Discovery and Sequencing of *Hox* Genes

Materials and Methods

Degenerate primers were designed from alignments of vertebrate *Hox* genes for *Hox*C4 and *Hox*D9 (C4 Forward 5'-ATG ACG TCG TAT TTG ATG-3'; C4 Reverse 5'-TGA TTT GCC TCT CGG AG-3'; D9 Forward 5'-GTA TTG GTA AAT ATG ATC ACG-3'; D9 Reverse 5'-CKG TTC TGA AAC CAG ATT TT-3'). PCR reactions were setup using reagents from Applied Biosystems™ AmpliTaq® DNA Polymerase kit (N8080153) in the following concentrations: 1X buffer, 3.5 mM $MgCl_2$, 0.8 mM dNTP, and 0.5 uM primers. 50 ul reactions were made with 3 ul genomic DNA (unknown concentration), 2.5 units Taq DNA polymerase and diluted with sterile water. Negative controls were made under identical conditions except genomic DNA was not added. The PCR was run using a MJ Research PTC-200 DNA Engine thermalcycler under the following conditions: 95 C for 5 minutes, 30 cycles of (95 C for 1 minute, 47 C (*Hox*C4) or 52 C (*Hox*D9) for 1 minute, 72 C for 2 minutes), 72 C for 10 minutes. 5 ul of each PCR reaction was mixed with 1 ul 6X loading dye and loaded onto a 2% agarose gel (1X TAE) along with 4 ul of 1X Invitrogen 1kb DNA ladder (15615-024). Gels were run on an Owl brand electrophoresis unit at 60 milliwatts for 1.5 hours. Bands of correct size were excised from the gel and purified using Qiagen Gel Extraction Kit (28706) under standard protocol except DNA was eluted using 10 ul sterile water. Gel purified DNA was cloned into Promega pGEM®-T Vector System II (A3610) using standard protocol except reaction was incubated at 16 C for 20 hours. 2 ul of ligation reaction was transformed into 25 ul *Escherichia coli* Promega JM109 Competent Cells (L2001) under

the following protocol: on ice for 20 minutes, 45 C (water bath) for 42 seconds, on ice for 2 minutes, add 240 ul LB broth, 37 C for 1 hour at 200 r.p.m. The transformed cells were plated on to LB agar plates containing 50mg/mL ampicillin, 1M IPTG and X-gal and grown at 37 C for 16 hours. Colony PCR was performed on white colonies under identical conditions as original PCR except in 20 ul reactions. Colonies with band of correct size were grown in 3 mL of LB broth at 37 C for 16 hours at 200 r.p.m. Plasmids containing insert of correct size were purified using Qiagen QIAprep Miniprep Kit (27106) using standard protocol except DNA was eluted using 30 ul sterile water. 300 ng of plasmid DNA was sent for sequencing using the T7 primer to the DNA core facility at the University of Medicine and Dentistry of New Jersey, Pisctaway, NJ.

Chromatograms of the DNA sequences for *Hox*C4 and *Hox*D9 were analyzed for errors. Resulting DNA sequences were then trimmed of vector and primer and blasted against the non-redundant database at http://blast.ncbi.nlm.nih.gov/. Verification of *Hox* paralog identity was based on identity of top matches from this blast. After verification of paralog identity, sequences were analyzed via phylogenetic inference. Alignments were created of all known paralogs using ClustalW algorithm provided by MacVector® 9.5.2. The alignment was corrected by eye so that all sequences started at the same position. Neighbor-joining, Maximum Parsimony, Maximum Likelihood and Bayesian trees were created from these alignments. Neighbor-Joining and Maximum Parsimony trees were created using default parameters in PAUP* v4.0b10 (Swofford 2003). Maximum Likelihood trees were created using GARLI v0.951 (Zwickl 2006), which can be downloaded from www.bio.utexas.edu/faculty/antisense/garli/Garli.html. The starting tree was obtained using heuristic search under the likelihood optimality criterion in

PAUP* v4.0b10 (Swofford 2003) with settings as specified by Modeltest (Posada and Crandall 1998, see Model Selection). The same settings were inputted into GARLI v0.951 (Zwickl 2006). Node confidence was assessed by using the bootstrap resampling method with 2,000 replications. The "number of generations without improving topology" setting was changed from 10,000 to 5,000 as suggested in the GARLI manual when doing bootstraps, as it will "shorten the run time without significantly affecting the results". A consensus tree was created using majority rule with a cutoff of 50% in PAUP* v4.0b10 (Swofford 2003). Bayesian trees for the nucleotide alignments were constructed using MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003) and the parallel version of MrBayes v3.1.2 (Altekar et al. 2004) under the settings specified by MrModeltest. Two independent Markov Chain Monte Carlo analyses were run with the following settings: number of generations was set to 1,000,000, sample frequency was taken every 1,000 steps, number of chains was set to 4 and the temperature was set 0.2. 'Burnin' was assessed after the run using the sum parameters command. The 'burnin' for the nucleotide analysis was set to 1 for all 3 datasets, which is equal to the first 1,000 steps or tree topologies. A majority rule consensus tree was created disregarding the 'burnin' trees using the sum trees command with a cut off of 0.50 posterior probability.

2.2.2 BAC Library Screening

Materials and Methods

A 5-fold coverage bichir genomic bacterial artificial chromosome (BAC) library (Chiu et al. 2004) was screened for presence of bichir *Hox* genes using a DNA probe of *Hox*C4 and *Hox*D9. High-density 5 X 5 arrayed filters were made by the RZPD (http://www.rzpd.de). Southern hybridization was performed on these filters. DNA

probes were labeled with a non-radioactive marker, digoxigenin. We DIG (Digoxigenin)-labeled the *Hox*C4 and *Hox*D9 fragment using Roche Diagnostics DIG DNA labeling kit (11175033910). The DIG labeling reaction includes boiling 800 ng (15 ul) of the *Hox*C4 or *Hox*D9 fragment in water for 10', placing on ice (H_2O) for 10', briefly spinning down, then adding 2 ul of the hexanucleotides, 2ul dNTP labeling mix, and 1ul Klenow enzyme provided in Roche kit. We then placed the labeling reaction at 37 C for 20 hrs. Southern hybridization was performed on the 5 X 5 filters in a Lab-Line rotisserie incubator at 65 C. Briefly, filters were "pre-hybridized" at 65 C for 5 hrs. DIG-labeled probe was added to pre-hybridization solution at a concentration of 1ul/ml. Pre-hybridization solution was then replaced with hybridization solution and incubated for 20 hours at 65 C. Filters were then washed with the following steps all at room temperature: rinsed in 2X SSC (300 mM sodium chloride, 30 mM sodium citrate), washed in first hybridization wash (2X SSC, 0.1% *v/v* SDS-sodium dodecyl sulphate) for 20 minutes while rocking, washed in second hybridization wash (1X SSC, 0.1% SDS) for 20 minutes while rocking, rinsed in diluent buffer (300 mM NaCl, 100 mM Tris ph 8.0) for 20 minutes while rocking. Filters were then saturated by 1% blocking solution (Roche 11096176001) for 1 hour while rocking. Alkaline phosphatase conjugated antibody to DIG (Roche 11093274910) was applied to filters in a 1/1000 ratio in 1% blocking buffer for 30 minutes while rocking. Filters were once again washed in a detergent solution (0.3% *v/v* Tween in diluent buffer) 3 successive times for 10 minutes while rocking. Filters were then incubated in detection buffer (1mM Tris pH 9.5, 0.1M NaCl) for 5 minutes. Filters were then affixed to old film panel and CDP star substrate (Roche 12041677001) was applied and covered with thin sheet of poly vinyl chloride. The substrate reacts with the alkaline

phosphatase to emit light, which is detected by film in a dark room. Coordinates of positive clones are determined in duplicate on a 2 clone per array basis.

A group of 4 clones were discovered with 1 or both *Hox*C4 and *Hox*D9. Due to the conflicting data from these hybridization experiments the bichir BAC library was screened via PCR for presence of *Hox*C4 containing clones using bichir specific primers (PseC4forward 5'-ATG AGG TCG TAT TTG ATG-3' and PseC4reverse 5'-TGA TTT GCC TCT CGG AG-3').

The bichir BAC library consists of 216 pools each of which contains 384 clones. The PCR reactions were setup similar to that of the genomic DNA isolation except they were scaled to 16 ul. The same PCR program was performed as for the genomic DNA isolation. No pool contained a single distinct band corresponding to the expected product therefore this analysis was not extended.

Earlier, bichir BAC clone CLN84 (AC135508) was isolated that contained the *Hox*D1 gene. This BAC clone was sent to Eric Lander at the Whitehead Institute/MIT Center for Genome Research to be sequenced. Annotation of the clone was performed to yield 3 *Hox* genes; *Hox*D1, *Hox*D2 and *Hox*D3. *Hox*D2 is the first D2 paralog that has been found among bony-fishes and supports the idea that bichir may represent a basal ray-finned fish. Phylogenetic analysis clearly defines *Hox*D1, *Hox*D2 and *Hox*D3 as 'D' paralogs of their specific *Hox* groups (Figure 7). There is no evidence of a *Hox*D4 gene upstream of *Hox*D3 despite 20 kb of sequence. The *Hox*D3 gene has the largest known intron of any vertebrate *Hox* gene. The *Hox*D3 intron is 16,255 bp and contains one intact Tc1 DNA transposon as well as remnants of Rex like non-LTR retrotransposons (see below). Bichir BAC clone M919 (AC138742) was annotated and found to contain 2

of the 3 exons of EVX2 (Figure 7) a vertebrate ortholog of the *Hox* domain containing gene in *D. melanogaster*, even-skipped. A *Hox*B containing bichir BAC clone (AC138147) was also found. This BAC clone contains the entire open reading frame (ORF) of *Hox*B5, *Hox*B8, *Hox*B10, *Hox*B13 and partial sequence for *Hox*B4 and *Hox*B7 (Figure 7). The partial sequences are most likely due to incomplete sequencing of the BAC clone and not due to a pseudogene status for *Hox*B4 and *Hox*B7.

2.3 *Hox* Gene Data Suggest Bichir is a Basal Ray-finned Fish

Phylogenetic results indicate that *Hox*C4 and *Hox*D9 of bichir were assigned to the correct clusters and paralog groups and each supports bichir's assignment as the most basal living ray-finned fish (Figure 5). Previous work has suggested that bichir does not have its own specific genome duplication (Chiu et al. 2004; Mulley et al. 2006), which is seen in other basal ray-finned fishes such as paddlefish, sturgeon, and gar (Ohno et al. 1969; Fontana and Colombo 1974; Dingerkus and Howell 1976; Ludwig et al. 2001; Fontana et al. 2008). *Hox*C4 is the first discovery of a *Hox*C paralog gene in bichir. *Hox*D9 adds to the annotation of *Hox*D1, *Hox*D2, *Hox*D3, *Hox*D12 and EVX2 in bichir that were found via clone identification and sequencing.

The discovery of several linked *Hox*B and D cluster genes as well as a *Hox*C cluster gene has helped verify that bichir has not undergone its own genome duplication (in agreement with Chiu et al. 2004 based on the *Hox*A cluster) and has also lent information to the evolution of *Hox* gene clusters in vertebrates as bichir has the first instance of a group 2 gene on the *Hox*D gene cluster in **Osteichthyes** and also the first instance of a group 1 gene on the *Hox*D gene cluster in **Actinopterygii** (Figure 6).

2.4 Conclusion

Bichir is in an important phylogenetic position among ray-finned fishes. Diverging from the rest of the ray-finned fishes prior to the FSGD is just one important factor. Bichir also has the distinction of being the most basal living ray-finned fish and according to the data gathered during this and previous work (Chiu et al. 2005; Mulley et al. 2006) does not appear to have undergone a separate lineage-specific genome duplication. Because of these factors it is important and advantageous to include bichir in all analyses determining the effects, or lack there of, of the FSGD on the evolution of ray-finned fishes. As will be presented in the following chapter, data from bichir *Hox* genes and *Hox* gene clusters has lent valuable information to the evolution of *Hox* gene clusters prior to the FSGD that supports the notion the duplication alone is not sufficient in explaining the genomic and phenotypic plasticity seen among the teleost lineage of ray-finned fishes.

Chapter 3- Re-analyzing Teleost Uniqueness

3.1 Introduction

3.1.1 *Cis*-regulatory Elements

In this chapter we present an argument of why *cis*-regulatory elements are important to evolution, how *cis*-regulatory elements have evolved in the conserved *Hox* gene clusters and how the addition of bichir affects the analyses of the changes observed in the non-coding DNA of teleosts *Hox* gene clusters.

We will start with a background of *cis*-regulatory elements, their detection and their importance. The majority if not all eukaryotic genes contain expression level controlling DNA sequences called *cis*-regulatory elements (CRE). These elements are normally found flanking the transcribed portion of the gene but may also be found within introns and even exons (including the coding region). CREs are supplementary to the promoter sequence that is responsible for binding RNA polymerase and its associated proteins.

The first CRE described was the *lac* operon (Jacob and Monod 1961). This operon is responsible for controlling the response to the presence or absence of glucose/galactose in the environment. From this knowledge came the idea that CREs are switches that could turn on or off the production of a protein. We now know that regulation can occur at many levels from DNA to protein. We also know that CREs are not simply ‘on/off’ switches. CREs may also work more like a dimmer switch controlling not only if RNA will be made but also how likely and how often RNA will be made. This is possible because CREs are made up of a bundle of transcription factor binding sites (TFBS) that bind proteins responsible for either enhancing or suppressing

transcription. CREs respond specifically to the transcription factor environment allowing them to regulate their genes in a context dependent manner, in many instances at specific time-points during ontogeny and in specific tissues.

Complimentary to their role in development and tissue maintenance, CREs are likely to be targets of natural selection at the population level (see 3.1.4) and hence play an important role in genomic (Vavouri and Lehner 2009) and phenotypic evolution (Gompel et al. 2005). Despite their importance, little is known about CREs. This is due to the difficulty in identifying CREs in the genome because they do not code for a specific gene product and may be located far from the gene(s) whose action they control. Multiple methods have been developed to identify CREs. First we will look at traditional wet lab approaches and then computer-based approaches derived from them.

3.1.2 Traditional Approaches to Locating *Cis*-regulatory Elements

Traditional approaches to finding CREs include DNase footprinting, electromobility shift assays (EMSA) and deletion analysis, which can all be followed by transgenics. The first approach, DNase footprinting, consists of binding specific transcription factors (TFs) to a known sequence and digesting away the exposed, single-stranded DNA leaving behind only the DNA protected by the TF; termed the TFBS (transcription factor binding site), Using this approach, Galas and Schmitz (1978) first reported that TFBS are often clustered in genomes, which is one important trait of CREs.

The second approach, EMSA, involves incubating a specific DNA sequence with either proteins hypothesized to bind to the sequence or whole nuclear protein extracts and identifying DNA-protein complexes by their retarded rate of migration in gels, or ‘shift’. While the EMSA procedure alone cannot identify the specific protein that is bound to the

known DNA sequence, expansion of this method (supershift; Kristie and Roizman 1986) has proven powerful in identifying CREs of known TFs.

Using these traditional wet lab approaches many TFBS have been identified. Several databases such as TRANSFAC (Matys et al. 2003) and JASPAR (Sandelin et al. 2004) store all these known TFBS and allow one to submit an unknown DNA sequence thought to contain TFBS and have a search algorithm identify any similarities. The search algorithm works by creating a degenerate binding matrix for a transcription factor with known binding sites and then aligning that to the sequence in question. The degenerate binding matrix is designed by giving scores to the presence of one of the four nucleotides (A, C, G, T) at each position relative to the frequency they appear in known binding sites. Sometimes certain nucleotide positions are weighted heavier overall (core sequence) due to their inclusion in most if not all TFBS for their particular TF. These programs can do this for every transcription factor in their database so as to provide one with as complete of an array of possible TFBS for their sequence as possible.

The third approach, deletion analysis, has been used to detect *cis*-regulatory elements for many genes including *Hox* genes (Shashikant et al. 1995; Nonchev et al. 1996; Maconochie et al. 1999). Deletion analysis involves removing portions of the DNA upstream of a gene and looking for indication of loss of expression. After expression is lost the region deleted can be narrowed for higher resolution. All of these methods are, of course, indirect evidence of *cis*-regulatory activity and can be supplemented by transgenics.

Transgenics uses reporter gene assays, which involve placing a putative *cis*-regulatory upstream of a minimal promoter and a marker gene (coding for a visible

protein; ie. green fluorescent protein or a protein that can be secondarily visualized; ie. β-galactosidase). This construct is then transiently or permanently transfected into a model organism. Via transfection one can test for recapitulation of the endogenous expression pattern of the gene thought to be regulated by the putative *cis*-regulatory element in question and/or identify other possible regulated genes.

3.1.3 Comparative Genomic Approaches to Locating *Cis*-regulatory Elements

In the post-genomic era, comparative genomics has become a powerful approach to identifying CREs. In this method, orthologous non-coding DNA sequences from two to several groups of organisms, which collectively add up to enough evolutionary time that similarities in their sequences can be considered not likely by chance, are aligned. This approach is based on the observation that CREs evolve at a slower rate than surrounding non-coding DNA, presumably because they have a function(s) and are subject to purifying selection (Keightley et al. 2005). Evolutionarily conserved noncoding sequences (CNS; Hardison 2000) are putative CREs. In this approach, the depth of evolutionary time necessary for this analysis differs based on parameters including relative mutation rate, genome size, and generation time (review, Martin and Palumbi 1993). Protocols implementing this method include rVista (Loots and Ovcharenko 2004), TRACKER (Prohaska et al. 2004), and MONKEY (Moses et al. 2004).

Phylogenetic Footprint Clusters (PFCs) are a stringently defined type of CNS with specific methods of detection. PFCs are defined as CNSs that between any two species contain 2 or more “footprints” within 100 nucleotides of each other (Chiu et al.

2002). A footprint is defined as a string of at least 6 nucleotides that are conserved 100% between two or more species being compared that, for mammalian lineages, represent at least 250 million years of additive evolutionary time (Tagle et al. 1988). This method was modeled after DNAse footprinting (see 3.1.2; Galas and Schmitz 1978). Evolutionary biologists determine putative TFBS by allowing evolution to do the "digestion" that can be seen when comparing homologous regions of two distantly related organisms whereby there has been enough evolutionary time that the surrounding non-binding sites have diverged and only the TFBS are still conserved (Tagle et al. 1988).

Phylogenetic footprinting has shown to be a useful method in the search for CREs. The majority of known CREs in *Hox* clusters can be found by this method, although, the amount of false positives has not been assessed due to the amount of time necessary to empirically test each PFC that does not already correspond to a known CRE. A system such as the *Hox* clusters lends itself well to this type of CRE inquiries for several reasons. 1. Identifying orthologous non-coding DNA is highly accurate due to the highly conserved nature of the cluster formation. 2. Because of the importance *Hox* genes have to development, their expression patterns have been well researched and documented. 3. Despite this seeming understanding of the system very few *Hox* genes have had their CREs identified in much detail.

3.1.4 *Cis*-regulatory Elements vs. Transcription Factors

Cis-regulatory element mutations are thought to be responsible for the majority of novelty needed for adaptation in comparison to coding sequence mutations. This hypothesis was first put forth based on the structure and function of CRE vs. coding sequence (Monod and Jacob 1961; Britten and Davidson 1971). *Cis*-regulatory elements

are modular in nature and hence may easily overcome the problem of pleiotropy that acts against changes in protein coding sequences. *Cis*-regulatory elements can drive the expression of a particular gene in many different stages of development as well as in many different tissue types. Depending on which TFs are being expressed in any particular cell or at a given time (developmental, seasonal, etc.), the CRE can respond in kind to either suppress or enhance transcription of the gene it regulates. Because of this cell environmental plasticity, it is possible to change the expression of the gene in one cellular context without affecting the expression in another cellular context (reduced pleiotropy). Coding sequences on the other hand incur mutations that 69% (148/216) of the time change the amino acid sequence that it codes for and thus the protein that it produces.

It is thought that changes in the amino acid sequence would be highly pleiotropic in that a change to the protein structure would affect the function of the protein in all contexts (of course this can be overcome somewhat by alternative splicing and domain specific contexts). Based on these ideas it is proposed that CRE evolution can play a large role in adaptation. There are many experiments to show that this is indeed the case; well-known examples include pigmentation in the fruit fly (Wittkopp et al. 2002; Gompel et al. 2005) and pelvic reduction in sticklebacks (Shapiro et al. 2004). At the population level, CREs have shown to play a role in the adaptation of malaria resistance (Tournamille et al. 1995), lactase persistence (Bersaglieri et al. 2004), and blood clotting in humans (Rockman et al. 2005).

This is an important distinction because all evolution must take place at the population level where changes in allele frequencies cause changes in the species

phenotype as a whole. That CREs would be a more likely source of functional variation/adaptation at the population level is not surprising. As summarized by Wray (2007), expression differences in alleles are independent which means that CREs controlling the expression are co-dominant, whereas coding sequence mutations are more often than not recessive. This is important because for a beneficial allele to reach a frequency in a population that is measurable and effectual it must be selected for in the heterozygote. This is the case because before a sufficient number of homozygotes appear in the population the allele will likely be lost to drift.

There are many good arguments for why CREs play a larger role in adaptation and it appears at first glance that the empirical evidence agrees. But, it is also important to assess the assumption that is made to reach this hypothesis; changes to the amino acid sequence will have pleiotropic effects. This is not necessarily the case. Alternative splicing is one such form that may overcome pleiotropy. Also, proteins have many domains especially transcription factors. Transcription factors in particular have a DNA binding domain that should stay well conserved, changes in this region will almost surely be pleiotropic. But, transcription factors also have protein-protein interaction domains that help them bind to other proteins that are apart of the enhancer/suppressor complex that regulates transcription for the gene they are helping to control. Changes in these regions may or may not be completely pleiotropic such that a transcription factor will not always be binding with the same protein partners in different context.

Lending evidence to this type of modularity are motifs known as Short Linear Motifs (SLiMs) that help regulate these protein-protein interactions (Neduva and Russel 2005). SLiMs are normally 6-10 amino acids in length and often can be modular. Re-

arrangements of the amino acids can occur giving them a different binding profile and only a few changes can completely inactive or cause a *de novo* binding. In this way, SLiMs can overcome the pleiotropic effects that plague proteins coding mutations and be selected for very much the same way as a CRE.

Despite the possibility of a protein to overcome pleiotropy there is very little evidence that this occurs. There may be other factors that are not considered that keep proteins from overcoming pleiotropy. One reason may have to do with the difficulty to select for these methods at the population level. Which as is pointed out earlier in this paper is necessary to become a contributing allotype. SLiMs may be able to overcome this evolutionary snag but to date very little is known about them and there is no such evidence to support that they do indeed overcome pleiotropy. So until SLiMs or another method is better classified it is still the consensus that the majority of selection on mutations that affect development should occur in CREs.

3.1.5 *Cis*-regulatory Elements and *Hox* Gene Cluster Evolution

CREs are important to the evolution of *Hox* gene clusters. CREs are the most likely targets of natural selection, especially for higher order regulatory genes such as *Hox* genes. This is apparent in how the *Hox* genes accomplish their role. The function of *Hox* genes is to bind to their downstream targets (CREs) and either enhance or suppress the expression of the genes these CREs control. The main result of this transcriptional regulation is locating in space and time the formation of particular structures along an axis. *Hox* genes accomplish this spatiotemporal specificity via overlapping expression profiles setting up specific domains during development of the embryo. For this reason, changing the *Hox* gene's own spatiotemporal expression is the most effective way to alter

their function. This is accomplished by altering the CREs of the *Hox* genes themselves thus changing their ability to respond to particular sets of upstream TFs. The result of this alteration will be the *Hox* gene and all of its downstream effectors being expressed in a different location in the embryo and thus altering the positioning along the axis.

The importance of the CREs of *Hox* genes can also be seen via work in the McGinnis lab (McGinnis et al. 1990; Malicki et al 1993) which has shown swapping out the coding region of *Hox* genes between *Drosophila* and mammals still allows them to function normally because they are being expressed at the right time and location. The importance of CREs also comes up in the discussion of the maintenance of the cluster formation. Several evolutionary biologists (Peifer et al. 1987; Kmita and Duboule 2003; Santini et al. 2003) have postulated that one reason the *Hox* genes maintain a clustered formation is because they locally share CREs (Gould et al. 1997). The question of whether or not or how much of a role local CRE sharing plays in maintaining the cluster formation is still open to debate and it is important to understand the maintenance pattern of CREs to determine the likelihood of this hypothesis.

3.2 *Hox* cluster Conserved Non-coding Sequence Database

3.2.1 Need for Comprehensive Database

Previous attempts have been made to catalog the CNSs within the *Hox* gene clusters of gnathostomes (Chiu et al. 2002; Santini et al. 2003; Chiu et al. 2004). These attempts were made in an effort to understand the relationship between species that contain 4 *Hox* gene clusters and those that contain more. In these studies a representative species is chosen from 2 or 3 clades of gnathostomes and then their *Hox* gene clusters are aligned and scanned for CNSs. The issue with these comparisons, which we hope to resolve, was the exclusion of species due to lack of sequencing data or the desire for simplicity.

As described in Chiu et al. (2004) it is important to include basal ray-finned fishes. Also apparent in this study is the knowledge that, at least for the teleost fishes, species thought to be similar may have a quite distinct retention patterns with the same disparately related species. For this reason we have attempted to create a comprehensive database of CNSs for all available gnathostome *Hox* gene clusters in an attempt to better understand the pattern of retention and loss of CNSs in these otherwise highly conserved gene clusters. Understanding this pattern will lead to a better understanding of the forces that act to resolve (or not) redundancies in conserved genetic systems that undergo replication. We also hope that establishing a comprehensive database with an informative and flexible naming system will lead to better communication between labs and help expedite searches for both the function of these CNSs as well as their contribution to the cluster formation of the genes they regulate.

3.2.2 Annotating *Hox* clusters

Materials and Methods

The *Hox* clusters used in this study were constructed by a variety of methods. *Homo sapiens* (human) *Hox*A cluster was taken from a shorter fragment of a contig made from clones with accession numbers AC004079, AC004080, AC010990. Human *Hox*B, *Hox*C, and *Hox*D clusters were taken from genomic DNA from the UCSC genome browser http://genome.ucsc.edu/ (NCBI35/hg17, release May 2004). The *Hox*B cluster is taken from chromosome 17, nucleotides: 43,961,813-44,161,040. The *Hox*C cluster is taken from chromosome 12, nucleotides: 52,618,296-52,735,253. The *Hox*D cluster is taken from chromosome 2, nucleotides: 176,772,385-176,881,142.

Xenopus tropicalis (frog) *Hox* clusters were taken from Ensembl web browser http://nov2005.archive.ensembl.org/Xenopus_tropicalis/index.html (JGI 3, release Jan 2005 v35.1b). The *Hox*A cluster is taken from scaffold 29 nucleotides: 1,777,789-2,133,531. The *Hox*B cluster is taken from scaffold 329 nucleotides: 415,000-1,016,000. The *Hox*C cluster is taken from scaffold 280 nucleotides: 199,492-581,365. The *Hox*D cluster is taken from scaffold 353 nucleotides: 474,676-800,000.

Latimeria menadoensis (coelacanth) *Hox*A cluster from EVX1 to *Hox*A9 was taken from BAC clone VMRC4-19C10 with genbank accession number 147788. The rest of the *Hox*A cluster and the entire *Hox*B, C, and D clusters were isolated from BAC clones and sequenced.

Heterodontus francisci (horn shark) *Hox*A cluster was constructed from two clones with accession numbers AF479755 and AF224262. The *Hox*D cluster was a clone with accession number AF224263.

Polypterus senegalus (bichir) *Hox*A cluster was made from two clones -22F22 and -164C2 with accession numbers AC126321 and AC132195, respectively. The partial *Hox*B cluster is constructed from clone L28995. *Hox*B13 is on segment L28995.23. *Hox*B10 is on segment L28995.7. *Hox*B8 and *Hox*B7 are on segment L28995.17. *Hox*B5 and *Hox*B4 are on segment L28995.1. The partial *Hox*D cluster is constructed from clone CLN84 which includes *Hox*D3, *Hox*D2, and *Hox*D1.

Danio rerio (zebrafish) *Hox* clusters were made from clones. *Hox*Aa cluster was constructed completely from clone 241I17 accession number AC107364, modifications included trimming the first 26,896 and the last 4,176 nucleotides and inserting a C at position 76,071. The *Hox*Ab cluster was made from clone 10O019 accession number AC107365 with an alteration of nucleotide 79,324 from T to C to avoid a premature stop codon. The *Hox*Ba cluster was constructed with two clones CH211-72A16 and BUSM1-254O17 with accession numbers BX927395 and AL645782, respectively. Clone CH211-72A16 was spliced to BUSM1-254O17 immediately after the *Hox*B4a gene. The *Hox*Bb cluster was simply clone BUSM1-227H09 accession number AL645798 used in its entirety with no alterations. The *Hox*Ca cluster was made from clones DKEY 148F24 and clone DKEY 81P22 with accession numbers BX465864 and BX005254, respectively. The *Hox*Cb cluster was taken from a scaffold from the Ensembl genome browser http://nov2005.archive.ensembl.org/Danio_rerio/index.html (WTSI Zv5, release July 2005 35.5b). The *Hox*Da cluster was clone RP71-78H1 accession number BX322661 used in its entirety.

Oreochromis niloticus (nile tilapia) *Hox*A cluster is the entire clone with accession number AF533976.

Morone saxatilis (striped bass) *Hox*A cluster is the entire clone with accession number AF089743.

Oryzias latipes (medaka) *Hox* clusters were taken from genbank in their entirety. *Hox*Aa, Ab, Ba, Bb, Ca, Da, and Db are under accession numbers AB232918-AB232924, respectively.

Tetraodon nigroviridis (spotted-green pufferfish) *Hox* clusters were all taken from Tetraodon Genome Browser at http://genome.ucsc.edu (Genoscope 7/tetNig1, release Feb 2004). The *Hox*Aa cluster is taken from chromosome 21, nucleotides: 2,878,001-3,153,406. The *Hox*Ab cluster is taken from chromosome 8, nucleotides: 6,506,471-6,727,504. The *Hox*Ba cluster is taken from chromosome Unknown, nucleotides: 37,928,410-38,293,032. The *Hox*Bb cluster is taken from chromosome 2, nucleotides: 1,321,876-1,537,033. The *Hox*C cluster is taken from chromosome 9, nucleotides: 4,083,941-4,353,227. The *Hox*Da cluster is taken from chromosome 2, nucleotides: 10,975,763-11,218,409. An alteration was made at nucleotide position 11,134,740; a T was deleted in order to shift back to correct frame. The *Hox*Db cluster is taken from chromosome 17, nucleotides: 9,471,355-9,694,740.

Takifugu rubripes (Japanese pufferfish) *Hox* clusters were acquired from the http://nov2005.archive.ensembl.org/Fugu_rubripes/index.html (FUGU 2.0, release May 2004 v35.2g). The *Hox*Aa cluster is constructed from the entire scaffold 47. The *Hox*Ab cluster is constructed from scaffold 330.

3.2.3 Create Conserved Non-coding Sequence Database

To obtain PFCs, pair-wise alignments were made between orthologous intergenic regions of human *Hox* clusters and all other species *Hox* clusters, bichir *Hox* clusters and

all ray-finned fish *Hox* clusters, and among all teleost *Hox* clusters using MacVector version 8.0 (Accelrys®) which uses the clustalW algorithm. For PFC analysis the pair-wise alignments between acanthomorph 'a' paralog *Hox* clusters were excluded due to their overall high identity. PFCs in this study were defined by the canonical definition of 2 or more sequences with 100% identity over at least 6 base pairs each of which are within 100 base pairs of each other. This analysis consisted of over 900 pair-wise alignments of homologous intergenic regions with an average alignment size of 10 kilobases. Due to the large number of PFCs obtained by this method I further restricted the criteria to include the alignment of each pair-wise PFC to be at least 60% identical and/or contain at least 5 PFs. All *Hox* clusters used in this study were scanned for repetitive and mobile DNA elements using the RepeatMasker database at http://www.repeatmasker.org/. PFCs that corresponded to masked elements were excluded.

In total I identified 563 unique PFCs that fit my criteria. After this I carefully determined which pair-wise PFCs overlapped across other species and numbered according to the relative position in the cluster with 1 starting at the posterior end of the cluster and ascending towards the anterior end of the cluster. This was accomplished by starting with the 'seed' species of human, bichir and zebrafish and determining overlapping PFCs by their overlapping nucleotide positions. After overlapping PFCs were determined, any PFC that existed in more than 2 species was subjected to a multiple alignment that was then trimmed to maintain the criteria of a PFC (Table S1).

PFCs were then broken down into sections based on differential retention of that section in the species that contain that PFC. A section may be as small as an individual

PF or as large as all but one PF. The PFCs were named based on their location in the cluster and for PFCs conserved among more than 2 species, their section content. The name consists of a 3 letter abbreviation for the species that it is found in, followed by 1 or 2 capital letters denoting the *Hox* cluster it resides in, followed by a 2 letter subscript denoting the intergenic region it resides in (see below), followed by the number of the PFC which are roughly ordered by their position along the cluster starting at the 5'end and sometimes followed by a 1-6 letter subscript describing the section content of that PFC (e.g. a PFC that is found between *Hox*A9 and *Hox*A8 in human is named $HsaA_{GH}74_{abe}$ which signifies that it is found in human between genes A9 and A8, it is the 74th PFC on the *Hox*A cluster and it shares sections a, b, and e with other species; Figure 8). For nomenclature purposes *Hox* genes were coded A-O starting with EVX and ending with group 1 genes. We propose that if new PFCs are found within a *Hox* cluster used in this analysis that they be named and numbered based on the following formula:

$$X = \frac{distAP}{distAB} + A$$

X = number of new PFC
P = new PFC
A = PFC 5' to P
B = PFC 3' to P

(e.g. if a new PFC is found 430bp downstream of $HsaA_{GI}103$ and $HsaA_{GI}104$ is 670bp downstream, then the new PFC should be named $HsaA_{GI}103.64$). I then compiled this information into two types of charts. One containing the name, location and sequence of PFCs in each individual *Hox* cluster and the other containing all PFCs in one *Hox* cluster paralog.

3.2.4 Analyze Conserved Non-coding Sequence Database

After the PFC database was created we searched for retention and loss patterns that may be pertinent to the evolution of gnathostomes with special emphasis on the ray-finned fishes and teleosts.

The first pattern looked at was 'deeply' conserved PFCs. 'Deeply' conserved PFCs are defined as found in at least 2 of the following groups, 1) horn shark, 2) $\geq$2 lobe-finned fish, 3) bichir, 4) $\geq$2 teleost fish. This criterion limits 'deeply' conserved PFCs to those that have maintained identity over at least 733 million years of additive evolutionary time. Of the 563 unique PFCs, 78 (13.9%) fit this criterion (Figure 9). Given a random distribution and an average PFC divergence time of at least 270 million years (low end estimate) one would expect 37% of the PFCs to fit the criteria of at least 733 million years so this is a significantly small amount (p-value=0.001). All PFCs were blasted against the expressed sequence tag (EST) database at http://www.ncbi.nlm.nih.gov/BLAST/. If human or zebrafish contained the PFC then their sequence was preferentially used to perform the blast search. All PFCs were also blasted against the nr database to identify untranslated regions and microRNAs. Additionally, PFCs considered 'deeply conserved' were blasted against the non-coding RNA database at http://research.imb.uq.edu.au/rnadb/default.aspx (Pang et al. 2005; Table S2).

Literature searches were done to identify all known *Hox* CREs and micro-RNAs. All known *Hox cis*-regulatory elements and micro-RNAs were detected as deeply conserved PFCs. This speaks to the robustness of this method at detecting conserved functional non-coding sequences. To help determine those PFCs that would be good

candidates for *cis*-regulatory elements efforts were made to identify PFCs that existed within sequences of known RNA transcripts. Of the 78 'deeply' conserved PFCs, 8 corresponded to sequences of known function: C8 early enhancer (Shashikant et al. 1995), A2 enhancer (Nonchev et al. 1996), D4 retinoic acid response element (RARE; Morrison et al. 1996), A4 RARE (Doersken et al. 1996), A5 MES enhancer (Larochelle et al. 1999), mir-196b (Yekta et al. 2004), mir-196-2 (Berezikov et al. 2005) and mir-10b (Berezikov et al. 2005). Of the remaining 70 'deeply' conserved PFCs, 22 were found in untranslated regions (UTR) of the RNA transcripts of *Hox* genes, 21 were found in the expressed sequence tag (EST) database and 27 have an unknown function. Of those 27, 6 lie within 500 bp up or downstream of the coding sequence of a *Hox* gene, which implies they could possibly be part of an unsequenced UTR or basal promoter element. That leaves 21 'deeply' conserved PFCs that could possibly act as *cis*-regulatory elements. It is also possible that those PFCs that lie within RNA transcripts and UTRs could also serve a *cis*-regulatory role.

Deeply conserved PFCs seem to have a biased towards the 3' end of the cluster. Between the paralogous *Hox* clusters the only difference that is apparent is the seemingly few deeply conserved PFCs on the *Hox*C cluster which is mainly due to the lack of PFCs with no known function. This may be due to sampling bias as the *Hox*C cluster is the only cluster that contains only 2 of the 4 groups considered (see figure 9 for explanation of groups). The finding of deeply conserved PFCs are not the only pattern worth exploring however, many unique patterns can be found when the phylogeny of the species used in this study are considered.

Several patterns of retention/loss of PFCs throughout gnathostome phylogeny can be seen in this database. First we will address the previous findings of Chiu et al. (2004), where it was determined that bichir showed a pattern of 'mosaicism' in its PFC content.

Previous studies found a trend of mosaicism in the retention of PFCs on the *Hox*A cluster of bichir when comparing it to the *Hox*A clusters of non-ray finned fish and teleosts, the more derived ray-finned fish clade. This mosaicism showed alternatives to the apparent pronounced loss of PFCs in teleost fishes, as seen by Chiu et al. 2002, when comparing them to the *Hox*A clusters of human and horn shark whose *Hox* clusters shared a remarkable number of PFCs. In the current study the *Hox*A cluster of bichir shared 17 PFCs exclusively with non-ray finned fishes (lobe-finned fishes and cartilaginous fishes) and 48 PFCs exclusively with teleost fishes which is 26% and 74% respectively.

This appears to be in contrast with the numbers shown in Chiu et al. 2004 because bichir shares much more with the teleosts but still supports the claim of bichir showing a mosaic pattern between non-ray finned fishes and teleosts. When doing this comparison however it is important to take into consideration the number of *Hox* clusters compared in each category. When scaled by the number of pair-wise comparisons made the numbers are a little closer with 47% of the PFCs found exclusively between bichir and non-ray finned fishes and 53% found between bichir and teleosts.

The pattern is different for the *Hox*B cluster where bichir shared 21% of its PFCs with the non-ray finned fishes and 79% of its PFCs with the teleosts even after scaling for number of pair-wise comparisons (Table 4). In order to see if this pattern may be by chance and also to assess the amount of conservation in the derived ray-finned fishes we

also calculated the number of PFCs shared exclusively between human and other non-ray finned fishes and human and teleosts. After scaling, the *Hox*A cluster of human shared 74% of its exclusive PFCs with other non-ray finned fishes and 26% with teleosts. These numbers were comparable on the *Hox*B cluster with 82% and 18% respectively, again showing the conservative nature of non-ray finned fishes *Hox* clusters.

Another category needed to determine the true mosaicism of bichir's *Hox* clusters is the number of PFCs that are shared only among the teleosts species. This becomes a problem however because the **Acanthomorpha**, which consists of all the teleosts in this study excluding zebrafish, are too closely related. This makes it difficult to assign PFCs because of the high identity between their non-coding regions. For this reason the category of "exclusive to teleosts" must be reduced to "shared exclusively between zebrafish and any other teleosts".

The *Hox*Aa cluster of zebrafish shared 49% of its PFCs with non-teleosts (lobe finned fishes, cartilaginous fishes and basal ray-finned fishes) and 51% with teleosts with or without scaling. The *Hox*Ab cluster was 71% non-teleosts and 29% teleosts; *Hox*Ba 29% non-teleosts and 71% teleosts; *Hox*Bb 56% non-teleosts and 44% teleosts all with scaling. For the *Hox*Aa cluster andBb cluster zebrafish appears completely mosaic. For the *Hox*Ab cluster zebrafish appears to be more non-teleost like and for the *Hox*Ba cluster zebrafish appears to be more teleost like. If the 'b' paralog clusters are ignored due to their low number of PFCs (and reduced gene number) then the zebrafish takes on a cluster specific pattern that mirrors that of bichir. The *Hox*A cluster of bichir and its ortholog the *Hox*Aa cluster of zebrafish both share around 50% with non-teleosts and 50% with teleosts. In contrast to the *Hox*A cluster but in agreement with one another, the

*Hox*B cluster of bichir and the *Hox*Ba cluster of zebrafish both share around 25% with non-teleosts and 75% with teleosts (Figure 10).

These patterns and others that will be explained in the following paragraphs show evidence that not only do bichir *Hox* clusters show 'mosaic' patterns but also these patterns are different for different clusters. When one combines this data of turnover in the ray-finned fishes (starting in bichir and continuing into teleosts) with the data that show the lack of retention of ray-finned fish specific PFCs it highlights the dynamic genomic environment that exists in the teleosts that started to take root in the basal ray finned fishes. This is in stark contrast to the extremely conserved nature of the *Hox* cluster regulatory environment seen in and among the sister lobe-finned fishes and more basal cartilaginous fishes.

Combining the information of deeply conserved PFCs with the idea of looking for patterns of gain/retention/loss in the *Hox* clusters we constructed a phylogenetic tree with apparent acquisition and loss numbers for deeply conserved PFCs at each node (Figure 11). The criteria of a deeply conserved PFC requires that the PFC has been conserved for ~1 billion years of additive evolutionary time making this type of analysis very important in identifying large (time) scale changes in non-coding sequence conservation within the *Hox* clusters of the gnathostomes. The identity of whether the PFC was gained or lost in each specific lineage was determined on a parsimony basis. Currently only the *Hox*A cluster can be analyzed using this method because it is the only cluster that we have complete sequences for in horn shark and bichir, two pivotal species in the analysis (Figure 11).

For the *Hox*A cluster it is evident that at least 30 PFCs were acquired sometime before the divergence of the 3 major gnathostome lineages and differential loss of these PFCs can be seen with the majority being absent in the teleosts' *Hox*Ab cluster. 4 PFCs were gained sometime after the divergence of cartilaginous fish but prior to the divergence of the lobe-finned and ray-finned lineages. Again the majority of these are absent in the teleosts' *Hox*Ab cluster. The lobe-finned fish lineage appears to have acquired 2 PFCs at the base which have been maintained throughout lobe-finned phylogeny. The lobe-finned fishes have also maintained all but one of the gnathostome acquired PFCs. The ray-finned fish lineage appears to have gained 7 new PFCs after they diverged from the lobe-finned fishes. They also lost two of the gnathostome PFCs. Bichir does not appear to have lost any of the gnathostome or bony-fish PFCs. The teleosts appears to have lost 7 more of the gnathostome PFCs and 2 of the bony-fish PFCs sometime prior to the genome duplication and did not maintain any PFCs that were gained after the divergence of bichir. The *Hox*Aa cluster is more conservative than the *Hox*Ab cluster in that it only lost 1 more of both the gnathostome PFCs and ray-fin PFCs after the genome duplication. The *Hox*Ab cluster appears to be very degenerate in its PFC conservation in that it lost 14 additional gnathostome PFCs, 1 additional bony-fish PFC, and 5 additional ray-fin fish PFCs after the genome duplication. Also the *Hox*Aa and *Hox*Ab clusters appear to have gained 1 PFC each after the genome duplication. It is equally parsimonious to assume that each of these PFCs was gained after bichir diverged but sometime prior to the genome duplication and then subsequently lost in the other paralogous cluster.

These results highlight the turnover of highly and anciently conserved PFCs (putative CREs) that began at the base of the ray-finned fish clade and continued through the derived ray-finned fish clade and the lack of turnover in the lobe-finned and cartilaginous clades.

3.3 Mutation Rate Estimation

Materials and Methods

Intergenic distances analysis was done on select intergenic sequences. I chose intergenic sequences from the 5' and 3' end of each of the 4 *Hox* clusters based on the availability in phylogenetically important taxa. The intergenic sequences chosen for analysis were sequences between *Hox*A13 and *Hox*A11, *Hox*A5 and *Hox*A4, *Hox*B9 and *Hox*B8, *Hox*B3 and *Hox*B2, *Hox*C12 and *Hox*C11, *Hox*C8 and *Hox*C6, *Hox*D12 and *Hox*D11, and *Hox*D4 and *Hox*D3. Comparisons were made between teleost fishes; Japanese pufferfish (Tru) and spotted green pufferfish (Tni), medaka (Ola), and zebrafish (Dre). Comparisons were then made between mammal species with approximately similar divergence dates; human (Hsa) and baboon (*Papio hamadryas*, Pha), dog (Cfa), and opossum (*Monodelphis domesticus*, Mdo). Intergenic sequences were aligned using clustalW algorithm in MacVector version 9.0 (MacVector, Inc.). Jukes-Cantor D-values were calculated from these alignments using Mega version 3.1.

Approximate divergence times between two species that are necessary when performing comparative studies for conservation of non-coding DNA sequences can differ depending on generation times and mutation rates of the taxa selected. For mammals a general time of at least 250 million years divergence was determined by Tagle et al. (1988) for performing phylogenetic footprinting analysis. For other taxa divergence times have not been well worked out. In this study most of the species comparisons fit into the criteria of at least 250 million years divergence except for the comparisons among teleosts (zebrafish, acanthomorphs). For this study it appeared that the intergenic regions of the *Hox* clusters of the acanthomorph species (striped bass, nile

tilapia, pufferfishes; diverged ~80-100 mya) were too similar to do a phylogenetic footprint comparison but when aligned with zebrafish which diverged ~135 million years ago (mya) from the acanthomorphs the intergenic regions appeared to be dissimilar enough to do phylogenetic footprint comparisons. For this reason we decided to compare intergenic region distances between teleosts with a mammalian proxy to see if indeed their intergenic regions were diverging quicker.

Figure 12 shows the Jukes-Cantor D-values for teleosts fish versus their mammalian proxies. Immediately apparent is the lesser values for the mammalian proxies which implies that either the teleosts have a much higher mutation rate, much shorter generation time, or the proxies are incorrect due to inaccurate divergence time estimates. The D-values of zebrafish versus Japanese pufferfish allowed us to determine that these species were divergent enough to perform a phylogenetic footprinting analysis. Therefore pair-wise alignments containing zebrafish versus the acanthomorphs were included despite the relatively recent divergence date.

Another pattern that is not as easily noticeable is the difference between the 5' and 3' ends of the clusters. The odd numbers which correspond to the 5' ends of the clusters all have greater D-values than the even numbers which correspond to the 3' ends of the cluster. This greater conservation at the 3' end is also noticeable in Figure 3 where the deeply conserved PFCs are biased towards the 3' end. This data supports the posterior flexibility hypothesis put forward by Ferrier et al. (2000) which states that the posterior *Hox* genes (9-14) are evolving faster than the anterior (1-8) and implies this is occurring, in general, at the posterior end of the *Hox* gene cluster.

3.4 Conclusion

Evolution at the most basic level is change in allele frequency that becomes fixed over time. Empirical and theoretical evidence suggest that selection at the population level is most likely to occur in *cis*-regulatory elements. Because of this it is important to understand the code and evolution of these non-coding elements. Without a standardized code for *cis*-regulatory elements and since they do not code for an identifiable alternative molecule, their identification is not as straightforward as it is for coding DNA.

In an effort to uncover these molecular gems among the other DNA, unique methods have been developed. One such method is detection of evolutionary conservation of non-coding DNA. This method has proved to be advantageous and accurate although not without weaknesses or hurdles. One such hurdle is the fact that sequence conservation may not be necessary for functional conservation (Fisher et al. 2006). Unfornutately, this hurdle cannot currently be overcome. Another such hurdle is finding homologous non-coding sequences to compare since by definition the DNA surrounding the CNSs must not be conserved. A unique molecular phenomenon that can help overcome this issue is the clustering of genes that remains intact over millions of years in disparate species resulting in stretches of definitive homologous non-coding DNA. One such system is the *Hox* gene cluster.

Despite the convenience, few authors attempt to find CNSs within the *Hox* gene clusters and those that have, concentrate on only a few species. In this study, we have done a complete scan of gnathostome *Hox* gene clusters' intergenic regions to identify as many putative CREs as possible while gaining insight into the evolution of these

sequences in jawed-vertebrates with specific focus on the ray-finned fishes and the highly speciose, diverse and derived group, teleosts.

First, it is apparent that CNSs in the gnathostome *Hox* gene clusters adhere to the posterior flexibility hypothesis of *Hox* genes put forth by Ferrier et al. (2000). Second, we have extended the work done by Chiu et al. (2004) showing the mosaicism in basal ray-finned fishes *Hox* gene clusters. The complete scan of gnathostome *Hox* clusters has revealed additional data supporting this mosaicism but has also shown that the turnover that causes this mosaicism has been continuous throughout the ray-finned fishes from the basal bichir to the highly derived teleosts family, **Percomorpha**. This is shown by the fact there are no CNSs that are conserved exclusively in teleosts. Zebrafish, an intermediate teleost, appears to be as mosaic as bichir when comparing its conservation to species that diverged before to that of species that diverged after. The turnover of CNSs contributing to this mosaicism is cluster specific indicating that the *Hox* gene cluster system does not necessarily evolve under the same pressures and circumstances across all clusters. Third, in an effort to formalize the identification of CNSs within the *Hox* gene cluster system we have proposed a flexible naming system as Scott (1993) did for the *Hox* genes.

Chapter 4- Defining a Cause of Uniqueness

4.1 Introduction

4.1.1 Hindbrain Development

What is happening at the molecular and population level that may explain this turnover in conserved non-coding sequences in ray-finned fishes? To help answer this question we have turned to a well-characterized enhancer controlling the expression of *Hox*A2 in the developing hindbrain. We will start with a brief introduction of hindbrain development and the known expression of *Hox*A2 in vertebrates. We will then perform both macro- and micro-evolutionary studies of the enhancer to help determine possible mechanisms of its seemingly dynamic expression patterns in teleosts.

The vertebrate central nervous system (CNS) develops from neuroectoderm that thickens into the neural plate that rolls into the neural tube. This process occurs in an anterior to posterior pattern. The CNS is ultimately divided into the forebrain, midbrain, hindbrain and spinal cord. During the development of the hindbrain the neural tube is divided into distinct segments called rhombomeres much like the mesoderm is divided into distinct segments termed somites. *Hox* genes are expressed in the developing CNS starting with an anterior border at rhombomere 2 (r2) in the hindbrain and ending at the posterior end of the spinal cord. More specifically group 1-4 genes are restricted to anterior expression limits in the hindbrain whereas group 5-13 genes all have an anterior limit in the spinal cord (Gaunt et al. 1989; Hunt et al. 1991; Keynes and Krumlauf 1994). The peripheral nervous system (PNS) develops from a group of cells that is derived from an area between the thickened neural plate and the overlying ectoderm called the neural crest. Neural crest cells migrate from a dorsal to ventral position and convert from

neuroectoderm into mesenchymal cells which give rise to numerous structures. Pharyngeal arches (or branchial arches in fish) develop from these migrating neural crest cells. The first pharyngeal arch (PA1) arises from r2, the second pharyngeal arch (PA2) arises from r4 and the third pharyngeal arch arises from r6. The first pharyngeal arch will become the lower oral jaw and the second pharyngeal arch will become the hyoid in tetrapods and structures supporting the lower oral jaw in fish. Pharyngeal arches 3-6 or 7 become parts of the pharynx and trachea in tetrapods and gills and pharyngeal jaws in fish.

4.1.2 *Hox* A2 Expression

*Hox*A2 is expressed in the developing hindbrain with an anterior limit of expression at the r1/r2 boundary and continues to be expressed through r7 in mouse and chick embryos (Hunt et al. 1991; Krumlauf 1993; Prince et al 1994; Figure 13). Mouse *Hox*A2 has stronger expression in r3, 5 whereas chick has strong expression r3-7. Immediately following the FSGD there were two *Hox*A2 paralogs dubbed *Hox*A2a and *Hox*A2b. Since the FSGD, both paralogs have been independently pseudogenized in fish with a MRCA of 135 mya (zebrafish *Hox*A2a, Prince et al. 1998; medaka *Hox*A2b, Davis et al. 2008) as well as maintained as paralogs with differing expression patterns in fish with a MRCA of ~100 mya (Japanese pufferfish, Amores et al. 2004; Tumpel et al. 2006; striped bass, Scemama et al. 2006; Nile tilapia, Le Pabic et al. 2007). Zebrafish *Hox*A2b gene maintains a near identical ancestral expression pattern in r2-5 with stronger expression in r2, 3 (Prince et al. 1998). The same is true for the *Hox*A2b gene of Japanese pufferfish (Amores et al. 2004, Tumpel et al. 2006). The *Hox*A2a gene of Japanese pufferfish however has lost almost all hindbrain expression and appears in a

small stripe of cells in r1, 2. Striped bass and Nile tilapia *Hox*A2a genes are expressed in r2-7 with stronger expression in r2, 3 similar to the *Hox*A2b genes of zebrafish and pufferfish (Scemama et al. 2006, LePabic et al. 2007). The *Hox*A2b genes of striped bass and Nile tilapia are more weakly expressed. Striped bass *Hox*A2b gene is expressed in r2-5 and Nile tilapia *Hox*A2b is expressed in r2-5, 7 with stronger expression in r3, 5 (Scemama et al. 2006; LePabic et al. 2007; Figure 14).

*Hox*A2 is also expressed in the peripheral nervous system (PNS), specifically neural crest cells derived from r4 that pattern PA2 (Hunt et al. 1991; Krumlauf 1993; Prince et al 1994). This expression is maintained among tetrapods *Hox*A2 genes but again has various patterns in the teleosts' *Hox*A2 paralogs. In zebrafish, which has a pseudogenized *Hox*A2a, only the *Hox*A2b paralog is expressed and maintained in neural crest cells (Prince et al. 1998). The same is true for pufferfish *Hox*A2b expression (Amores et al. 2004; Tumpel et al. 2006). In striped bass and Nile tilapia, both paralogs are expressed weakly at the onset but only *Hox*A2a is maintained in detectable levels (Scemama et al. 2006, LePabic et al. 2007). Medaka, which has a pseudogenized *Hox*A2b, follows the same pattern as striped bass and Nile tilapia (Davis et al. 2008).

Controlling the expression of *Hox*A2 genes are 3 distinct enhancers (Figure 15). Expression in r2 is controlled by an enhancer in exon 2 of the *Hox*A2 gene, expression in r4 is controlled by an enhancer within the intron of the *Hox*A2 gene, and expression in r3, 5 and in r4 derived neural crest cells is controlled by an enhancer ~1-3kb upstream of the *Hox*A2 gene transcription initiation site (Nonchev et al. 1996; Maconochie et al. 2001; Tumpel et al. 2006). Within each of these enhancers are distinct regulatory elements that drive partial expression.

The r3/5 enhancer is the only known *Hox*A2 enhancer that lies outside the transcriptional start and stop sites. The r3/5 enhancer is divided into 5 elements that control rhombomere expression that are known as 'response elements' (RE1-5, Nonchev et al. 1996; Maconochie et al. 2001; Tumpel et al. 2006) and 4 elements that control neural crest and pharyngeal arch expression that are known as 'neural crest elements' (NC1-4, Maconochie et al. 1999). Some upstream transcription factors have been identified that bind to these *cis*-regulatory elements and are necessary for proper *Hox*A2 expression and rhombomere and pharyngeal arch development. There appears to be 1 specific transcription factor for each of the three compartments where *Hox*A2 expression is driven by the r3/5 enhancer. Krox-20 is normally expressed in tissue that will become rhombomeres 3 and 5 (Wilkinson et al. 1989; Nieto et al. 1991; Bradley et al. 1993; Oxtoby and Jowett 1993). Krox-20 (a.k.a. Kruppel, egr2) has two binding sites in the r3/5 enhancer sequence and directly regulates expression of *Hox*A2 in r3 and partially in r5 (Nonchev et al. 1996). Mafb (a.ka. Kreisler, valentino) is normally expressed in and is necessary for r5/6 development (Prince et al. 1998b; Sadl et al. 2003). Mafb has multiple binding sites within the r3/5 enhancer and though it has not been shown that it directly regulates *Hox*A2, *Hox*A2 expression is absent in r5 in Mafb $^{(-/-)}$ zebrafish (Prince et al. 1998b). AP-2 is expressed in the neural crest cells and their derivatives (Mitchell et al. 1991). AP-2 has multiple binding sites in the r3/5 enhancer and, though it has not been shown to directly regulate *Hox*A2 expression most likely due to redundant AP-2 family members, the r3/5 enhancer has been shown to drive expression of a reporter gene in the same pattern as AP-2 expression (Maconochie et al. 1999). The combination of these 3 main transcription factors and several co-factors that have not yet been identified are

responsible for the total expression of *Hox*A2 in r3 and 5 and neural crest cells derived from r4 in the developing hindbrain of all studied vertebrates.

The r3/5 enhancer has been studied in much detail via deletion and swap assays. Deletion assays in mouse done by Nonchev et al. (1996), Maconochie et al. (1999) and Maconochie et al. (2001) identified RE1-4 and NC1-4 as well as two essential Krox-20 binding sites and an element known as Box A. Both Krox-20 sites, Box A and RE1-3 are necessary for proper expression in r3 and 5 although r5 expression is only lost completely in Krox-20 site deletions (Maconochie et al. 2001). RE4 is shown to stimulate r5 expression in the presence of Krox-20 sites. Both RE1 and RE3 contain a TCT motif (TCTNAC) that appears to be the essential binding site for their function (Maconochie et al. 2001). A similar experiment was done for expression in the neural crest cells. NC1, 3 and 4 were shown to be essential for neural crest cell expression but did not affect r3, 5 expression (Maconochie et al. 1999). NC2 was shown to be essential for neural crest cell expression and although it did not affect r3 expression it did lower expression in r5 (Maconochie et al. 1999). It is important to note that NC2 and RE4 overlap as well as NC3 and RE3. NC3 and RE3 do not appear to overlap in essential positions though because absence of NC3 does not abolish r3, 5 expression like absence of RE3 (Maconochie et al. 1999; Maconochie et al. 2001). Most likely it is the TCT motif of RE3 that is responsible for driving expression in r3, 5. NC2 and RE4 both give a similar phenotype of reducing or ablating expression in r5.

Swap assays were done in the Japanese pufferfish (Tumpel et al. 2006). The Japanese pufferfish has both a *Hox*A2a and *Hox*A2b paralog although the *Hox*A2a paralog has lost almost all expression (Amores et al. 2004). Tumpel et al. (2006)

assumed that the loss of expression of *Hox*A2a was due to inactive enhancer elements. To take advantage of this they swapped out RE1-4 individually from the functional *Hox*A2b r3/5 enhancer into the non-functional *Hox*A2a r3/5 enhancer to determine how much each individual response element contributed to the overall r3, 5 expression levels. From this analysis they determined the fraction of embryos each response element was able to drive expression in and decided that this was the efficiency of that particular element. The results were as follows: Box A 0%, RE2 6%, RE3 67%, RE4 32% and a newly discovered RE5 33%. Although this percentage was calculated based on the fraction of embryos, the intensity of the staining was directly correlated. Combining these analyses it shows that RE3 is a major factor in r3, 5 expression whereas RE4 and RE5 are minor factors.

4.2 Expression Analysis of *Hox* A2 Paralogs

Materials and Methods

We used *in situ* hybridization using mRNA probes on whole mounted embryo to detect expression of *Hox*A2a and *Hox*A2b in a tissue and stage specific manner. *Hox*A2a, *Hox*A2b and EGR2 were isolated from *F. heteroclitus* gDNA (gift from D. Duvernell) using PCR protocol as described in section 2.2.1 with the following modifications: *Hox*A2a was isolated with *Hox*A2a exon 2 forward (5'-TGACCGAGAGGCAGGTCAAGG-3') and *Hox*A2a exon 2 reverse (5'-AGGGCAGAGGGGCTGTCACC-3') primers at an annealing temperature of 56 C; *Hox*A2b was isolated with *Hox*A2b forward (5'-CCTGACATCTCTCGCTAACC-3') and *Hox*A2b reverse (5'-AGAGGTCAGACGCTGCTGC-3') primers at an annealing temperature of 59 C; EGR2 was isolated with EGR2 forward (5'-CCAGACCTTYACCTAYATGGG-3') and EGR2 reverse (5'-TGTGTCTCTTYCTCTCRTCGC-3') primers at an annealing temperature of 59 C to yield a 400bp, 250bp and 955bp fragment, respectively. DNA fragments were cloned into the pGEM T-vector (Promega). RNA probes were synthesized using Roche T7 RNA polymerase (10881767001) that primes from the pGEM T-vector T7 site. *Hox*A2a andA2b RNA were synthesized using Roche DIG RNA labeling mix (1127703910). EGR2 was synthesized using Roche Fluoroscein RNA labeling mix (11685619910). RNA was purified using standard Phenol/Chloroform extraction. *In-situ* hybridizations were performed as described (Thisse and Thisse 2008) with the following modifications. Embryos were de-chorionated and de-yolked after rehydration and before digestion. Embryos were then digested in proteinase K for 50 seconds per somite. Antibodies were

pre-absorbed in blocking buffer for 2 hours at 4 C while rocking. Following overnight incubation with antibody, embryos were washed 6 times for 30 min each with gentle agitation. Embryos were stained in ceramic spot plates in the dark.

For two-color staining, both EGR2 and the *Hox*A2(a or b) RNA probe were hybridized to the embryo simultaneously. The 1st stain was conducted for EGR2 via Roche anti-fluorscein-AP Fab fragments with Sigma Fast*TM* FastRed (Sigma). The embryo was then washed 2 times for 30 minutes each in 1X PBT followed by a 30 minute wash in AP inactivation solution (0.1 M glycine-HCl pH 2.2, 0.1% Tween 20), followed by 4 times for 30 minutes in 1XPBT. The 2nd stain was conducted for *Hox*A2(a or b) via Roche anti-DIG-AP Fab fragments with NBT/BCIP stain providing a contrasting blue/purple color. Stained embryos were imaged using a Leica DFC290 digital camera attached to a Leica MZ12.5 stereo dissecting microscope.

In-situ hybridization analysis was performed with mRNA of *Hox*A2a and *Hox*A2b genes of mummichog (*Fundulus heteroclitus*) on *F. heteroclitus* embryos from 4 distinct developmental stages to determine the pattern of expression the *Fundulus sp.* group has compared to the patterns seen in other teleosts lineages. Both the *Hox*A2a and *Hox*A2b gene have a similar onset of expression in r2, 3 seen in figures 16m, n respectively. At the 8-somite stage *Hox*A2a and *Hox*A2b show weak expression in r2-4 and PA2 (Figures 16a-d). The strongest expression is seen at the 14-somite stage (as is in other teleosts; Prince et al. 1994; Scemama et al. 2006; Le Pabic et al. 2007; Davis et al. 2008) where *Hox*A2a is expressed in r2-7 with stronger expression in the more anterior rhombomeres and *Hox*A2b in r2-5 with weak expression in r5 (Figures 16e-h). In difference to other teleosts both *Hox*A2a and *Hox*A2b are strongly expressed in PA2 at

the 14-somite stage (Figures 16g, h). Expression was also detected at the 21-somite stage to determine maintenance of expression. At this stage *Hox*A2a and *Hox*A2b are again similar with expression in r2-5 and PA2 although *Hox*A2b appears to have slightly stronger expression in PA2 (Figures 16i-l).

Compared to other teleosts *F. heteroclitus* maintains the ancestral expression of the *Hox*A2a gene. This is deduced from known expression data where striped bass, Nile tilapia and medaka all have expression in r2-7 with stronger expression more anteriorly and also strong expression in PA2 (Scemama et al. 2006; Le Pabic et al. 2007; Davis et al. 2008). This assumes that the psuedogenization of *Hox*A2a in zebrafish and the near complete loss of expression in Japanese pufferfish were independent events, which is the most parsimonious conclusion (Prince et al. 1998a; Amores et al. 2004). The *Hox*A2b paralog however shows continued variation that is seen in other teleosts. In *F. heteroclitus* the *Hox*A2b gene is expressed strongly in r2-4 but weaker in r5 (Figure 16f, h). For the other known teleosts expressions, striped bass has weak expression in r2-5, Nile tilapia has weak expression in r2, 4, 7 and strong expression in r3, 5, medaka has a pseudogene, Japanese pufferfish has strong expression in r2-5 and zebrafish has weak expression in r4, 5 and strong expression in r2, 3 (Scemama et a. 2006; LePabic et al. 2007; Davis et al. 2008; Amores et al. 2004; Prince et al. 1998). Despite the differing expressions among the *Hox*A2b genes of teleosts in the rhombomeres they all show weak expression in PA2 except for *F. heteroclitus*, which has strong expression. This is the first instance of strong expression of *Hox*A2b in PA2 and also the first instance of definite and strong expression of both *Hox*A2 paralogs in PA2. Due to the differing expression patterns of *Hox*A2b and the strong conservation of *Hox*A2a expression we

furthered our study to look for evidence of selection in the enhancer sequences of these paralogs.

4.3 Mutation Rate Analysis of *Hox* A2 Loci

4.3.1 Phylogenetics of *Hox* A2 Loci

Materials and Methods

To determine mutation rates among *Hox*A2 loci we first created phylogenetic trees of each locus. Nucleotide alignments were performed using the clustalW algorithm provided by MacVector® 9.5.2. The alignments were corrected by eye so that all sequences started at the same position. 4 types of gene trees were constructed including neighbor-joining, maximum parsimony, maximum likelihood and Bayesian. Although neighbor-joining and maximum parsimony are considered antiquated there are some cases in which they have been shown to be more accurate (Landan and Graur 2007). For the more sophisticated tree building methods it was necessary to find an optimal model and prior inputs.

ModelTest v3.7 (Posada and Crandall 1998) was used to select the likelihood model for the nucleotide data. The model selected for the *Hox*A2 alignment by the hierarchical likelihood ratio tests (hLRT) (Gaut and Weir 1994) was HKY+G and by the Bayesian information criterion (BIC) (Raftery 1986a, 1986b) was HKY+G. We chose to use the HKY+G model which estimated the following parameters: Base frequencies; A=0.2333, C=0.3466, G=0.2580, T=0.1620. Ti/Tv ratio of 1.4986. Among site rate variation; proportion of invariable sites=0, Gamma distribution shape parameter=0.8828. The model selected for the *Hox*A2 enhancer alignment by the hierarchical likelihood ratio tests (hLRT) (Gaut and Weir 1994) was K80+G and by the Bayesian information criterion (BIC) (Raftery 1986a, 1986b) was K81+G. We chose to use the K80+G model, which estimated the following parameters: Base frequencies; equal frequencies. Ti/Tv

ratio of 0.9952. Among site rate variation; proportion of invariable sites=0, Gamma distribution shape parameter=1.0289.

MrModelTest v2.2 (Nylander 2004) was used to select the likelihood model for the nucleotide data that could be implemented by MrBayes. The model selected for the *Hox*A2 alignment by hLRT, hLRT2, hLRT3 and hLRT4 was HKY+G. The model selected by AIC was HKY+G. We chose to use the HKY+G model which when implemented in MrBayes gives a prior state frequency of dirichlet (1,1,1,1) and uses the basic model nst=2 with among site rate variation set to estimate rates based on a gamma shaped distribution. The model selected for the *Hox*A2 enhancer alignment by hLRT, hLRT2, hLRT3 and hLRT4 was K80+G. The model selected by AIC was HKY+G. We chose to use the K80+G model which when implemented in MrBayes gives a prior state frequency of fixed (equal) and uses the basic model nst=2 with among site rate variation set to estimate rates based on a gamma shaped distribution.

Neighbor Joining trees for the nucleotide alignments were constructed using PAUP* v4.0b10 (Swofford 2003). The *Hox*A2 and *Hox*A2 enhancer trees were constructed under the following assumptions; distances were calculated using the Tamura-Nei method (Tamura and Nei 1993), using parameters specified from models selected above. The objective function was set to minimum evolution and TBR was selected as the swapping algorithm. Trees were constructed using the Neighbor Joining method (Saitou and Nei 1987) and then 2000 bootstraps were conducted to assess node confidence. A consensus tree was created using majority rule with a 50% cutoff.

Maximum parsimony trees for nucleotide alignments were constructed using PAUP* v4.0b10 (Swofford 2003). All nucleotide trees were constructed under the

default settings using a heuristic search and node confidence was assessed with 2000 bootstraps. A consensus tree was constructed using majority rule with a 50% cutoff.

Maximum likelihood trees for the nucleotide data of both alignments were constructed using GARLI v0.951 (Zwickl 2006), which can be downloaded from www.bio.utexas.edu/faculty/antisense/garli/Garli.html. The starting tree was obtained using heuristic search under the likelihood optimality criterion in PAUP* v4.0b10 (Swofford 2003) with settings as specified by Modeltest (Posada and Crandall 1998, see Model Selection). The same settings were inputted into GARLI v0.951 (Zwickl 2006). Node confidence was assessed by using the bootstrap resampling method with 2,000 replications. The "number of generations without improving topology" setting was changed from 10,000 to 5,000 as suggested in the GARLI manual when doing bootstraps as it will shorten the run time without significantly affecting the results. A consensus tree was created using majority rule with a 50% cutoff in PAUP* v4.0b10 (Swofford 2003).

Bayesian trees for the nucleotide alignments were constructed using MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003) and the parallel version of MrBayes v3.1.2 (Altekar et al. 2004) under the settings specified by MrModeltest (see Model Selection). Two independent Markov Chain Monte Carlo analyses were run with the following settings: number of generations was set to 1,000,000, sample frequency was taken every 1,000 steps, number of chains was set to 4 and the temperature was set 0.2. 'Burnin' was assessed after the run using the sum parameters command. The 'burnin' for the nucleotide analysis was set to 1 for both datasets, which is equal to the first 1,000 steps or tree topologies. A majority rule consensus tree was created disregarding the 'burnin' trees using the sum trees command with a cut off of 0.50 posterior probability.

Phylogenetic analysis indicated that *Hox*A2a, *Hox*A2b, *Hox*A2a enhancer and *Hox*A2b enhancer sequences isolated from *F. heteroclitus* as well as northern studfish (*Fundulus catenatus*) and the two species that were later used in population level studies, black-striped topminnow (*Fundulus notatus*) and black-spotted topminnow (*Fundulus olivaceus*) were true paralogs and were the result of the same duplication as the rest of the teleosts' *Hox*A2 sequences, FSGD (Figure 17). Overall, gene trees appear to agree with accepted species phylogeny.

4.3.2 Mutation Rate Analysis

Gene trees and alignments were used as input data for a character state tree constructed using MacClade v4 (Maddison and Maddison 1989) for the 156 bp spanning NC2 to RE3 which also encompasses the entire RE4 and 5' half of NC3. From this analysis it is apparent that the 'b' paralog has acquired more changes than the 'a' paralog and at a seemingly steady rate (Figure 18). The changes in the 'a' paralog are concentrated at the base of the teleosts (35) prior to the rapid speciation with significantly (p=0.00468) fewer changes in all of the following internal and terminal branches (34 total) as compared to identical branches in the 'b' paralog (53 base and 122 other). The 'b' paralog however, has a continuous change in characters with no less than 23 on any branch. The *Fundulus sp.* 'b' paralog has acquired more changes than any other 'b' paralog terminal with at least 36 changes. This evidence of continued turnover in the 'b' paralog binding sites well into the *Fundulus sp.* lineage and lack thereof in the 'a' paralog is in contrast to the more conserved nature of the 'b' paralog sequence between the two in-group species of *Fundulus sp*. This is more apparent taking into consideration changes in highly conserved binding sites. When comparing only those sites that have one step in

the tree (those sites where all taxa except for one share the same state) the *Fundulus sp.* ‘b’ paralogs have at least twice as many steps as any other fish/paralog. This indicates that leading to the *Fundulus sp.* lineage changes in highly conserved nucleotides were at least twice as likely to become fixed then in any other fish lineage. This is further supported by the fact that changes within the *Fundulus sp.* group at the ‘b’ paralog enhancer are concentrated around functional sites. This same pattern can be seen in the *Hox*A2 coding sequences as well.

4.4 Population Genetic Study of *Fundulus Hox* A2 Enhancer Evolution

4.4.1 Specimen Acquisition

Materials and Methods

To perform a population genetic study on *Fundulus* it was necessary to acquire wild *Fundulus* fish because no lab strains exist, we needed individuals from two sister species and lab strains are unlikely to maintain the variation that existed before they were institutionalized. For the purpose of the population genetic study we collected individuals from two sister species of killifish, black-striped topminnow and black-spotted topminnow from Mississippi River drainages in Illinois and Missouri with the help of Dr. David Duvernell of University of Southern Illinois-Edwardsville and Robert Lynch (lab member). *F. olivaceus* was collected from the west side of the Mississippi River (Missouri) from 3 separate locations comprising two separate river drainage systems. These species were chosen due to the glut of knowledge of their phylogeography, well studied relationship, location within the United States and availability of expert on-site help. From the 1st drainage system we collected 93 individuals from one site, which was Rockford Beach at Big River at House Springs (38° 25’ 20.27” N, 90° 35’ 23.58” W, from here on known as Rockford Beach). From the 2nd drainage system we collected 29 individuals from two sites which were Little Piney Creek (5 individuals) at Highway 44 bridge at Jerome (37° 55’ 04.68” N, 91° 58’ 16.66” W, from here on known as Little Piney) and Gasconade River (24 individuals) at Jerome (37° 56’ 05.34” N, 91° 58’ 39.24” W, from here on known as Gasconade River). *F. notatus* was collected from the east side of the Mississippi River (Illinois) from 3 separate locations comprising two separate river drainage systems. From the 1st drainage system

we collected 42 individuals from one site, which was Salt Creek at Highway 40 bridge east of Effingham (39° 07’ 15.49” N, 88° 31’ 24.00” W, from here on known as Salt Creek). From the 2nd drainage system we collected 64 individuals from two separate locations which were Kaskaskia River at Highway 2700N north of Vandalia (39° 07’ 53.79” N, 90° 35’ 23.58” W, from here on known as Kaskaskia River) and Big Creek at Highway 2100E and 2600N intersection north of Saint Elmo (39° 06’ 57.32”N, 88° 51’ 42.81” W, from here on known as Big Creek). Approximately 12 individuals of an out-group taxa northern studfish were also collected at Little Piney. The fish were collected on September 16th and 17th, 2006. Collection of individuals was done using seines and hand-nets. We preserved individuals on-site in 95% ethanol. Individuals were later placed into separate, numbered storage bottles in 95% ethanol.

We performed DNA isolation of individuals using fin clips under standard tissue preparation protocols using the AquaPure Genomic DNA Tissue Kit by Bio-Rad.

4.4.2 Gene Isolation

Materials and Methods

We isolated *Hox*A9b, A3a, A2a, A2b and B2a genes; *Hox*A2a and *Hox*A2b r3/5 enhancers as well as phosphoglycerate kinase 1 (PGK1) from *F. notatus* and *F. olivaceus* using PCR protocols as described in section 2.2.1 with the following modifications: *Hox*A9b was isolated with A9b forward (5’-CTTTGGAGACGCACACWCC-3’) and A9b reverse (5’-TTCTTCATCTTCATYCTGCGG-3’) primers at an annealing temperature of 57 C; *Hox*A3a was isolated with A3a forward (5’-ARTACAARAAGGATCAGAAAGG-3’) and A3a reverse (5’-TTRCCCATTGTGATTGCTCC-3’) primers at an annealing temperature of 57 C;

*Hox*A2a was isolated with A2a forward (5'-GAATTCGARCGAGAGAGCG-3') and A2a reverse (5'-ACRGGTCCGTTKGAGATGG-3') primers at an annealing temperature of 56 C; *Hox*B2a was isolated with B2a forward (5'-CATTTCAAACTTCATCAATCAAGG-3') and B2a reverse (5'-CTCYTTCATCCAKGGRAACTC-3') primers at an annealing temperature of 58 C; PGK1 was isolated with PGK1 intron 7 forward (5'-GCRAAGGTGAAAGATAAGATYCAGC-3') and PGK1 intron 7 reverse (5'-TCTCSGCTTTGGCCATCAGG-3') primers at an annealing temperature of 62 C.

*Hox*A9b and *Hox*A3a were subjected to the same phylogenetic analysis as *Hox*A2a and *Hox*A2b to determine their paralog status with the following gene specific input data: Regarding ModelTest v3.7 (Posada and Crandall 1998), the model selected for the *Hox*A3a alignment by the hierarchical likelihood ratio tests (hLRT) (Gaut and Weir 1994) was HKY+G (Hasegawa et al. 1985) and by the Bayesian information criterion (BIC) (Raftery 1986a, 1986b) was HKY+G. We chose to use the HKY+G model which estimated the following parameters: Base frequencies; A=0.2236, C=0.3432, G=0.2357, T=0.1975. Ti/Tv ratio of 1.4676. Among site rate variation; proportion of invariable sites=0, Gamma distribution shape parameter=0.6010. The model selected for the *Hox*A9b alignment by the hierarchical likelihood ratio tests (hLRT) (Gaut and Weir 1994) was HKY+I+G and by the Bayesian information criterion (BIC) (Raftery 1986a, 1986b) was HKY+I+G. We chose to use the HKY+I+G model which estimated the following parameters: Base frequencies; A=0.2923, C=0.2753, G=0.2391, T=0.1933. Ti/Tv ratio of 1.4986. Among site rate variation; proportion of invariable sites=0.2648, Gamma distribution shape parameter=1.7812. Regarding

MrModelTest v2.2 (Nylander 2004), the model selected for the *Hox*A3a alignment by hLRT, hLRT2, and hLRT4 was HKY+G, hLRT3 was GTR+G (Lanave et al. 1984). The model selected by AIC was GTR+G. We chose to use the HKY+G model which when implemented in MrBayes gives a prior state frequency of Dirichlet (1,1,1,1) and uses the basic model nst=2 with among site rate variation set to estimate rates based on a gamma shaped distribution. The model selected for the *Hox*A9b alignment by hLRT2, hLRT3, and hLRT4 was HKY+I+G, hLRT was GTR+I+G. The model selected by AIC was HKY+I+G. We chose to use the HKY+I+G model which when implemented in MrBayes gives a prior state frequency of dirichlet (1,1,1,1) and uses the basic model nst=2 with among site rate variation set to estimate rates based on a gamma shaped distribution and proportion of invariable sites.

To empirically link the *Hox*A2 enhancers to the genes they regulate the entire intergenic region of *Hox*A3a to *Hox*A2a was isolated in *F. catenatus*. We used Takara LA Taq due to its ability to amplify longer DNA fragments with the following volumes and concentrations of reagents: 5.0 ul 10X buffer, 4.0 ul $MgCl_2$ (25 mM), 4.0 ul dNTP (10 mM), 1.0 ul each of primers (25 pmol/ul), 1.0 ul gDNA (250 ng/ul), and 1.0 ul LA Taq (5 U/ul). We conducted the PCR reaction under the following conditions: 95 C for 2’, 30 cycles of (95 C for 30”, 55 C for 1’, 68 C for 4’), 68 C for 5’ and 10 C hold. Due to unknown product size and multiple bands we performed a southern blot and hybridization.

We performed the southern blot procedure by transferring the PCR reaction from a 1% agarose gel to Amersham pharmacia biotech’s Hybond-N+ nylon membrane. We performed transfer of DNA from the agarose gel to the nylon membrane as follows:

Soaked gel in southern denaturation solution (0.2 M NaOH, 0.6 M NaCl) for 30' on a rocker. Soaked gel in neutralization solution for 30' on a rocker. Placed two sheets of Whatman's paper cut to size of gel in a glass dish, filled to top of Whatman's paper with 20X SSC (3.0 M NaCl, 0.3 M sodium citrate), place gel on top of paper, place Hybond-N+ membrane on gel, place two more pieces of Whatman's paper followed by 5 inches of paper towels and a ~1lb weight, add 20X SSC 1 hr later, wrapped entire assembly in plastic wrap and let sit overnight. We cut all paper and membrane to the exact size of the gel. We then let the membrane dry and pre-hybridized overnight in 10ml hybridization solution (5X SSC, 0.1% *v/v* N-Lauroylsarcosine, 0.2% *v/v* sodium dodecyl sulfate, 1% *w/v* blocking solution (Roche Diagnostics Blocking reagent)) at 65 C in rotisserie incubator along with a positive filter consisting of *Hox*A3a fragment on the nylon membrane. We prepared the positive filter by placing 1 ul (100 ng/ul) of fragment on a nylon membrane, drying, then floating in southern denaturation solution for 3' followed by floating in neutralization solution for 3'.

We used the 451 bp *Hox*A3a fragment isolated from *F. catenatus* genomic DNA as a probe. We DIG (Digoxigenin)-labeled the *Hox*A3a fragment using Roche Diagnostics DIG DNA labeling kit (11175033910). The DIG labeling reaction includes boiling 800 ng (15 ul) of the *Hox*A3a fragment in water for 10', placing on ice (H_2O) for 10', briefly spinning down, then adding 2 ul of the hexanucleotides, 2ul dNTP labeling mix, and 1ul Klenow enzyme provided in Roche kit. We then placed the labeling reaction at 37 C for 20 hrs.

After pre-hybridization and probe labeling we re-boiled the probe for 10' and chilled on ice for 10'. We then added the 20 ul of probe to 10 ml of hybridization

solution and hybridized membranes overnight in rotisserie incubator. We then washed the membranes briefly in 2X SSC followed by 2 washes in 2X SSC, 0.1% *v/v* SDS for 20' at room temperature on a rocker followed by 2 washes in 0.5X SSC, 0.1% *v/v* SDS for 20' at 65 C in rotisserie incubator. We then rinsed filter in diluent buffer (300 mM NaCl, 0.1 M Tris pH 8.0). We followed this by a "blocking" step by placing membranes in 40 ml of diluent buffer, 4 ml 10% blocking solution for 1 hr on a rocker then an "anti-blocking" step by placing membranes in 30 ml of diluent buffer, 4 ml 10% blocking solution, 3 uL of Roche Diagnostics Anti-DIG-AP Fab fragments (0.75 U/ul). We followed this with 3 washes in a tween solution (150 ul tween 20, 50 ml diluent buffer) for 10' on a rocker. We followed this with washing in 50 ml of detection buffer (100 mM Tris pH9.5, 100 mM NaCl) for 5' on a rocker. We then prepared the membranes for exposure to Kodak Biomax light film. We prepared the membranes for exposure by adding Roche Diagnostics CDP-star to their surface. We then exposed the membranes to the film for 10' and then the film was developed on a Kodak developer.

Following identification of correct band (~6.5 kb) by southern hybridization we followed the protocol explained above to sequence the fragment. We had to create 3 sets of species-specific nested primers to sequence the entire fragment We used Takara LA Taq with the following volumes and concentrations of reagents: 5.0 ul 10X buffer, 4.0 ul $MgCl_2$ (25 mM), 4.0 ul dNTP (10 mM), 1.0 ul each of primers (25 pmol/ul), 1.0 ul gDNA (250 ng/ul), and 1.0 ul LA Taq (5 U/ul). We conducted the PCR reaction under the following conditions: 95 C for 2', 30 cycles of (95 C for 30", 55 C for 1', 68 C for 4'), 68 C for 5' and 10 C hold.

4.4.3 Population Level Gene Sequencing

Materials and Methods

*Hox*A2a enhancer, *Hox*A2b enhancer and a shorter fragment of PGK1 intron 7 were isolated from ~100 individuals of the *F. olivaceus* populations. These were done with Finnzymes Phusion Taq due to its 50 times greater accuracy over AB amplitaq. Phusion Taq PCR reactions, volumes and concentrations were as follows: 10.0 ul 5X HF buffer, 4.0-7.0 ul $MgCl_2$ (25mM), 1.0 ul dNTP (10 mM), 1.0 ul each of primers (25 pmol/ul), 1.0 ul gDNA (250 ng/ul), 0.5 ul Phusion Taq (2 U/ul) and sterile water to 50 ul. The PCR reaction for Phusion Taq was conducted under the following conditions: 98 C for 1', 30 cycles of (98 C for 30", annealing temperature (see table 2) for 30", 72 C for 30"), 72 C for 5' and 10 C hold. All PCR reactions were carried out in an Eppendorf Mastercycler® ep thermocycler. Following PCR, the PCR products were sent directly from the gel purification step to be sequenced. Approximately 70 ng of purified PCR product was sent. This method of direct PCR sequencing reduces the effect of base pair mispairing that can occur with DNA polymerases due to the fragments with the mispairing representing a minority of the fragments in the sequencing reaction.

4.4.4 Tests of Selection

We conducted both interspecies and intraspecies tests of selection on the *Fundulus sp.* group (Table 1). For interspecies comparisons we calculated the Ka/Ks ratio (Nei and Gojobori 1986). This test was performed on all coding regions that were sequenced (Table 1). These results were directly compared with the Kb/Ki ratio (Hahn 2007, see materials and methods), which was implemented for the non-coding *Hox*A2 enhancers. These ratios are an indicator of selective forces acting at these loci. A ratio of

one is an indicator that there is no selection and the nucleotides are neutrally evolving. More than one denotes positive selection and less than one denotes negative or purifying selection. The majority of loci indicate negative or purifying selection.

Since the *Hox* loci are master regulatory genes found throughout the animal kingdom, selection to maintain nucleotide sequence and thus protein function is likely. The non-*Hox* locus PGK1 also appears to be under purifying selection. The lone exception to the neutral theory of evolution is the comparison between in-group species *F. olivaceus* and *F. notatus* at the *Hox*A2b enhancer. This comparison yields an exceptionally high Kb/Ki ratio of 7.78 (p=0.03) indicating strong positive selection as described in Hahn et al. (2004). The positive selection is isolated to this node as the Kb/Ki ratios between either of these two sister species and the out-group species, *F. catenatus,* is close to and not significantly different than 1. Further indication of positive selection can be noted by comparing the non-coding/non-binding percent difference for each locus, which is an indicator of neutral evolution at each site.

As seen in table 1 the two sister species, *F. olivaceus* and *F. notatus* have a lower percent difference than either do to the out-group, *F. catenatus,* most likely due to the divergence dates of these species. The more notable difference is the divergence seen between sister species when comparing the two *Hox*A2 enhancers; *Hox*A2b enhancer 0.48% and *Hox*A2a enhancer 2.02%. Compared to the *Hox*Ab cluster genes which are 2.63% and 2.95% and the *Hox*Aa cluster genes which are 0.66% and 0.90% there seems to be an inverse relationship between enhancer and gene neutral mutation rates at the ‘a’ and ‘b’ paralog loci (Figure 19). This inverse relationship is not seen however in the comparison between either of the sister species to the out-group species. At this level the

*Hox*A2a enhancer is more conserved at 4.03% for both comparisons than is the *Hox*A2b enhancer, 4.76% and 5.24%.

Also, between the sister species no other locus has more than 1 fixed difference in the functional region whereas the *Hox*A2b enhancer has 4. This evidence that the neutral mutation rate was slowed at the non-binding sites in the *Hox*A2b enhancer between *F. olivaceus* and *F. notatus* strengthens the claim that nucleotides at binding sites were preferentially fixed in the population (positive selection) at this enhancer.

Following the method used by Hughes and Nei 1988, we also applied the Kb/Ki ratio test to the *F. olivaceus* within species data. The *Hox*A2a enhancer had a Kb/Ki ratio of 0.00 (p=0.59) indicating purifying selection. The *Hox*A2b had a Kb/Ki ratio of 1.94 (p=0.63) suggesting balancing selection although neither test was statistically significant. The neutral polymorphism at these two sites was less than the non-*Hox* locus with the *Hox*A2a enhancer at 0.76%, the *Hox*A2b enhancer at 0.48% and intron 7 of the PGK1 gene at 2.4%. We also performed conventional intraspecies tests of selection using the software programs Neutrality Test v1.1 (Li et al. 2003) and DnaSP v4.0 (Rozas et al. 2003) on the *Hox*A2b enhancer. These test were not performed on the *Hox*A2a enhancer because alleles could not be confidently assigned to individuals. These programs gave a Tajima's D score of 1.681 and 1.13474 standard deviations from the mean for the *Hox*A2b enhancer indicating an excess of high frequency polymorphisms that is interpreted as balancing selection, which was shown to not be significant. Fu and Li's D and F statistics were 0.68090 and 0.96107 respectively without an out-group and changed less than 0.001 when an out-group was included. Again these tests did not have significance ($p > 0.05$) but fell in the direction of balancing selection. In order to take

advantage of the multiple loci and species an HKA test was performed (results not shown) but the results again were not statistically significant. It appears that within species the *Hox*A2b enhancer is evolving under no or slightly balancing selection but between *F. olivaceus* and *F. notatus* there was strong positive selection.

We also performed a Hardy-Weinberg (H-W) equilibrium test on the *Hox*A2b enhancer of the *F. olivaceus* population to show that our sample was a random set of the population. This is an important step as many tests of selection assume random sampling. The *F. olivaceus* population has two distinct sub-populations gathered from three sites. The first sub-population was gathered from two sites along the same river drainage that are named Little Piney and Gasconade River (see materials and methods). The second sub-population was gathered from one site and is named Rockford Beach. The Little Piney/Gasconade River sub-population showed no allele variants at the *Hox*A2b enhancer. The Rockford Beach sub-population had two allele variants that were distinct from the one found in Little Piney/Gasconade River for a total of three alleles. These three alleles were found to be in Hardy-Weinberg equilibrium within the Rockford Beach sub-population (p=0.62). This verified that our sample of the population was not biased which could alter tests of selection that assume random mating.

Given the data we collected and the lack of information about the divergence dates between our two sister species we performed a divergence date estimation. The nuclear gene estimate of the divergence date between *F. olivaceus* and *F. notatus* was calculated by calculating the distance between 7 homologous regions (*Hox*B2a exon 1, *Hox*A9b exon 1, 2 and intron, *Hox*A2b exon 1, *Hox*A3a exon 2, and *Hox*A2a exon 1, *Hox*A2a enhancer and *Hox*A2b enhancer) totaling 1,395 bp. Divergence date estimations for the

three *Fundulus sp.* involved in this study were performed based on the % difference of the sequences accumulated. This was then compared to two pufferfish species (*Tetraodon nigroviridis* and *Takifugu rubripes*) with a known divergence date to estimate divergence date for the *Fundulus sp.* assuming an equal rate of mutation for these two **Acanthomorpha** clades. Overall *F. olivaceus* and *F. notatus* appear to have diverged around ~4.76 mya. This is the first nuclear data used to estimate divergence dates for these two species to our knowledge and agrees with the date derived from mitochondrial data ~2.75 mya (personal communication, Brian Kreiser).

4.4.5 *In-silico* Detection of Upstream Binding Factors

To better understand the expression patterns and evidence of selection it is necessary to know the nature of the enhancer. The *Hox*A2 R3/5 enhancer consists of 7 modules. RE2-5 help code for expression in r3 and 5 and NC1, 2 and 3 code for expression in neural crest cells derived from r4 and their derivative (PA2) (Nonchev et al. 1996; Maconochie et al. 2001; Tumpel et al. 2006). Three main transcription factors are thought to control *Hox*A2 expression via the R3/R5 enhancer: Krox-20, which controls expression in r3, 5 (Nonchev et al. 1996); Mafb, which control expression in r5 (Prince et al. 1998); and AP-2, which controls expression in PA2 (Maconochie et al. 1999). Since the domains of the enhancer that code for both r3 and 5 expression (RE4) and neural crest cell expression (NC2) overlap in the area of strong positive selection it is possible that we are seeing strong purifying selection elsewhere in the enhancer (RE2, 3, 5) (see divergence comparisons above) to maintain rhombomere expression in the presence of strong positive selection (RE4/NC2) to increase expression in the neural crest cells and their derivatives. This is supported by Tumpel et al. 2006, who showed RE4 is 32%

efficient at driving R3/5 expression and RE3 and RE5 were 67% and 33% efficient, respectively with RE1 and RE2 contributing negligible amounts. So maintaining strong purifying selection in RE3 and RE5 could compensate for the loss of expression in RE4 due to its overlap with the strong positive selection to modify NC2. It is also possible that upstream factors that normally bind to RE4 to drive expression in r3 and 5 have co-evolved with the nucleotide changes to maintain expression. A third possibility that compensatory mutations elsewhere in the enhancer are compensating for the loss of specificity at RE4 seems less likely since the rest of the enhancer is so well conserved. To help differentiate between the three possible scenarios TFBS prediction program JASPAR was used to analyze possible TFBS changes between *F. notatus* and *F. olivaceus* at the changes within RE4/NC2 (Sandelin et al. 2004; Figure 20). All 4 fixed differences resulted in different TFBS profiles and this region is flanked by two RUSH-1α sites (5' site and 3' site) that are present in all 4 species (Figure 20). Both 5' and 3' RUSH-1α sites have shifted toward each other by 2 nucleotides in the two sister species due to two mutations shared by these species that are not present in either *F. heteroclitus* or *F. catenatus* (Figure 20). Despite these mutations both sites have maintained similar scores in the JASPAR database. RUSH-1α is a SWI/SNF-related matrix-associated protein, which possesses nucleosome remodeling activity and may be necessary for basal transcription of target genes (Hewetson et al. 2002). RUSH transcription factors also contain a RING finger domain at the c-terminal end that plays a role in protein-protein interactions (Mansharamani et al. 2001). RUSH-1α binding sites share sequence similarity with FOXL1 and FOXC1 binding sites.

FOXL1 and FOXC1 are forkhead family transcription factors that play roles in anterior segment development (Hacker et al. 1995). All four species maintain either 1 FOXC1 or FOXL1 at both 5' and 3' sites but *F. notatus* has binding sites for both forkhead family member at both 5' and 3' sites (Figure 20). The 3' site also has affinity for Ultrabithorax (Ubx). Ubx is orthologous to *Hox*A7 in vertebrates and although there is no evidence that *Hox*A7 regulates hindbrain development it is worth noting that *F. olivaceus* loses this Ubx TFBS at the 3' site but gains it back at the 5' site. Both *F. olivaceus* and *F. notatus* independently gain an HMG-1 binding site. HMG-1 helps regulate transcription, differentiation and DNA repair (Bianchi and Beltrame 2000; Mitsouras et al. 2002). HMG-1 contains two DNA binding sites along with a C-terminal protein interaction domain that has been shown to interact with *Hox* proteins and enhance their DNA binding and transcriptional activation (Zappavigna et al. 1996). The Ubx and HMG-1 site are of importance due to the discovery that *Hox* genes often auto and cross-regulate (Hafen et al. 1984; Kuziora and McGinnis 1988). The most pertinent TFBS change occurs in *F. notatus* midway between the 5' and 3' site. At this site all species except for *F. notatus* maintain an Mafb binding site. Mafb also known as kreisler (mouse) or valentino (zebrafish) is expressed in the developing hindbrain and is essential for r5, 6 development and patterning (Prince et al. 1998b; Sadl et al. 2003). The 1 fixed difference between *F. olivaceus* and *F. notatus* that lies outside of any known functional element causes *F. notatus* to have an Mafb binding site that the other three *Fundulus sp.* do not. From this data there is evidence of both positive selection to gain binding sites as well as compensatory mutations to maintain them.

If we subject all other vertebrate *Hox*A2 enhancer sequences to the same type of analysis for Mafb binding sites as well as AP-2 and Krox20 binding sites a unique pattern emerges. Comparing binding sites of all three major R3/R5 enhancer transcription factors across vertebrate phylogeny allows us to see that RE3 is the most well conserved response element, whereas RE4 appears to be less conserved (Figure 21). This is in agreement with the idea that RE3 is responsible for a more basal expression whereas RE4 can act as a "fine-tuning" element. Tumpel et al. (2006) also showed this in their swap assays where RE3 recapitulated 67% whereas RE4 recapitulated 32% of the normal expression. Despite RE3 being the more conserved region the *Fundulus sp*. appear to be more variable than other species, specifically in the loss of highly conserved AP-2 binding sites. The two sister species have both lost an AP-2 binding site that is conserved among all enhancers except the *Hox*A2a enhancers, including the two other members of the *Fundulus sp.* group. It also appears that all of the Fundulus sp. have lost another AP-2 binding site in RE3 that is shared among all ray-finned fish including both 'a' and 'b' paralogs and also with chicken, frog and coelacanth. The region between the Krox-20 binding site and RE4 has been shown to be non-functional in both deletion analysis and swap assays (Maconochie et al. 2001; Tumpel et al. 2006). In this region a mutation in *F. notatus* adds back an Mafb binding site that no other *Fundulus sp*. member has. Also, the entire *Fundulus sp*. group lacks any AP-2 binding sites. This is possibly due to the ability for spurious binding sites to 'sequester' transcriptions factors away from the functional binding sites since the *Fundulus sp*. group specifically has lost 2 well-conserved AP-2 binding sites in RE3.

4.5 Discussion

*Hox*A2 is the most anteriorly expressed *Hox* gene in anterior-posterior patterning. The expression pattern of this gene has been shown for many vertebrates including birds, mammals and fish. In ray-finned fishes the expression pattern is made more complex due to the retention of two paralogs following the FSGD. Much work has gone into discovering the CREs responsible for driving expression of *Hox*A2 in the developing nervous system and derived structures. From this work a detailed map of the CREs has emerged. In this study we set forth to gather data that would help shine light on the evolution of one of these particular CREs. From this study we have discovered that the r3/r5 enhancer of the *Hox*A2 gene has undergone different rates of mutation throughout gnathostome evolution. More specifically, the retention of *Hox*A2 paralogs in the teleosts resulted in asymmetrical evolution. The *Hox*A2a r3/r5 enhancer appears to have undergone an accelerated rate of mutation fixation after the FSGD but prior to the **Ostariophysii/Acanthomorpha** species radiations. In contrast, the *Hox*A2b r3/r5 enhancer appears to have maintained an accelerated rate of mutation fixation prior to and throughout the teleost species radiation. The highest r3/r5 enhancer mutation rate is maintained in the *Hox*A2b locus of the *Fundulus* sp. complex. Classification of these mutation fixations indicates that the mutations in TFBS were preferentially retained. Evolutionary maintenance of the r3/r5 enhancers along with support from earlier expression assays indicates there are certain elements that may be more important to basal expression whereas other elements may act as ‘fine-tuning’ mechanisms. The *Fundulus* sp. complex is again unique with more loss of TFBSs in the basal elements than any other species/group.

In an effort to discover why/how the *Fundulus* sp. complex has so much nucleotide turnover in the otherwise more conserved and seemingly more basal regions of the r3/r5 enhancer we performed micro-evolutionary tests. From these test we discovered support for evidence of strong positive selection occurring within this group between two closely related sister species. More specifically we saw 4 fixed differences within 38 base pairs in the RE4/NC2 region. TFBS analysis indicates these changes resulted in the two sister species each independently gaining the same *Hox* protein-binding partner. The loss of a *Hox* protein-binding site was regained in one species but the loss of the TFBS for Mafb (a TF necessary for *Hox*A2 expression in r5) in the other species was not regained. This shows that turnover of CREs at the macro-evolutionary level that appear to be due to relaxation of constraint and probable neutral evolution may actually be driven at least in part by strong positive selection at the micro-evolutionary level. This reiterates the need for more research into the evolutionary mechanisms that are shaping our CREs rather than research aimed at detecting the results of those mechanisms. Emphasis on the level at which evolution actually occurs may help shed a brighter light on how non-coding DNA contributes to adaptation and ultimately lead us to a more accurate and useful genetic code for regulatory DNA.

If the selective pressure is indeed due to the need to increase PA2 expression of *Hox*A2b, then what might be the possible phenotypes on which selection is acting? *Hox*A2 has been shown to directly regulate bone morphogenetic protein 4 (BMP4; Smith et al. 2009). BMP4 is expressed in the lower oral jaw (LOJ) and is directly associated with altering the ratio between flexibility and bite force of the oral jaw (Albertson et al. 2005).

Liem's Hypothesis contains the idea, that since the adaptation of posterior pharyngeal arches into a pharyngeal jaw, that oral and pharyngeal jaws have become uncoupled and may evolve independently (Liem 1973). While we are not testing this hypothesis in this study it does help explain how such adaptations can occur without detriment to the species and how such a variation in *Hox*A2 expression could be tolerated within the teleosts.

Fundulus olivaceus and *F. notatus* have been well studied in their physical characteristics and habitat distribution (Petifils 1986). The only defined distinguishing characteristic between the two species is the presence of dark regular spots above the lateral line in *F. olivaceus* (Thomerson 1966). A possible explanation for the phenotypic difference that may be driving the positive selection for altered expression of *Hox*A2b in PA2 has to do with courtship. *F. olivaceus* males have been seen defending a $3m^2$ territory from other males (Harper 1992). During this defense the male will flare its opercles and gular area when a female and another male are present (Baugh 1981). This behavior is not seen in *F. notatus*. It is possible that differing PA2 development could lead to the ability/inability to flare this area. Albertson et al. (2005) state the lever mechanism that is directly correlated with BMP4 expression is reliant on the ligament attaching the lower jaw to the opercle. QTL mapping in cichlids has also shown a correlation between directional selection at the genetic level and phenotypic effects in the LOJ (Albertson et al. 2003).

Other possible explanations include bite speed/force or digging ability for egg depositing. *F. olivaceus and F. notatus* both feed at the same trophic level on the same food stuff so the need for bite speed/force differences is unlikely (Thomerson and

Wooldrige 1970). The two sister species do prefer habitats with different floor surfaces although neither bury their eggs. Since *Hox*A2 is both partially responsible for the musculature and innervation of the LOJ several other phenotypes may also be affected.

In conclusion, we have found evidence of positive selection followed by balancing selection acting at the microevolutionary level in an enhancer that is differentially driving expression between two paralogs at the macroevolutionary level. Coincident with this selection is evidence of compensatory mutations. To our knowledge, this is the first study to research the evolutionary forces acting on a *Hox cis*-regulatory element within and between sister species.

Chapter 5 - Independent Evolutionary Mechanisms of Teleosts Uniqueness

5.1 Introduction

5.1.1 Transposable Elements Role in Genomic Evolution

Transposable elements (TE) were discovered by McClintock (1956) and are endogenous sequences of DNA that have the ability to transpose from place to place within the genome. There are two main classes of transposable elements. Class I transposons (a.ka. retrotransposons) transpose by the copy and paste method. First RNA is transcribed and then reverse transcribed by a reverse transcriptase encoded for by the element itself. Finally, the new DNA sequence is integrated back into the genome by an integrase also coded for by the element itself. Class II transposons (a.k.a. DNA transposons) transpose by the cut and paste method. A transposase protein coded for by the element itself excises the element, or any other element that contains the correct repeat sequences, from the genome and integrates it somewhere else in the genome.

Class II but not class I transposons have been found to be positively correlated with areas of high recombination within the genome (Duret et al. 2000). This association is theorized to be a cause-effect relationship rather than an exploitation of the endogenous double stranded DNA repair process, but not all class II transposons seem to have this ability. In a study done on class II transposable elements in *C. elegans,* Tc1 transposon family members were found in areas of high recombination whereas closely related Tc3 and Tc5 transposons were not (Rizzon et al. 2003).

The introduction of transposable elements near a gene can interrupt or in some cases introduce *cis*-regulatory sequences (McClintock 1956). TEs can also become exons

or introduce alternate splice sites within a gene (Kim et al. 2006; Huh et al. 2009). TEs can also contribute to the overall size of a genome. In some cases TEs compose nearly 50% or more of the entire genome (Lander et al. 2001). Because of the ability to affect genetic structure at both the gene and genome level, TEs are considered a major force in genomic evolution. Because TEs have the ability to reorganize, lengthen intergenic regions, as well as create general havoc to gene sequences it has been theorized that they would be preferentially excluded from well-organized and conserved gene clusters such as the *Hox* gene cluster (Fried et al. 2004).

5.1.2 Tc1 Transposons

As mentioned in the previous section Tc1 transposons have been found to associate with areas of high recombination (Duret et al. 2000). Despite this we found them to be prevalent in ray-finned fishes *Hox* clusters. For this reason we pursued several methods aimed at detecting their possible role in *Hox* cluster evolution that relies on a basic understanding of their mechanism of action.

The Tc1 transposable element was originally found in a species of worm, *Caenorhabditis elegans* (Emmons et al. 1983, X01005.1). This Tc1 was found to be a high copy number transposable element with precise excision from the DNA sequence in which it was residing (Emmons et al. 1983). It was later discovered that Tc1 not only had sites of precise excision but also sites of imprecise excision resulting in the element leaving remnants of its own insertion or taking small sections of DNA from its insertion site (Ruan and Emmons 1987). The Tc1 transposable element also differs not only in its excision precision but also in its excision prowess i.e., the rate at which it excises or inserts. This difference in excision rate is apparent when comparing germ line excision

rate to the rate of excisions in somatic tissue. Excisions occur more than 1,000 times more often in somatic than in germ line tissue (Eide and Anderson 1988). This is important because the number of excisions has a direct effect on the mutagenic ability of the element, also only germ line "jumps" can be seen in the next generation. Earlier in the same decade Eide and Anderson (1985) found that Tc1 transposition was strain specific. This shows that the response to Tc1 differs between cell types as well as between individuals within the same species.

The structure of the Tc1 transposable element and the Tc1 transposase and their interaction were discovered not long after the Tc1 sequence itself was obtained. Tc1 is known to be a 1,610 base pair transposable element with 54 base pair terminal inverted repeats (Liao et al. 1983; Figure 23). There is one continuous open reading frame that codes for a 343 amino acid transposase (Vos et al. 1993). Structurally Tc1 transposase contains a bipartite DNA binding domain. The first domain is from amino acids 1-68 and the second domain is from amino acids 68-142 (Vos and Plasterk 1994). There is also the catalytic domain, characterized by its DDE (Aspartate, Aspartate, Glutamate) motif at positions 157, 247 and 282 respectively (Doak et al. 1994). The bipartite DNA binding domain of the Tc1 transposase binds the Tc1 transposable element by first binding to nucleotides 12-26, with the amino-terminal domain and then binding nucleotides 7-13 with the carboxy-terminal domain, if and only if the amino-terminal domain binds first (Riddle et al. 1997).

Upon excision the Tc1 transposable element inserts preferentially nearby and in some cases to specific sequences (Luo et al. 1998). In the case of Tc1, that integration site has been shown to be a TA/AT palindrome sequence (Rosenzweig et al. 1983b). A

preference has also been shown for the TA palindrome to be flanked by specific nucleotides. The consensus sequence for Tc1 integration is GARA**TA**TGT (the R represents A/G) (Mori et al. 1988).

In order for the transposable element to insert into the target sequence it must bring its 5' and 3' ends into close proximity. This can be accomplished by forming an intermediate loop, as some introns excised from mRNA do. This is known to occur in the Tc1 excision reaction (Rose and Snutch 1984). Proteins from the host cell can also be involved in Tc1 transposition. Since there are differences in the "jumping" abilities of Tc1 transposable elements in germ line and somatic tissue, it is fair to assume that different cell types differentially regulate its transposition. Shortly after Tc1 was discovered, it was shown that Tc1 elements in 5 different loci on different chromosomes were more active in the somatic tissue opposed to the germ line tissue. The study also showed that this was caused by tissue-specific factors that either activated or suppressed the transposition (Emmons et al. 1986). Although there may be factors that help suppress or up-regulate transposition, they are not necessary for transposition. The Tc1 transposase, itself along with the terminal 26bp of the repeat region, are sufficient for activity (Vos et al. 1996).

Although Tc1 was originally found in *C. elegans* and most of the early research went into the mechanisms of transposition, recent effort has been invested in characterizing the phylogenetic extent of this transposable element. Tc1 has been found in fungi (Langin et al. 1995), nematodes (Emmons et al. 1983), arthropods (Jacobson et al. 1986) and vertebrates (Radice et al. 1994). Tc1 also shares its DDE motif with bacterial IS transposases and retroviruses (Fayet et al. 1990), leading to the theory that

Tc1 is a very ancient and conserved transposon. The discovery of the mariner transposable element in *Drosophila mauritiana* led to the realization that Tc1 and mariner constituted a 'family' (Jacobson et al. 1986). Transposable elements belonging to this family subsequently have been discovered in species representing several lineages, several with Tc1 elements that are still active. One lineage shown to lack such a Tc1-like element is the ray-finned fish clade. Almost all remnants of elements found have been shown to have stop codons within their open reading frame. There is evidence, however, for activity in a few species. In zebrafish, active transposition of Tc1-like elements have been shown to occur, although no Tc1-like element with a complete open reading frame has been found to date (Lam et al. 1996). There have also been transcribed Tc1-like transposons found in two groups of salmonid fish, though these Tc-1 like transposable elements have premature stop codons and have not been shown to be translated into functional transposases (Krasnov et al. 2005).

5.2 Transposable Elements in *Hox* Clusters

Transposable elements were originally thought to be excluded from gnathostome *Hox* gene clusters due to their ability to cause genomic rearrangements, alter intergenic distances and ablate or introduce *cis*-regulatory elements (Fried et al. 2004). But, recent work has shown there are indeed TEs within the *Hox* gene clusters (Di poï et al. 2009). We have systematically searched *Hox* gene clusters of the following gnathostomes for transposable elements: horn shark, human, coelacanth (*Latimeria menadoensis*), bichir, zebrafish, medaka, Nile tilapia, spotted-green pufferfish, Western clawed frog, and chicken (Figure 24).

From this analysis we discovered that TEs do reside in and around *Hox* gene clusters of gnathostomes although the type of TEs differs between lineages. Human and shark and for the most part coelacanth *Hox* clusters contain class I transposons whereas the ray-finned fishes including bichir *Hox* clusters contain mostly class II transposons that increase in frequency in the teleost division. This shows a distinction between ray-finned fishes and cartilaginous/lobe-finned fishes. This distinction becomes relevant in terms of how these two types of transposons transpose. Class II transposons “cut and paste” to adjacent regions of DNA with sloppy precision taking with and leaving behind DNA sequences whereas class I transposons “copy and paste” resulting only in insertions. Because of this class I transposons have a greater ability to alter intergenic distances.

One specific class II transposon was of interest due to finding two open reading frames in and immediately flanking the bichir *Hox* cluster. The Tc1/mariner family of DNA transposons is found throughout animal phylogeny from *C. elegans* to Drosophila

to frog (Emmons et al. 1983; Harris et al. 1988; Lam et al. 1996). It is also found among many teleosts fishes (Radice et al. 1994). However a Tc1 with an intact open reading frame has never been found in a vertebrate. We have found two Tc1 transposons in and around the bichir *Hox* clusters. The first is found flanking bichir's *Hox*A cluster just downstream of *Hox*A1. The second is found within the 16,254 base pair intron of *Hox*D3. This is the largest intron found to date for a canonical *Hox* gene. A typical *Hox* gene intron is ~25X smaller (200-700 base pairs). These two Tc1 transposons both contain complete open reading frames that code for amino acid sequences that share homology with known functional Tc1 transposons including the DDE active motif, glycine rich box, nuclear localization signal and paired DNA binding domains (Figure 25). They also contain the inverted repeats flanking the open reading frames that are necessary for self-excision. Although these two Tc1 transposons both appear functional they do differ from each other significantly and do not appear to be the result of a recent transposition from one location to the other (Figure 25).

Tc1 family members are grouped into different sub-families. Within the teleost division Tc1-like transposons are grouped into A, B and C subfamilies with A and B being more closely related (Ivics et al. 1996). Other, more ancient, Tc1-like sequences from a diverse array of organisms (*C. elegans*, frog, fungus, etc.) group outside of the recent teleosts expansions and appear more closely related to one another than to any of the 3 teleost subfamilies (Figure 26). The Tc1-like transposon flanking bichir's *Hox*A cluster (bichir *Hox*A Tzf) groups with subfamily C, which is mainly composed of zebrafish Tc1-like sequences. The Tc1-like transposon found with bichir's *Hox*D3 intron (bichir *Hox*D Tc1) groups with the seemingly more ancient clade. The *Hox*D Tc1 also

differs in the spacing of its catalytic domain. Tc1 family members normally have a DD34E catalytic domain with 34 amino acids between the second aspartate and the glutamate (Shao and Tu 2001). Other variations have also been found for related transposons such as the DD35E motif shared by the bacterial insertion sequence (IS) and human immuno-deficiency virus (HIV, Plasterk et al. 1999). DD37E transposons have also been found (Shao and Tu 2001). The *Hox*D Tc1 has 36 amino acids between the second aspartate and glutamate (Figure 25). The only other transposon with the DD36E configuration is the Tc1-like transposon *Maya*2 found in the African clawed frog (Sinzelle et al. 2005). *Hox*D Tc1 also shares the short inverted repeats with *Maya*2 and the original Tc1 found in *C. elegans*.

Tc1-like sequences are found in and around *Hox* clusters of the teleost fishes as well (Figure 27). They maintain the pattern seen in bichir where subfamily C members flank the *Hox* gene clusters and the seemingly more ancient Tc1 transposons are located within *Hox* gene clusters. In zebrafish, 4 subfamily C members are found within 100 kilobases upstream of the *Hox*Aa cluster and 2 are found within 50 kilobases downstream and all contain inverted repeats. This is significant because Tc1-like transposons normally transpose to an area within 100 kilobases from where they are excised (Fischer et al. 2001; Carlson et al. 2003). Also, intact inverted repeats within 12.5 kilobases of each other are all that are necessary for the Tc1 transposon protein to excise and transpose DNA (Fischer et al. 1999). At first glance this may not seem an issue in zebrafish since no intact Tc1 has been found that is capable of producing a full-length protein. However, as previously mentioned it has been shown that zebrafish has Tc1 transposon activity (Lam et al. 1996). Whether there is another transposon capable of

utilizing the Tc1 inverted repeats is unknown, but it is also possible that the intact Tc1 has yet to be annotated due to the incompleteness of the genome sequence. There is also a subfamily C member flanking the anterior end of the zebrafish *Hox*Da cluster. Within the *Hox* clusters are only members that group with the seemingly more ancient clade. There is one between zebrafish's *Hox*A11a and A9a, *Hox*B13a and B10a, *Hox*B2a and B1a, *Hox*C3a and C1a and medaka's Eve1 and *Hox*D13a and *Hox*D13a and D9a. It appears that at the time the more ancient Tc1-like transposon entered the teleost fishes' genomes the *Hox* gene clusters were more amenable to invasion but during the more recent expansion of the teleosts-specific Tc1-like transposon expansion they have been excluded.

5.3 Testing Tc1 Transposon Activity

Since we found two intact Tc1 transposons in and around bichir's *Hox* clusters we decided to develop both theoretical and functional tests to determine the ability of these sequences to create a functional tranpsosase as well as their ability to alter the structure of the *Hox* clusters in which they reside.

First, we determined if there was a difference in the likelihood that a Tc1 would jump into *Hox* clusters of different species since we had previously determined that there is a difference in the types of transposons that invade ray-finned fishes *Hox* clusters and lobe-finned and cartilaginous fishes *Hox* clusters. In order to perform this analysis we implemented the ProTIS program developed by Liu et al. (2005). In this program 4 types of Tc1-like insertion sites are given weights based on empirical frequencies of insertion. The four categories are 13, 5, 4, and 1 where 13 is 13 times more likely to incur an insertion than 1. This script was run in Perl on the *Hox*A clusters of horn shark, human, mouse, coelacanth, bichir, zebrafish (Aa and Ab) and Japanese pufferfish (*Takifugu rubripes*; Figure 28, 29). From this analysis it is apparent that it is twice as likely that a Tc1-like transposon will jump into an intergenic or intronic region rather than an exonic region. It is also apparent that the species that actually have Tc1-like sequences in an around their *Hox* clusters are twice as likely to have an additional insertion of a Tc1-like transposon within their *Hox* clusters. Bichir and zebrafish have a relative average weight of 0.40 for their non-coding regions whereas mouse and human have a relative average weight of 0.15. This discrepancy disappears outside of the *Hox*A cluster where bichir and zebrafish have a relative average weight of 0.25 and mouse and human 0.22. From this data we can conclude that the reason ray-finned fishes have invasions of class II

transposon where lobed-finned fishes do not is partly due to their respective *Hox* cluster sequence preference. It appears that mouse and human actively maintain an environment that is not conducive to Tc1 transposon insertion (average weight of 0.15 inside cluster to 0.22 outside cluster) and bichir and zebrafish are maintaining an environment that is conducive to Tc1 transposon insertion (average weight of 0.40 inside cluster to 0.25 outside cluster).

Second, we determined what would be the results of a Tc1-like transposon jumping in and out of the bichir *Hox*A cluster. For this analysis a program was written in Perl and JAVA with the help of lab members (Yu-Kang Cheng, Richard David-Rus and Robert Lynch). In this implementation the program takes into consideration the location of the ORF of the transposon, the inverted repeats, *Hox* coding regions and pattern of insertion sites (obtained by previously mentioned script). Variables for the program include rate of fixation in generations, nucleotide mutation rate, population size, number of fixation events and number of runs to execute. For our analysis we started out with the following parameters:

Fixation Rate: 1
Mutation Rate: 0.000624855
Population Size: 500
Fixation Events: 25
Current number of runs used: 1000

In summary the algorithm would jump the transposon and check for sequence differences. If the mutation was in an exon then it was considered inviable and a new mutation was performed. If the mutation hit in a PFC it was considered inviable with an 80% rate of survival. If the mutation hit another area of the *Hox* cluster it was considered 100% viable and continued. In order to multiply the class II Tc1 transposon must jump

in front of the replication fork thus copying itself. We considered this would happen 1 out of every 1000 jumps. In order for the transposase to transpose it must find two inverted repeats within 12.5 kilobases of each other and in the correct orientation (Fischer et al. 1999). Once these are found the transposase will cut this sequence out and jump it either to the right or left with a 50% probability. If the transposition is outside of the reference sequence then it will be considered missing and the run will start over. When a transposition occurs there are 8 types of footprints left behind. Which type was considered with the frequency of the empirical weights given by van Luenen et al. (1994). Based on this program there appears to be no preference of any specific intergenic or intronic region in which the Tc1 transposase integrates (Figure 30). Also, the major changes in *Hox* gene cluster architecture were due to spacing and not sequence alteration.

Third, we developed an *in vitro* system to test the ability of the Tc1 *Hox*A and Tc1 *Hox*D transposons from bichir to mediate transposition. The first step of this experiment was to construct vectors (Figure 31). Two vectors contain the transposase coding sequence of each bichir Tc1 transposon. Four vectors contain reporter genes with inverted repeats from either transposase sequence. As a backbone pFastBac1 vector (GenBank accession number: AY598466.1) was used that contains its own polyhedron promoter, SV40 polyadenylation site as well as an F1 origin of replication, gentamicin-resistance gene and ampicillin-resistance gene. Before each vector was created we first integrated a multiple cloning site (synthesized by IDT) developed to specifically integrate our genes and promoters between the polyhedron promoter and SV40 polyadenylation site. The multiple cloning site also included the short inverted repeats of the Tc1 *Hox*D

transposon. To each reporter gene vector all additional sequences were added between the inverted repeats. We spliced in the EF1α eukaryotic promoter (GenBank accession number: EF203084.1) as well as the EM7 prokaryotic promoter (synthesized by IDT) to vectors 1-4. For vector 1 we added the reporter fusion gene HcRed-Blasticidin (GenBank accession numbers: AY92935.1 and AB364162.1, respectively) and for vector 2 we added Hygromycin-green fluorescent protein (GenBank accession number: AB289768.2) downstream of the two promoter sequences. For vector 3 and 4 we replaced the Tc1 *Hox*D inverted repeats of vectors 1 and 2, respectively, with Tc1 *Hox*A inverted repeats. For the expression vectors we added all sequences outside of the inverted repeats specifically downstream. We added a CAG promoter (GenBank accession number: EF59149.1) to both vectors 5 and 6. To vector 5 we added the Tc1 *Hox*D transposase coding sequence downstream of the CAG promoter. For vector 6 we replaced the Tc1 *Hox*D transposase coding sequence with the Tc1 *Hox*A transposase coding sequence. The total size of each vector was approximately and did not exceed the 7,000 -8,000 base pair range. The vectors are identified as follows: vector 1-pFBEEHcRed, vector 2-pFBEEHygroGFP, vector 3-pFB5EEHcRed3, vector 4-pFB5EEHygroGFP, vector 5-pFBCAGpTc1 and vector 6-pFBCAGpTzf.

Vector design, pFastBac1 vector and reporter gene mother vectors were obtained from Rick Cohen's lab of the W.M. Keck Center for Collaborative Neuroscience at Rutgers, The State University of New Jersey. The multiple cloning site and EM7 promoter were synthesized each as two single stranded DNA molecules by IDT. We annealed the sense and anti-sense strands of these sequences via the following protocol:

1. Dilute DNA oligo crystals in Fermentas 1X PNK Buffer A to a final concentration of

0.001 M. a. Add 20 ul of each strand of the multiple cloning site (50 ul each of EM7 promoter) to a 1.5 mL tube, hand shake and quick spin. b. Add 3 inches of water to a 150 mL beaker and bring to a boil, remove from heat and place on desktop. c. Place 1.5 mL tubes in a float and place in water, let come to room temperature (~45 minutes). 2. Phosphorylate double stranded oligos. a. add the following to a 1.5 mL tube (20 uL dsDNA oligo, 4 uL 10X PNK Buffer A, 0.4 uL 100uM dATP, 13.6 uL pure water, 2.0 uL T4 polykinase (Fermentas). b. incubate in 37 C water bath for 20 minutes. 3. Follow protocol for Enzyme reaction cleanup (Qiagen MinElute Kit 28204). Use spectrophotometer to measure concentration of purified DNA. Run 1% agarose gel to check for proper annealing.

The next step is to amplify the promoters, reporter genes and Tc1 transposons from their respective vectors. Each of these sequences are amplified from their vectors using primers specific for the sequence with additional sequences that can be cut by restriction enzyme to produce overhangs for insertion into vector. They were amplified with the following primers, annealing temperatures and extension times:

Ef1α (58 C, 45”), Ef1-alpha-ClaI-F (5’-ATATATCGATGTTTGCCGCCAGAACACAGGTAAGTGC-3’) and Ef1-alpha-SphI-R (5’-GTATGGCATGCTTTGGCTTTTAGGGGTAGTTTTCACGACAC-3’); HcRed-Blasticidin (58 C, 45”), HcRed-BSD-5-SalI-F (5’-CATACGTCGACCCACCATGGTGAGCGGCCTGCTGAAG-3’) and HcRed-BSD-3-SbfI-R (5’-GTATGCCTGCAGGTTAGCCCTCCCACACATAACCAGAG-3’); Hygromycin-green fluorescent protein (57 C, 1’), HygroGFP-5-SalI-F (5’-CATACGTCGACCCACCATGAAAAAGCCTGAACTCACCGCGACG-3’) and HygroGFP-3-SbfI-R (5’-GTATGCCTGCAGGTTACTTGTACAGCTCGTCCATGCCGAG-3’); CAG promoter (57 C, 1’), CAGpromoter-5-AscI-F (5’-CATACGGCGCGCCCTAGTTATTAATAGTAATCAATTAC-3’) and CAGpromoter-3-NsiI-R (5’-GTATGATGCATTTTGCCAAAATGATGAGACAGCAC-3’); Tc1 *Hox*D (57

C, 1'), Tc1-5-NsiI-F (5'- CATACATGCATCCACCATGCGGCAGTCGTGTGGGCGAAAATGC-3') and Tc1-3-XmaI-R (5'-GTATGCCCGGGCTAACACGGTGTTTGACCCCCTTTCG-3'); Tc1 *Hox*A (57 C, 1'), Tzf-5-NsiI-F (5'- CATACATGCATCCACCATGTCAGAGCACAAACCAAGCATG-3') and Tzf-3-XmaI-R (5'- GTATGCCCGGGTCAATACTTTGTCGATGCACCTTTG-3'); Tc1 *Hox*A 5' inverted repeat (55 C, 30"), 5IVRTzf-5-NotI-F (5'- CATACGCGGCCGCTACAGTGCATCCGGAAAGTATTCA-3') and 5IVRTzf-3-ClaI-R (5'-GTGATATCGATAGCAAAGGCTGTGAATACTTATGTA-3'); Tc1 *Hox*A 3' inverted repeat (55 C, 30"), 3IVRTzf-MluI-F (5'-CATACACGCGTGCAAAGGCTTTGAATACTTATGTA-3') and 3IVRTzf-AscI-R (5'-GTATGGGCGCGCCTACAGTGCATCCGGAAAGTATTC-3').

The PCR reactions were setup with the following volumes: 5.0 uL 10X LA Taq buffer (Takara), 4.0 uL 25mg/mL $MgCl_2$, 4.0 uL 10 mg/mL dNTP, 2.5 uL 10pmol/uL forward primer, 2.5 uL 10pmol/uL reverse primer, 30-60 ng DNA template, 0.5 uL LA Taq (Takara), dilute to 50 uL with pure water. The PCR reactions were ran through PCR in a Eppendorf EP gradient S thermalcycler with the following program: initial denaturation 95 C 2' 30 cycles of (95 C 30", variable annealing temperature for 1', 68 C for variable extension time), final extension 68 C 5'. PCR reactions were purified using Quiagen Gel Extraction Kit 28704 under standard protocol conditions.

The next step is to end digest the amplified sequences. Each sequence is digested with specific restriction endonucleases that will allow them to be ligated to the vector in a particular location. The following enzymes were used to end digest each sequence: Ef1alpha, 10 units ClaI (New England Biolabs, NEB) and 10 units SphI (NEB) with 1X bovine serum albumin and NEB buffer 4; HcRed-Blastocidin, 20 units SalI (Fermentas, F) and 20 units SdaI (F) with Fermentas buffer SdaI; Hygromycin-green fluorescent protein, 20 units SalI (F) and 20 units SdaI (F) with Fermentas buffer SdaI; CAG promoter, 20 units AscI (F) and 20 units NsiI (F) Fermentas buffer R; Tc1 *Hox*D, 20 units

NsiI (F) and 20 units XmaI (NEB) with Fermentas buffer SdaI; Tc1 *Hox*A, 20 units NsiI (F) and 20 units XmaI (NEB) with Fermentas buffer SdaI; Tc1 *Hox*A 5' inverted repeat, 15 units ClaI (NEB) and 10 units NotI (NEB) with NEB buffer 3; Tc1 *Hox*A 3' inverted repeat, 20 units AscI (F) and 20 units (F) MluI with Fermentas buffer R. Each reaction consists of 5.0 ug of DNA (except Tc1 *Hox*A 5' and 3' inverted repeats which uses only 500 ng) along with the restriction enzymes and buffers diluted to 100 uL with pure water. Reactions are placed in a 37 C for 1 hour. Reactions are then purified using Qiagen Enzyme Reaction Cleanup Kit 28204 under standard protocol.

The next step is to ligate the inserts to the vector. First the vector must be digested to receive the proper insert. For any particular insert digest the vector in the same reaction conditions including the same enzymes. Also set up 2 test digestions (10 uL total) one for each enzyme independently. Dephosphorylate the main digestion by the following protocol: 1. Add 5.0 uL 10X Fast AP buffer (Fermentas), 4.0 uL Fast AP (Fermentas), 43 uL sterile water to a 1.5 mL tube and mix. 2. Add mix directly to digestion reaction after digestion and place in 37 C water bath for 10'. Run gel of uncut, single digestions and main digestion on 1% agarose gel to check for proper digestion. Second the insert should be ligated to the vector. The insert to vector ratio should be 3:1. The ligation should be setup under the following conditions: 100 ng vector, variable amount of insert, 2 uL 10X ligase buffer (Fermentas), 1 uL T4 DNA ligase (Fermentas #EL0014) and dilute to 20 uL with sterile water. Place at room temperature for 1 hour. Vector can be transformed into JM109 competent cells ((Promega) using protocol outlined in chapter 4. Colony PCR and sequence verification are performed using protocol outlined in chapter 4.

After the 6 vectors were created they were transformed into human embryonic kidney cell (HEK) 293 cell lines and grown in ampicillin to test for the presence of vector. The vectors have been transformed into the HEK 293 cell lines in the following combinations: 1. negative control-unrelated vector with no expressed genes 2. pFBEEHygroGFP, pFBCAGpTc1 3. pFB5EEHygroGFP3, pFBCAGpTzf. The HcRed-Blastocidin reporter gene has yet to be used because transformation of both transposons simultaneously has yet to be performed.

After initial growth in ampicillin surviving cells were transferred to hygromycin plates and scored for fluorescence. Initial results suggest that the Tc1 *Hox*D transposase was able to mediate transposition of the reporter gene via its inverted repeats into the nuclear genome of the cell.

5.4 Conclusion

Transposable elements are themselves mutations (change in DNA sequence) but they also cause mutation. Because of this and the frequency at which they occur they are considered drivers of genome evolution (review, Biémont and Vieira 2006). The *Hox* gene cluster was thought to be devoid of such elements due to the abilities of these elements to cause large and small-scale genomic rearrangements that would seemingly disrupt the orderly and clustered nature. Shockingly, TEs were discovered in the *Hox* gene clusters of several species of gnathostomes. However, the types of TEs found in *Hox* gene clusters seem to be lineage specific. The ray-finned fishes have undergone an invasion of class II transposons that tend to associate with greater recombination rates as opposed to the lobe-finned and cartilaginous fishes, which have an invasion of class I transposons. One particular class II transposon that is highly associated with areas of high recombination, the Tc1 transposable element, is found in great frequency in and around the ray-finned fishes' *Hox* gene clusters. This invasion seems to have been specific to a particular type of Tc1 element and perhaps to a particular period in geological time. Closer inspection of these Tc1 elements in and around the ray-finned fishes *Hox* gene clusters reveals two intact transposons in the basal living ray-finned fish, bichir. Present knowledge indicates that these are the only known intact Tc1 elements in all of **Vertebrata**. Experiments show that discrimination between types of transposable elements found in *Hox* gene clusters of different species/lineages is the result of presence/absence of insertion sites within those clusters. These patterns are likely

maintained through purifying selection. This implies that invasion of the *Hox* gene clusters coincide with relaxed constraints on purifying selection.

As we have shown previously with the higher rate of turnover of CNSs and the faster rate of mutation in the r3/r5 enhancer of the *Hox*A2 gene, the ray-finned fishes and to a greater extent the derived teleost clade have what can best be described as an 'evolvable' *Hox* gene cluster system. This evolvability is apparent at both the molecular level as well as the phenotypic level and can be seen at both the micro as well as macroevolutionary levels. Such a system was likely to arise in a fashion similar to the following scenario.

For the ray-finned fish lineage to develop an "evolvable" genome based on Lynch (2007) it is likely that there was a bottleneck in population size(s) that allowed mutation and drift to act as the major evolutionary forces, which led to more complex genes. Following this bottleneck a population expansion necessarily occurred leading to an increase in heterozygotes, which is synonymous with an increase in standing variation. Either following or co-occurring with the population expansion was an increase in natural selection as the major evolutionary force due both to strong extrinsic factors (i.e. niche filling, resource competition, etc.) and to the increase in the effectiveness in selection due to the large population size. Strong extrinsic factors that coincide with the diversification of ray-finned fish include large continental plate movements exposing 1,000s of miles of new shoreline, rise of current coral reef species, sea level changes, etc. (Bellwood and Wainwright 2002). Species diversity associated with these extrinsic factors has been studied in pufferfishes and was shown to be significantly correlated with species number (Alfaro et al. 2007).

Chapter 6 – Overall Conclusions

6.1 Questions Addressed

My thesis has addressed the evolution of the intergenic sequences of *Hox* gene clusters among jawed-vertebrates. Specifically it has concentrated on the current state of *Hox* gene cluster intergenic sequences in the ray-finned fishes to obtain a better understanding of how these sequences changed starting at the base of the lineage leading into the duplicated paralogs of the crown group, teleosts. I found that the intergenic sequences of cartilaginous and lobe-finned fishes have maintained a relatively stable set of PFCs as has been reported previously. Chiu et al. (2004) found that the ray-finned fishes, however, have had a considerable turnover of PFCs in the *Hox*A cluster starting at the base of the lineage and this turnover continues well into the teleost division. My work extends this to the *Hox*B and *Hox*D clusters by adding data from the phylogenetically important ray-finned fish, bichir. I found that this rate of turnover appears to be different but steady for each *Hox* gene cluster paralog. This restructuring of *Hox* gene cluster intergenic sequences of the ray-finned fishes appears to be mostly the result of relaxed constraints on the conservation of PFCs due to the loss of several deeply conserved PFCs although positive selection for novel PFCs cannot be ruled out as a few novel ray-finned fish specific PFCs were also found.

Though the macro-evolution of the *Hox* gene clusters in the ray-finned fishes appears to be undergoing mostly relaxed constraints it is unclear what is happening at the micro-evolutionary level to obtain this result. To date and to my knowledge the question of what forces are acting at the micro-evolutionary level within *Hox* gene cluster intergenic sequences has not been addressed. To address this, at least for one specific

CRE, I performed a population level study. Curiously, the strong evidence of relaxed selective constraints seen at the macro-evolutionary level for this CRE appears to be partially the result of strong positive selection at the micro-evolutionary level.

Directional selection on CREs is just one possible force shaping the architecture of the *Hox* gene cluster intergenic sequences. To address what other possible genetic mechanisms have helped shape these non-coding sequences I also cataloged all of the mobile elements and remnants of mobile elements. I found that the conserved *Hox* gene clusters of cartilaginous and lobe-finned fishes contain mostly class I (retro-) transposons, but ray-finned fishes with their high rate of PFC turnover contain mostly class II (DNA) transposons.

6.2 Contribution to the Field

My work has provided the field of *Hox* gene cluster evolution with an important database to begin the cataloging of PFCs and other CNSs. My hope is this database will lead to a large consensus database that can be expanded upon as well as studied in a collaborative effort between evolutionary and developmental biologist. Specifically I have contributed to the knowledge that large turnover of PFCs seen in the ray-finned fishes *Hox*A cluster that appears to be steady from the stem to the crown is not confined only to the *Hox*A cluster and that this turnover proceeds at various rates between the *Hox* gene cluster paralogs. I have also contributed to the recent work done by Di poï et al. (2009, 2010) in lizards that shows mobile elements have invaded the *Hox* gene clusters. My work specifically adds data from the ray-finned fishes, which were not included in these studies. Finally, my population level study was the first such work done at the micro-evolutionary level for any *Hox* gene locus. I hope that my population level research will spur initiatives into similar studies for other regions of the *Hox* gene cluster system and open a door to a deeper understanding of its evolution.

6.3 Questions Raised

The findings of my work raise several intriguing questions. 1. What are the function of the 21 deeply conserved PFCs not associated with any known transcripts? The discovery of these deeply conserved PFCs with no known function desperately calls for research into their possible roles in the *Hox* gene clusters. 2. What are the roles of transposable elements in *Hox* gene cluster evolution? Research into the correlation between transposable element invasions, PFC turnover and morphological variability will help not only further the knowledge of *Hox* gene clusters but genome evolution in general. 3. What knowledge can be gained by studying the micro-evolution of the *Hox* gene clusters? Evidence of strong positive selection at the micro-evolutionary level in an enhancer of a gene that appears to be under relaxed constraints at the macro-evolutionary level brings into question what forces are acting at the micro-evolutionary level in other parts of the *Hox* gene clusters (microRNAs, coding regions, introns, etc…)

6.4 Future Research

Further work is necessary to further realize the impact of the findings of my work. Expression vector studies need to be done to obtain at least a rudimentary knowledge of the function of the deeply conserved PFCs with no known function. While these PFCs may be maintained for reasons other than *cis*-regulation this is the first logical step. Complete sequencing of *Hox* gene clusters of species in phylogenetically important positions (such as hagfish, lamprey, lungfish, bowfin, gold eye, etc…) needs to be completed to obtain a better picture of the evolution of the vertebrate *Hox* gene cluster system. This is exemplified in the recent work in lizards (Di poï et al. 2009, 2010). Collaborative efforts between the fields of mobile element evolution and *Hox* gene cluster evolution could result in the enrichment of both as each provides a unique mechanism for research and could enrich the other.

To follow up on the evidence of strong positive selection between the sister species of killifish there is a need to first establish the effects on expression of these nucleotide differences and then to link this to a phenotypic difference. Functional analyses of the nucleotide differences in the CRE proved difficult. Perhaps this is due to the relatively few differences between the sister species. It is possible that these changes do not alter whether or not a protein will bind but at what strength and length of time it will bind. If the differences are this slight then binding conditions would have to be optimal to maintain pristine protein structure. Freckleton et al. (2009) have designed one type of analysis that may help overcome this problem. In this study genomic DNA is fragmented and ligated into bacteriophages. These bacteriophages will produce amino acid sequences from the genomic DNA inserts and present these molecules on their

capsid. These phages can then be selected for by binding to a specific DNA sequence (*cis*-regulatory element). The genomic DNA sequence can then be amplified and either bound to a micro array or sequenced. This would allow the possible production of proteins that are normally only produced in developmental stages of species such as *F. olivaceus* and *F. notatus* where lab strains have not been established. After establishing expression differences, phenotypic correlation studies should also be performed to see if there is a correlation between genotype and phenotype in these fish. The possibility of a connection between the selection seen at the *Hox*A2b enhancer and the ability to flare the opercle in mate defense are intriguing. Bone morphometrics and behavior studies in these two species could help resolve the results of this selective pressure.

6.5 Concluding Remarks

Less than 10 years ago was an idea that the gnathostome *Hox* gene cluster intergenic sequences, specifically the CNSs, were as conserved as the genes themselves. This was challenged by Chiu et al. (2002) where they showed that in the teleosts *Hox* gene clusters intergenic sequences had lost several PFCs. This was then attributed to the FSGD. This was again challenged by Chiu et al. (2004) where they showed that the actual turnover of PFCs started well before the FSGD in the basal ray-finned fish, bichir. My research has shown that these changes exist across all *Hox* gene cluster paralogs although with varying rates of turnover. It was also the convention that mobile elements would be absent from the *Hox* gene clusters due to their ability to disrupt the order and orientation of the genes in the cluster as well as the spacing between them. My research in ray-finned fishes along with work done by the Duboule lab (Di poï et al. 2009, 2010) in lizards has shown that *Hox* gene clusters do in fact have mobile elements throughout and that the types of mobile elements that have invaded the clusters are lineage specific. Finally, prior to my research, there has never been a population level study of a *Hox* gene cluster locus. My findings, specifically the evidence of strong positive selection at the population level along with the unique evolution of ray-finned fishes *Hox* gene clusters, should focus attention not only on the macro-evolution of this unique system but also on its micro-evolution.

7. Legends

7.1 Table Legends

Table 1-Tests of Selection

A. Interspecies comparisons. Shown are the percent differences found between F. olivaceus, F. notatus and F. catenatus at several loci. *Total differences* is the number of differences seen for each comparison and the total length of that comparison followed by the percent difference. *Coding/in element* is the number of differences seen among nucleotides that either change the amino acid coded for or reside within a known response element. *Non-coding/out element* is the number of differences seen among nucleotides that either do not change the amino acid coded for or reside outside of a known response element. *Ka/Ks-Kb/Ki* is the results of synonymous/non-synonymous or within binding site/outside binding site ratio tests. The highest ratio is 7.78 seen between sister species F. olivaceus and F. notatus for the Hox A2b Enhancer. B. Intraspecies comparisons. Shown are the percent polymorphisms for each category within *F. olivaceus* populations as well as their Kb/Ki ratios for the Hox A2a and Hox A2b enhancers as well as a portion of intron 7 of the PGK1 gene. Also shown are the alleles for each locus and their percent in each population. There are no known binding elements within PGK1 intron 7 therefore no Kb/Ki ratio was done. Note the relatively high number of polymorphism within PGK1 intron 7. C. Hardy-Weinberg Test. Shown is the Hardy-Weinberg test for assortative mating for the Hox A2b enhancer within the Rockford Beach sub-population of *F. olivaceus*. Note that the observed deviation from expected frequencies is 61% probable to happen by chance thus this population is shown to be in Hardy-Weinberg equilibrium.

Table 2-Primers

List of primer names and sequences that were used to amplify regions of interest used in this study. Annealing temperatures are listed to the right. AB/LA stands for annealing temperatures used for Applied Biosystems Taq polymerase and Takara LA Taq polymerase. Phusion refers to annealing temperatures used for Finnzymes Phusion Taq polymerase.

Table 3-Divergence Estimation between *Fundulus sp.*

Divergence date estimations for the three *Fundulus sp.* involved in this study along with percent difference between two pufferfish species with an estimated divergence date of 5-30 my. Personal communication with Michael Alfaro.

Table 4-Exclusive PFC Patterns in *Hox* A and *Hox* B Clusters

Percentage of PFCs shared exclusively between indicated groups. Each comparison was also scaled to the number of pair wise comparisons made for that group of comparisons.

7.2 Figure Legends

Figure 1-Oral and Pharyngeal Jaws

Teleosts fishes have two sets of jaws. The oral or frontal jaws have mostly lost the ability to grind or smash foodstuffs. The pharyngeal jaws evolved secondarily and mostly grind and smash foodstuffs. Liem's Hypothesis states that this partitioning of work allows the two jaws to evolve independently.

Figure 2-Metazoan Tree

The Metazoan tree shown here has families as the terminal branches and an emphasis on the ray-finned fishes (Actinopterygii). Two insets show alternative phylogenies for the Actinopterygii.

Figure 3-Gnathostome Phylogeny and *Hox* Gene Cluster Content

This phylogenetic tree is shown to emphasize the gnathostome species that contain 4 *Hox* gene clusters versus those that contain 8 *Hox* gene clusters as a result of the FSGD. For each species or group a summary of known Hox genes are shown to the right. Paralogous group Hox genes are aligned vertically with the corresponding group number at the top (E = even skipped homologs). For every group/species their Hox clusters are labeled according to accepted nomenclature. A colored pie piece indicates presence of intact Hox gene, a gray pie piece indicates a pseudogene, a white pie piece and/or no pie indicates absence of Hox gene/s. For killifish and striped bass known and verified Hox gene content is listed. See figure 2 for scientific names corresponding to common names.

Figure 4-Bichir Anatomy

The bichir is a unique organism that shares traits with many different orders as well as some unique features.

Figure 5-Bichir *Hox* Gene Trees

Bootstrap or Bayesian posterior probability support is indicated for each branch. Protein trees clockwise from top left, neighbor joining, maximum parsimony and Bayesian. Nucleotide trees from top left, neighbor joining, maximum parsimony, maximum likelihood and Bayesian.

Figure 6-Bichir *Hox* Gene Content

Hox gene cluster size is roughly to scale. *Hox* genes are indicated by boxes with their paralog identity indicated by number. *Hox* genes in blue were previously annotated and those in yellow were annotated during this work. Curved lines on either side of a gene indicate incomplete sequencing of that gene. Unconnected *Hox* genes indicate unsequenced intergenic regions. *Hox* D3 has a 16,255 bp intron indicated by a line connecting the two coding regions.

Figure 7-Bichir *Hox* Gene Trees

Bootstrap or Bayesian posterior probability support is indicated for each branch. Protein trees clockwise from top left, neighbor joining, maximum parsimony and Bayesian. Nucleotide trees from top left, neighbor joining, maximum parsimony, maximum likelihood and Bayesian.

Figure 8-Phylogentic Footprint Clusters Naming Scheme

Proposed nomenclature for naming phylogenetic footprint clusters. The name consists of a 3-letter abbreviation for the species consisting of the first letter of the genus and first two letters of the species. A capital letter representing the Hox cluster, two subscript capital letters represent the position along the cluster, an identification number specific for each PFC, and 1-6 lower case subscript letters representing the section content of that species.

Figure 9-Deeply Conserved PFCs

A map of the locations of deeply conserved phylogenetic footprint clusters on reconstructed gnathostome ancestor HoxA, B, C, and D clusters. Black boxes denote Hox genes; open boxes denote Hox genes that are found in only 1 extant gnathostome to date. Red lines represent PFCs that are found within untranslated regions (UTRs) of Hox

gene mRNAs. Blue lines represent PFCs that are found within 500 nucleotides of a Hox gene proper region but have not been identified as UTRs. Green lines represent PFCs that have high or identical sequence identity to sequences in the EST library. Yellow lines represent PFCs that are not found in any databases or published papers. Black lines are sequences with known function and labeled accordingly. 'Deeply Conserved' refers to PFCs conserved between at least 2 of the following groups, 1) horn shark, 2) $\geq$2 lobe-finned fish, 3) bichir, 4) $\geq$2 teleost fish. The C8 early enhancer does not qualify as deeply conserved but has been found by the PFC method with the 5' end conserved in lobe-finned fish and the 3' end conserved in teleost fish. Information on blast results and PFC number for each PFC represented in this figure is available in supplementary table 1. Numbers in the figure refer to the following references: 1. Larochelle et al. 1999. 2. Nonchev et al. 1996. 3. Berezikov et al. 2005. 4. Shashikant et al. 1995. 5. Morrison et al. 1996. 6. Yekta et al. 2004. 7. Doersken et al. 1996.

Figure 10-PFC Retention Mosaicism

Select species showing the pattern of PFC retention with species containing 4 *Hox* gene clusters (red) or 8 *Hox* gene clusters (blue) for the *Hox* A (left) and *Hox* B (right) gene clusters. As you can see from the pie diagrams human retains mostly PFCs that are shared with other 4 *Hox* gene cluster species whereas bichir has a greater turnover sharing more PFCs with 8 *Hox* gene cluster species. The same pattern can be seen for zebrafish 'a' paralogs indicating further turnover within the teleosts. Also the rate of turnover is greater for the *Hox* B gene cluster than for the *Hox* A gene cluster.

Figure 11-*Hox* A Gene Cluster Retention

A simplified phylogenetic tree of the *Hox* A clusters. Solid circles denote acquisition of a new PFC in that lineage. Open circles denote loss of PFCs in that lineage. Color of open circles denote where the lost PFCs were originally acquired. Numbers indicate the amount of PFCs gained or lost. I only used PFCs that met the criteria of being conserved between at least 3 species. Notice that there is no apparent gain of any PFCs at the base of the teleost but there is evidence of 7 new PFCs gained before the divergence of bichir at the base of the ray-finned fish clade. This puts acquisition of new elements prior to FSGD at a ratio of 7:2. If the two gains in the teleost Aa and Ab clusters are considered to be gained prior to the divergence of the paralogous clusters and then lost in each of the paralogs; the acquisition of new elements in the ray-finned fish clade is in favor of pre-FSGD with a ratio of 9:0.

Figure 12-Intergenic Distances with Mammalian Proxies

Comparison of Hox cluster orthologous intergenic region distances between teleost fish with proxies from mammals. The random bar consists of an average distance for comparisons of 10 random intergenic regions, distances greater than distances between random sequences are considered by chance and therefore are not considered informative. No HoxC cluster is currently available for opossum, *Monodelphis domestica*. Notice the greater distances for teleost fish opposed to the mammalian proxies. This is either due to an overall increased mutation rate and/or incorrect divergence dates for teleost fishes resulting in incorrect proxies.

Figure 13-*Hox* A2 Gene Expression Pattern

In mammals and birds the *Hox* A2 gene is found to be expressed in rhombomeres 2-7 of the developing hindbrain as well as in neural crest cells derived from rhombomere 4. *Hox* A2 has the most posterior expression of any *Hox* gene.

Figure 14-*Hox* A2 Paralogs Expression Patterns

Expression patterns of the *Hox* A2a and *Hox* A2b paralogs in various teleost fishes. The *Hox* A2a paralog maintains a mores ancestral expression pattern in the Acanthomorpha but has been independently pseudogenized in both zebrafish and pufferfish. The *Hox* A2b paralog has a more varied expression with pseudogenization in medaka.

Figure 15-*Hox* A2 Gene Regulation Schematic

Hox A2 gene expression is controlled by 3 *cis*-regulatory elements. The r3/r5 enahncer lies 1-3 kb upstream of the translational start site and controls expression in rhombomeres 3 and 5 as well as neural crest cells derived from rhombomere 4. The r4 enhancer lies within the intron and controls expression in rhombomere 4. The r2 enhancer lies within the 3' portion of exon 2 and controls expression in rhombomere 2.

Figure 16-*In-situ* Hybridization of *Fundulus heteroclitus* Embryos

Hox A2 gene expression in *Fundulus sp.* was discovered using RNA *in-situ* hybridization assays. *Hox* A2a expression is on the left in each developmental stage and and *Hox* A2b expression is on the right. Dorsal views are on top and lateral views are on bottom. An arrow indicates the otic vesicle, which lays overlapping rhombomeres 4 and 5. An arrowhead indicates pharyngeal arch 2. EGR2, a krox-20 ortholog, stains rhombomeres 3 and 5 and is used to show location of rhombomeres in early embryos. M and N are 2-4 somite stage embryos with EGR2 staining in red and *Hox* A2 staining in

blue/purple. A-D are 8-somite stage embryos. E-H are 14-somite stage embryos. I-L are 21-somite stage embryos.

Figure 17-*Fundulus sp. Hox* Gene Trees

Nucleotide trees are shown. Bootstrap or Bayesian posterior probability support is indicated for each branch from top left, neighbor joining, maximum parsimony, maximum likelihood and Bayesian. The branch that supports paralog status for *Fundulus sp.* is circled. *Fundulus sp.* are circled.

Figure 18-Charater State Trees

This character state tree of the *Hox* A2 r3/r5 enhancer from NC2 to RE3 including RE4 and 5' portion of NC3 shows the number of nucleotide changes on each branch and branch lengths are proportional to number of changes. 'a' following common names refers to the *Hox* A2a enhancer and 'b' to the *Hox* A2b enhancer. Total number of changes on each branch is inside squares and number of single-step changes is in diamonds. The *Hox* A2b enhancers show significantly more changes among teleosts and those changes are continuously steady moving from the stem to terminal branches. The *Hox* A2a enhancers share the majority of their changes and have very few terminal branch changes. Killifish *Hox* A2b enhancers have the most number of changes as well as twice as many single-step changes as any teleost *Hox* A2 enhancer. The same pattern is seen for the *Hox* A2 gene region.

Figure 19-Ka/Ks Ratio and Synonymous Percent Difference

This figure shows the Ka/Ks ratios and synonymous percent differences between *F. olivaceus* and *F. notatus* for the *Hox* A2a an *Hox* A2b r3/r5 enhancers and surrounding genes.

Figure 20-RE4/NC2 *Hox* A2b r3/r5 Enhancer Variable Region

All 4 fixed differences and one polymorphism are located within 38 bps of each other in the RE4/NC2 region of the r3/r5 enhancer of the *Hox* A2b gene in the *Fundulus sp.* *F. notatus* has 4 differences that result in slight alteration of RUSH1-α site strength and loss of the Mafb binding site with gain of an HMG-1 binding site. *F. olivaceus* has 4 differences that result in slight alteration of RUSH1-α site strength and loss relocation of an Ubx site with. *F. olivaceus* allele #5 has an additional polymorphism that results in the loss of the RUSH1-α site and does not appear in the population in the homozygous state.

Figure 21-Master TFBS Retention

Shown are the TFBS for krox-20, AP-2 and Mafb. These three transcription factors control 3 major aspects of *Hox* A2 gene expression in the developing hindbrain. Response elements 2-5 are also indicated. Triangles indicate AP-2 binding sites and circles indicate Mafb binding sites. Dark color represents TFBS located in the same orientation as transcription and reddish color represents TFBS located in the opposite orientation of transcription.

Figure 22-*Fundulus sp.* Tree

3 major groups of *Fundulus* are shown along with their possible phylogenetic relationships. Pictures of *F. olivaceus* and *F. notatus* are also shown showing the presence or absence of spots above the lateral line. Approximate divergence dates are given.

Figure 23-Tc1 DNA Transposon Schematic

Schematic of the original Tc1 DNA transposon found in *C. elegans*. D refers to the aspartate amino acid and E to the glutamate amino acid. The Tc1 transposon bind directly to its own inverted repeat sequences as a dimer bringing the two inverted repeats within close proximity.

Figure 24-*Hox* Gene Cluster Transposon Content

The *Hox* gene clusters of 9 species are represented. The top line represents the *Hox* gene cluster of amphioxus which contains the most *Hox* genes of any cluster. The next four lines represent the *Hox* A, B, C and D gene clusters (respectively) of bichir (light blue), horn shark (red), coelacanth (green) and human (pink). The last four lines represent the *Hox* Aa/Ab, Ba/Bb, Ca/Cb and Da/Db gene clusters (respectively) of pufferfish (light blue), Nile tilapia (purple), medaka (green) and zebrafish (fuchsia). The bars indicated the percentage of the intergenic region that is composed of transposons. The bars are color coded to each species and designate 3 types of transposons; Class II transposons (solid bar), Class I LTR transposons (striped bar) and Class I non-LTR transposons (open bar).

Figure 25-Tc1-like Transposons in Bichir

Amino acid alignment of the artificially constructed and functional *sleeping beauty* Tc1-like transposon, the consensus reconstructed ancestral salmonid Tc1-like transposon, the intact Tc1-like transposon found flanking bichir's *Hox* A gene cluster and the intact Tc1-like transposon found within the intron of bichir's *Hox* D3 gene. 4 functional domains are highlighted; two DNA binding domains (Paired-like domain with Leucine Zipper and Homeo-like Domain), nuclear localization signal, glycine-rich box and DDE active motif. The DDE active motif is well conserved between the functional

sleeping beauty transposon and the two bichir Tc1-like transposons although the *Hox* D Tc1-like transposon has a hereto for unseen DD36E motif. Although the DNA binding domains do not appear to be conserved, the two Tc1-like transposons in bichir still are reported as having a Tc1-like DNA binding domain by NCBI protein-blast domain searching algorithm.

Figure 26-Tc1-like Transposon Tree

The Tc1-like transposons found in teleosts are separated into 3 main groups labeled A, B and C. The Tc1-like sequences found during this study are circled in red.

Figure 27-Tc1-like Tranposons in *Hox* Gene Clusters

All Tc1-like transposons found within the *Hox* gene clusters either group or have a greater similarity with Tc1-like sequences that group outside the three main teleosts groups. All Tc1-like transposons found flanking the *Hox* gene clusters group within teleosts group C. Tc1-like transposons that still maintain an identifiable inverted repeat are identified with a star.

Figure 28-*Hox* A Gene Cluster Tc1 Insertion Sites

Shown is the relative amount of Tc1 insertion sites found in intergenic regions of 8 *Hox* A gene clusters. The amount is based on the number and weight of insertion sites.

Figure 29-*Hox* A Gene Cluster Tc1 Insertion Sites supplement

A-D. Shown is the relative amount of Tc1 insertion sites found in intergenic regions of 8 *Hox* A gene clusters. These figures show the relative amount divided into the 4 weight categories. E and F. Shown is the relative amount of Tc1 insertion sites found in intron and exon regions, respectively, of 8 *Hox* A gene clusters.

Figure 30-Simulation Results of Tc1-like Transposition in Bichir *Hox* A Gene Cluster

Shown is the number of base pair alterations after 25 simulated fixations or "jumps" of the Tc1-like transposon found flanking bichir's *Hox* A gene cluster. There appears to be no discernable pattern or preference for location of base pair alteration.

Figure 31-Schematic of Tc1-like Activity Confirmation Vectors

All vectors used a pFastBac1 backbone with an integrated multiple cloning site designed by Rick Cohen. P_{PH} indicates the polyhedron promoter; SV40 polyA indicates location of the poly A signal. 1-6 shows the different multiple cloning site regions used to make the 6 types of vectors used in this study.

8. Tables and Figures

8.1 Tables

Table 1.

A. Interspecies Comparisons

1. PGK1 locus	*Total Differences*	*coding/in element*	*non-coding/out element*	*Ka/Ks-Kb/Ki*	p-value
F.olivaceus/F. notatus	58/1615 (3.59%)	0/92 (0.00%)	58/1523 (3.81%)	0.00	0.06
F. catenatus/F. olivaceus	124/1635 (7.58%)	1/92 (1.09%)	123/1543 (7.97%)	0.14	0.02
F. catenatus/F. notatus	103/1652 (6.23%)	1/92 (1.09%)	102/1560 (6.54%)	0.17	0.04
A3a locus (exon 2)					
F.olivaceus/F. notatus	2/451 (0.44%)	1/300 (0.33%)	1/151 (0.66%)	0.50	0.62
F. catenatus/F. olivaceus	13/451 (2.88%)	4/300 (1.33%)	9/151 (5.96%)	0.22	0.01
F. catenatus/F. notatus	15/451 (3.33%)	5/300 (1.67%)	10/151 (6.62%)	0.25	0.01
T. nigroviridis/T. rubripes	23/436 (5.28%)	5/290 (1.72%)	18/146 (12.3%)	0.14	<0.001
A2a locus (exon 1)					
F.olivaceus/F. notatus	2/333 (0.60%)	1/222 (0.45%)	1/111 (0.90%)	0.50	0.62
F. catenatus/F. olivaceus	3/333 (0.90%)	2/222 (0.90%)	1/111 (0.90%)	1.00	1.00
F. catenatus/F. notatus	5/333 (1.50%)	3/222 (1.35%)	2/111(1.80%)	0.75	0.75
T. nigroviridis/T. rubripes	13/364 (3.57%)	1/243 (0.40%)	12/121 (9.9%)	0.04	<0.001
2. A9b locus (exon1,2 intron)					
F.olivaceus/F. notatus	13/597 (2.18%)	1/190 (0.53%)	12/407 (2.95%)	0.18	0.06
F. catenatus/F. olivaceus	27/597 (4.52%)	1/190 (0.53%)	26/407 (6.39%)	0.08	0.002
F. catenatus/F. notatus	22/597 (3.69%)	0/190 (0.00%)	22/407 (5.40%)	0.00	0.001
T. nigroviridis/T. rubripes	23/285 (8.07%)	12/190 (6.32%)	11/95 (11.6%)	0.55	0.14
			intron 86/243 (35.4%)		
A2b locus (exon1)					
F.olivaceus/F. notatus	2/227 (0.88%)	0/151 (0.00%)	2/76 (2.63%)	0.00	0.05
F. catenatus/F. olivaceus	3/227 (1.32%)	3/151 (1.99%)	0/76 (0.00%)	N/A	0.22
F. catenatus/F. notatus	5/227 (2.20%)	3/151 (1.99%)	2/76 (2.63%)	0.75	0.76
T. nigroviridis/T. rubripes	8/216 (3.70%)	4/144 (2.78%)	4/72 (5.56%)	0.50	0.32
B2a locus (exon1)					
F.olivaceus/F. notatus	1/209 (0.48%)	1/138 (0.72%)	0/71 (0.00%)	N/A	0.47
F. catenatus/F. olivaceus	3/209 (1.44%)	2/138 (1.45%)	1/71 (1.41%)	1.03	0.98
F. catenatus/F. notatus	4/209 (1.91%)	3/138 (2.17%)	1/71 (1.41%)	1.54	0.70
T. nigroviridis/T. rubripes	26/478 (5.44%)	6/319 (1.88%)	20/159 (12.6%)	0.15	<0.001
3. A2a Enhancer					
F.olivaceus/F. notatus	9/538 (1.67%)	1/141 (0.71%)	8/397 (2.02%)	0.40	0.30
F. catenatus/F. olivaceus	17/538 (2.97%)	1/141 (0.71%)	16/397 (4.03%)	0.18	0.06
F. catenatus/F. notatus	16/538 (2.97%)	0/141 (0.00%)	16/397 (4.03%)	0.00	0.02
T. nigroviridis/T. rubripes	49/555 (8.83%)	3/181 (1.66%)	46/374 (12.3%)	0.13	<0.001
4. A2b Enhancer					
F.olivaceus/F. notatus	5/318 (1.57%)	4/108 (3.70%)	1/210 (0.48%)	7.78	0.03
F. catenatus/F. olivaceus	16/318 (5.03%)	6/108 (5.56%)	10/210 (4.76%)	1.17	0.77
F. catenatus/F. notatus	18/318 (5.67%)	6/108 (5.56%)	12/210 (5.71%)	0.97	0.96
T. nigroviridis/T. rubripes	61/420 (14.5%)	17/172 (9.88%)	44/248 (10.3%)	0.96	0.04

B. Intraspecies Comparisons

1. A2a Enhancer

a. F. olivaceus Polymorphism

Population	*Total Polymorphism*	*in elements*	*out elements*	*Kb/Ki*	*p-value*
Little Piney	0/519	0/141	0/397	0.00	
Rockford Beach	1/519	0/141	1/397	0.00	0.56
Gasconade River	0/519	0/141	0/397	0.00	
Total	3/519 (0.6%)	0/141 (0.00%)	3/397 (0.76%)	0.00	0.31

b. F. olivaceus Alleles

Allele	*Little Piney*	*Rockford Beach*	*Gasconade River*
#1	100%	0%	100%
#2 -133C, -177G, -300A	0%	73%	0%
#3 -133C, -177G	0%	27%	0%
Chromosome Count	6	148	42

2. A2b Enhancer

a. F. olivaceus Polymorphism

Population	*Total Polymorphism*	*in elements*	*out elements*	*Kb/Ki*	
Little Piney	0/318	0/108	0/210	N/A	
Rockford Beach	2/318	1/108	1/210	1.94	0.63
Gasconade River	0/318	0/108	0/210	N/A	
Total	2/318 (0.63%)	1/108 (0.92%)	1/210 (0.48%)	1.94	0.63

b. F. olivaceus Alleles

Allele	*Little Piney*	*Rockford Beach*	*Gasconade River*
#1	100%	49%	100%
#5 A218G	0%	13%	0%
#7 T68A	0%	38%	0%
Chromosome count	10	142	28

3 PGK1 intron7

a. **F. olivaceus Polymorphism**

Population	*Total Polymorphism*
Little Piney	6/333
Rockford Beach	8/333
Gasconade River	6/333
Total	8/333 (2.4%)

b. **F. olivaceus Alleles**

Allele	*Little Piney/Gasconade River*	*Rockford Beach*
#1	13%	65%
#2 AA55GT, G62T, C68T, A72C, A87G, CCGA264TCTT	86%	22%
#3 C179T	0%	8%
#4 C235T	0%	4%
Chromosome count	19	56

Table 2.

Primer Name	Primer Sequence	Annealing AB/LA-Phusion
A9b forward	5' CTT TGG AGA CGC ACA CWC C 3'	57 C
A9b reverse	5' TTC TTC ATC TTC ATY CTG CGG 3'	
A2b forward	5' CCT GAC ATC TCT CGC TAA CC 3'	59 C-60 C
A2b reverse	5' AGA GGT CAG ACG CTG CTG C 3'	
A3a forward	5' ART ACA ARA AGG ATC AGA AAG G 3'	57 C
A3a reverse	5' TTR CCC ATT GTG ATT GCT CC 3'	
A2a forward	5' GAA TTC GAR CGA GAG AGC G 3'	56 C-60 C
A2a reverse	5' ACR GGT CCG TTK GAG ATG G 3'	
B2a forward	5' CAT TTC AAA CTT CAT CAA TCA AGG 3'	58 C
B2a reverse	5' CTC YTT CAT CCA KGG RAA CTC 3'	
A3a-A2a int 1 forward	5' AAT GTG AAG CCG TCC GTC C 3'	
A3a-A2a int 1 reverse	5' CAG TAG CTC TCA TAT TTA GGC 3'	
A3a-A2a int 2 forward	5' CCA TGC CAC TAG ATT GCA AGC 3'	
A3a-A2a int 2 reverse	5' ACG TCT ACG TCT CCA AAT CCG 3'	
A3a-A2a int 3 forward	5' AAC AAA TGC TGG ATA TAG ATG GC 3'	
A3a-A2a int 3 reverse	5' CAC GTA ACA GTA ACA TGA AGG C 3'	
A2b enhancer forward	5' AAC ACC CAC TCA CCT CAG C 3'	56 C-60 C
A2b enhancer reverse	5' GTG TGA TCA ATC TTT CHM GCC 3'	
A2a enhancer forward	5' CAA AGA TAA GTG TGC AGA ACG C 3'	56 C-67 C
A2a enhancer reverse	5' TGG CTT GTT GAG TAA GAA GAG G 3'	
PGK1 intron 7 forward	5' GCR AAG GTG AAA GAT AAG ATY CAG C 3'	N/A-62 C
PGK1 intron 7 reverse	5' TCT CSG CTT TGG CCA TCA GG 3'	
PGK1 intron 7 short forward	5' CTG TTT CCA CCA CGA TCT CC 3'	N/A-60 C
PGK1 intron 7 short reverse	5' TCA AAA GAT TGC AAC CAA CG 3'	
A2a exon 2 forward	5' TGA CCG AGA GGC AGG TCA AGG 3'	56 C
A2a exon 2 reverse	5' AGG GCA GAG GGG CTG TCA CC 3'	
EGR2 forward	5' CCA GAC CTT YAC CTA YAT GGG 3'	59 C
EGR2 reverse	5' TGT GTC TCT TYC TCT CRT CGC 3'	

Table 3.

A3a locus (exon 2)	*non-coding/out element*	**5**	**30**
F.olivaceus/F. notatus	1/151 (0.66%)	0.27	1.61
F. catenatus/F. olivaceus	9/151 (5.96%)	2.42	14.5
F. catenatus/F. notatus	10/151 (6.62%)	2.69	16.1
T. nigroviridis/T. rubripes	18/146 (12.3%)	2.46/my	0.41/my
A2a locus (exon 1)			
F.olivaceus/F. notatus	1/111 (0.90%)	0.45	2.73
F. catenatus/F. olivaceus	1/111 (0.90%)	0.45	2.73
F. catenatus/F. notatus	2/111(1.80%)	0.91	5.45
T. nigroviridis/T. rubripes	12/121 (9.9%)	1.98/my	0.33/my
A9b locus (exon1,2 intron)			
F.olivaceus/F. notatus	12/407 (2.95%)	1.27	7.76
F. catenatus/F. olivaceus	26/407 (6.39%)	2.75	16.8
F. catenatus/F. notatus	22/407 (5.40%)	2.33	14.2
T. nigroviridis/T. rubripes	11/95 (11.6%)	2.32/my	0.38/my
A2b locus (exon1)			
F.olivaceus/F. notatus	2/76 (2.63%)	2.37	13.8
F. catenatus/F. olivaceus	0/76 (0.00%)	0	0
F. catenatus/F. notatus	2/76 (2.63%)	2.37	13.8
T. nigroviridis/T. rubripes	4/72 (5.56%)	1.11/my	0.19/my
B2a locus (exon1)			
F.olivaceus/F. notatus	0/71 (0.00%)	0	0
F. catenatus/F. olivaceus	1/71 (1.41%)	0.56	3.36
F. catenatus/F. notatus	1/71 (1.41%)	0.56	3.36
T. nigroviridis/T. rubripes	20/159 (12.6%)	2.52/my	0.42/my
A2a Enhancer			
F.olivaceus/F. notatus	8/397 (2.02%)	0.82	4.93
F. catenatus/F. olivaceus	16/397 (4.03%)	1.64	9.83
F. catenatus/F. notatus	16/397 (4.03%)	1.64	9.83
T. nigroviridis/T. rubripes	46/374 (12.3%)	2.46/my	0.41/my
A2b Enhancer			
F.olivaceus/F. notatus	1/210 (0.48%)	0.23	1.41
F. catenatus/F. olivaceus	10/210 (4.76%)	2.31	14
F. catenatus/F. notatus	11/210 (5.24%)	2.54	15.4
T. nigroviridis/T. rubripes	44/428 (10.3%)	2.06/my	0.34/my
TOTAL			
F.olivaceus/F. notatus	25/1423 (1.76%)	0.79	4.76
F. catenatus/F. olivaceus	63/1423 (4.43%)	2	12
F. catenatus/F. notatus	64/1423 (4.50%)	2.03	12.2
T. nigroviridis/T. rubripes	155/1395 (11.1%)	2.22/my	0.37/my

Table 4.

Human

Raw	4 cluster		teleost	
HoxA	48	**46%**	57	**54%**
HoxB	30	**60%**	20	**40%**

Scaled	4 cluster		teleost	
HoxA	16	**74%**	5.7	**26%**
HoxB	15	**82%**	3.33	**18%**

Bichir

Raw	4 cluster		Teleost		Both	
HoxA	17	**20%**	48	**56%**	21	**24%**
HoxB	3	**8%**	22	**59%**	12	**32%**

Scaled	4 cluster		Teleost		Both	
HoxA	4.3	**40%**	4.8	**45%**	1.6	**15%**
HoxB	1	**17%**	3.67	**61%**	1.3	**22%**

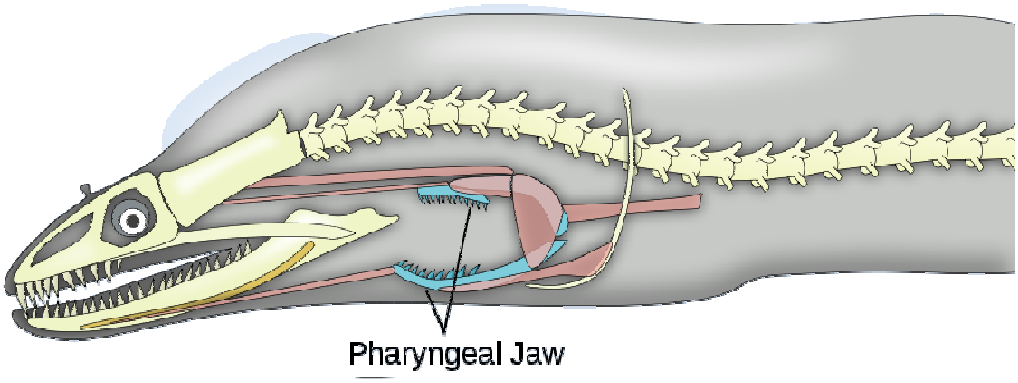
Scaled	4 cluster		Teleost	
HoxA	4	**47%**	4.8	**53%**
HoxB	1	**21%**	3.7	**79%**

Zebrafish

Raw	4 cluster		teleost	
HoxAa	18	**49%**	19	**51%**
HoxAb	12	**80%**	3	**20%**
HoxBa	23	**43%**	30	**57%**
HoxBb	5	**71%**	2	**29%**

Scaled	4 cluster		teleost	
HoxAa	3.6	**49%**	3.8	**51%**
HoxAb	2.4	**71%**	1	**29%**
HoxBa	5.8	**28%**	15	**72%**
HoxBb	1.3	**56%**	1	**44%**

Figure 1.



Credit: *Zina Deretsky, National Science Foundation (after Rita Mehta, UC Davis)*

Figure 2.

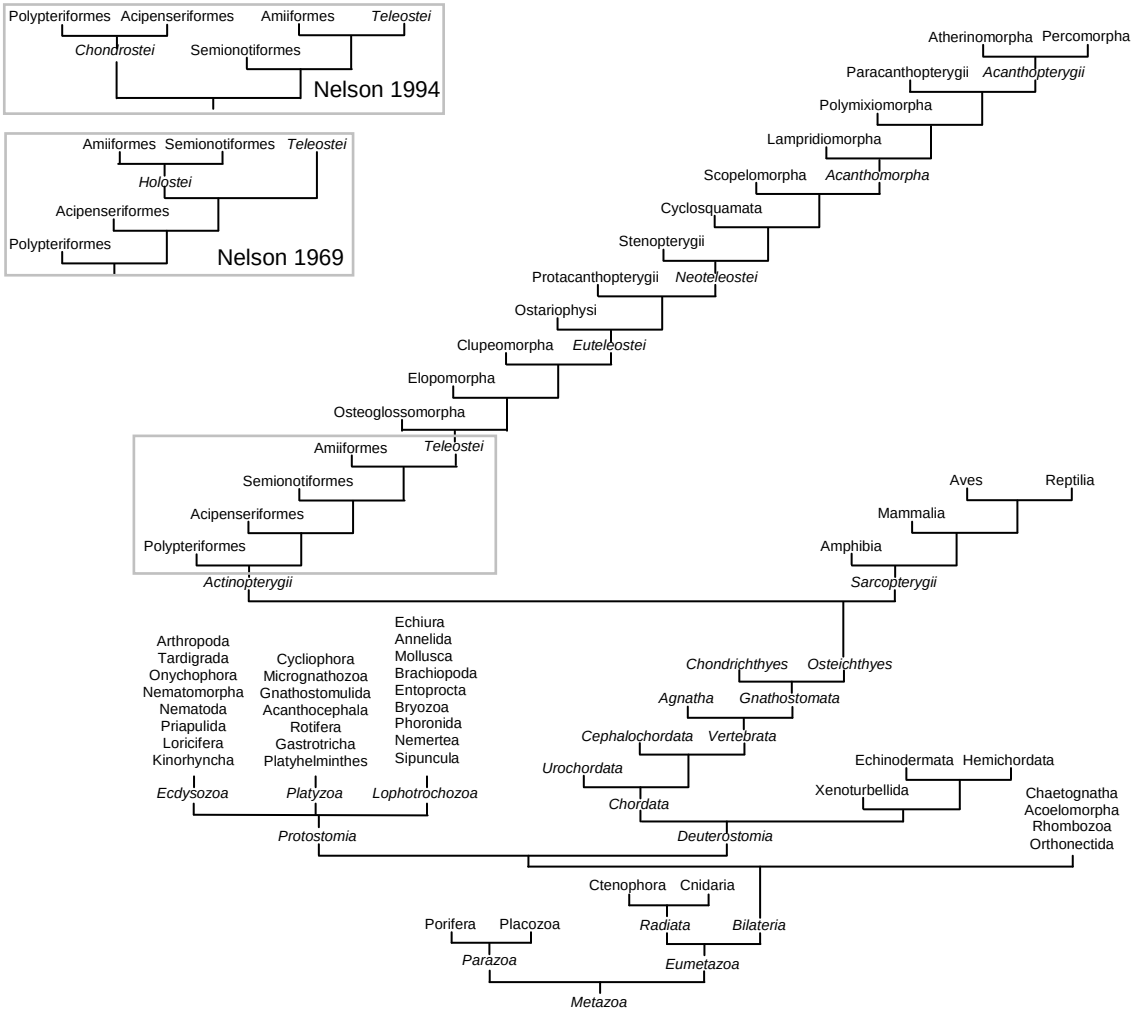


Figure 3.

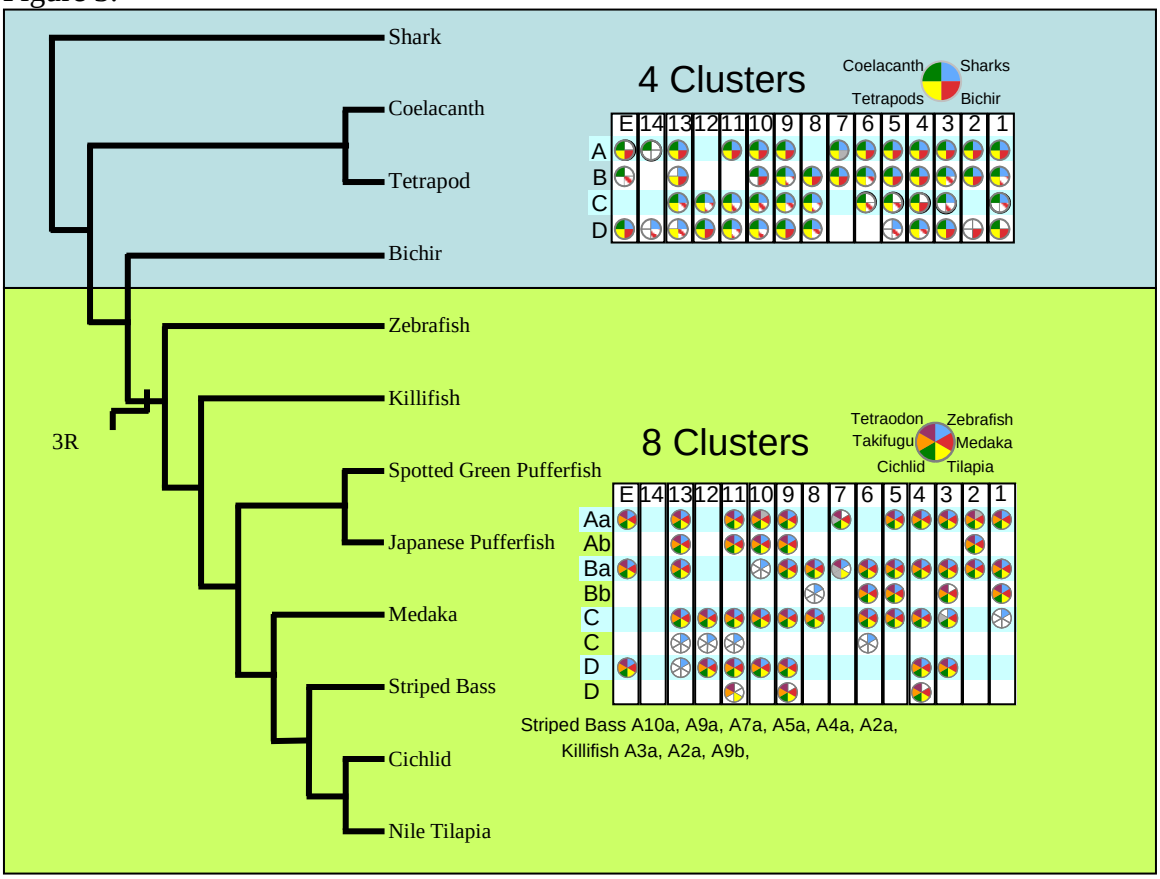
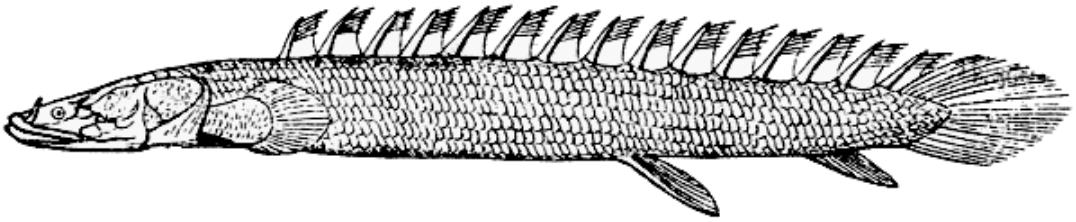


Figure 4.



(Nile Bichir ("Polypterus bichir bichir") from Günther, A.C.L.G., 1880. "An introduction to the study of fishes". Today & Tomorrow's Book Agency, New Delhi. Category:Polypteriformes)

Figure 5.

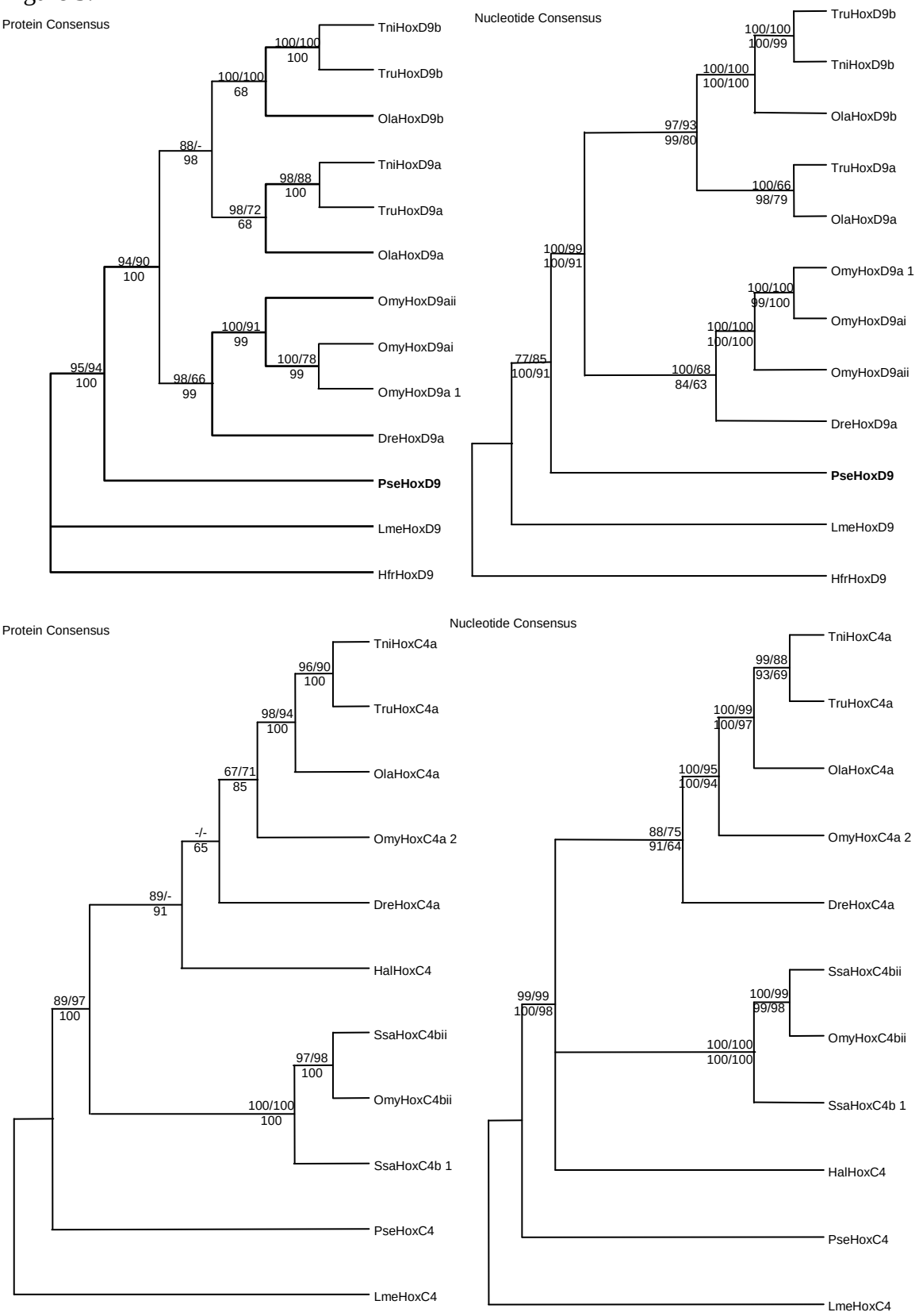

Protein Consensus
TniHoxD9b
TruHoxD9b
OlaHoxD9b
TniHoxD9a
TruHoxD9a
OlaHoxD9a
OmyHoxD9aii
OmyHoxD9ai
OmyHoxD9a 1
DreHoxD9a
PseHoxD9
LmeHoxD9
HfrHoxD9
100/100
100
100/100
68
88/-
98
98/88
100
98/72
68
94/90
100
100/91
99
100/78
99
95/94
100
98/66
99
Nucleotide Consensus
TruHoxD9b
TniHoxD9b
OlaHoxD9b
TruHoxD9a
OlaHoxD9a
OmyHoxD9a 1
OmyHoxD9ai
OmyHoxD9aii
DreHoxD9a
PseHoxD9
LmeHoxD9
HfrHoxD9
100/100
100/99
100/100
100/100
97/93
99/80
100/66
98/79
100/99
100/91
100/100
99/100
100/100
100/100
77/85
100/91
100/68
84/63
Protein Consensus
TniHoxC4a
TruHoxC4a
OlaHoxC4a
OmyHoxC4a 2
DreHoxC4a
HalHoxC4
SsaHoxC4bii
OmyHoxC4bii
SsaHoxC4b 1
PseHoxC4
LmeHoxC4
96/90
100
98/94
100
67/71
85
-/-
65
89/-
91
89/97
100
97/98
100
100/100
100
Nucleotide Consensus
TniHoxC4a
TruHoxC4a
OlaHoxC4a
OmyHoxC4a 2
DreHoxC4a
SsaHoxC4bii
OmyHoxC4bii
SsaHoxC4b 1
HalHoxC4
PseHoxC4
LmeHoxC4
99/88
93/69
100/99
100/97
100/95
100/94
88/75
91/64
99/99
100/98
100/99
99/98
100/100
100/100

Figure 6.

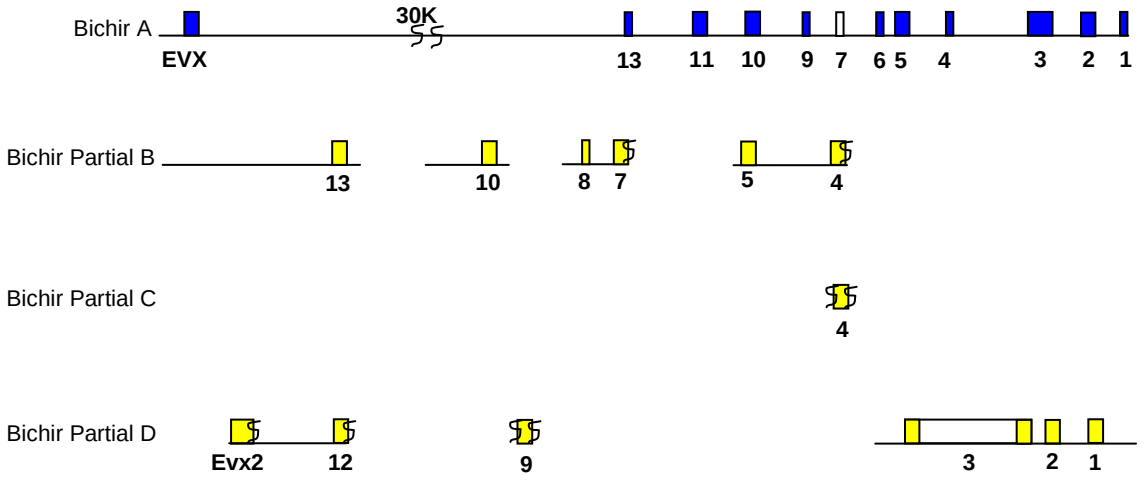


Figure 7.

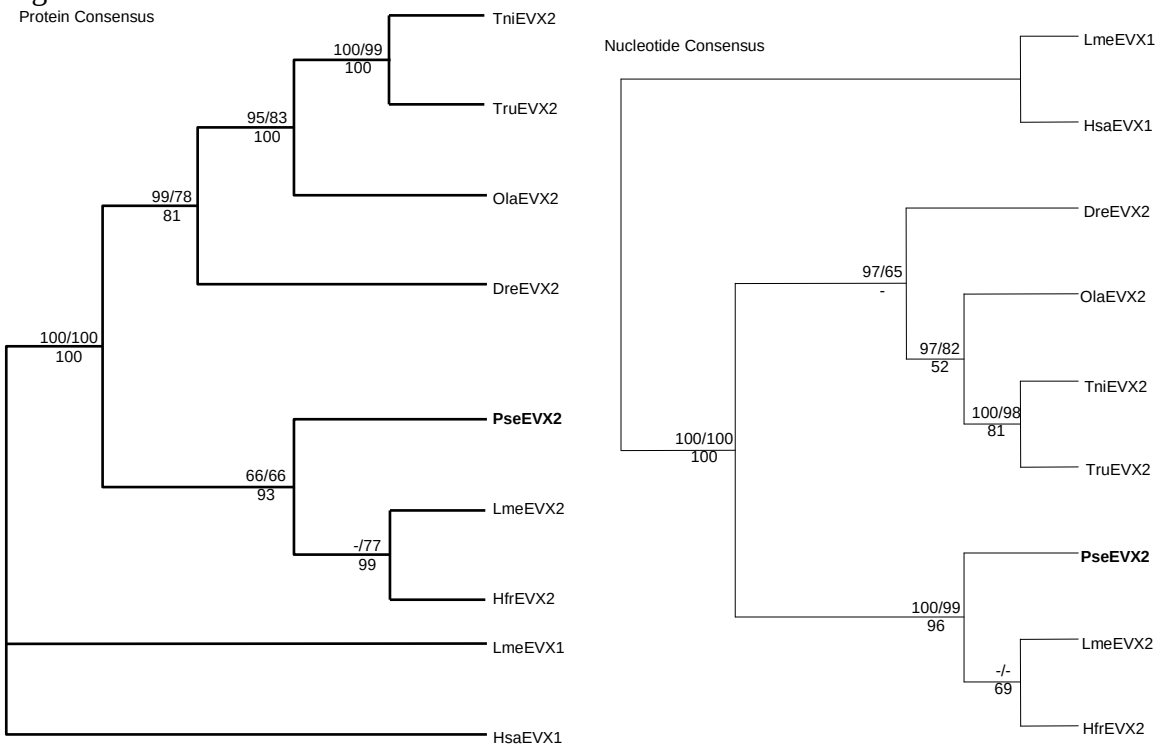


Figure 7. (continued)

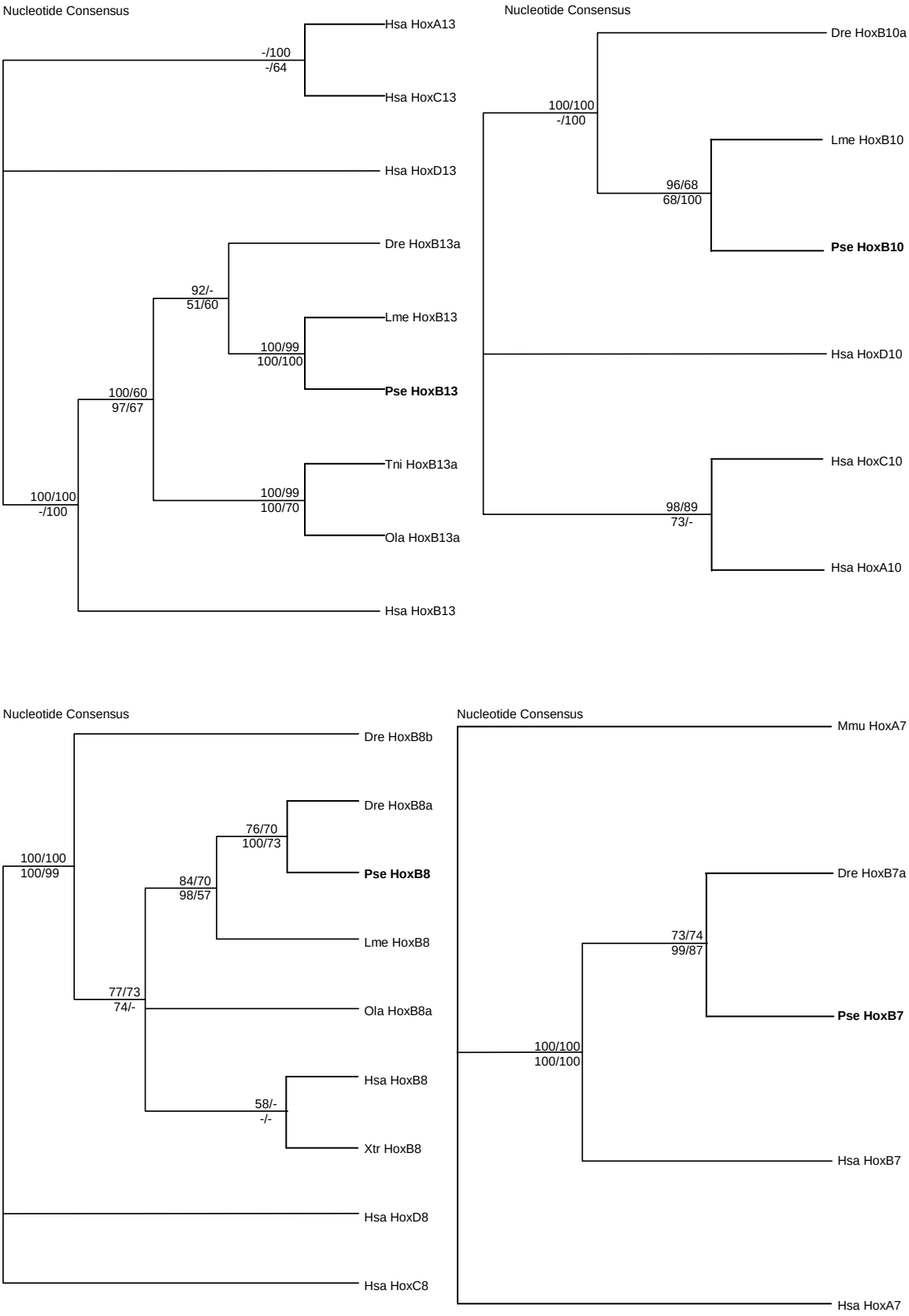
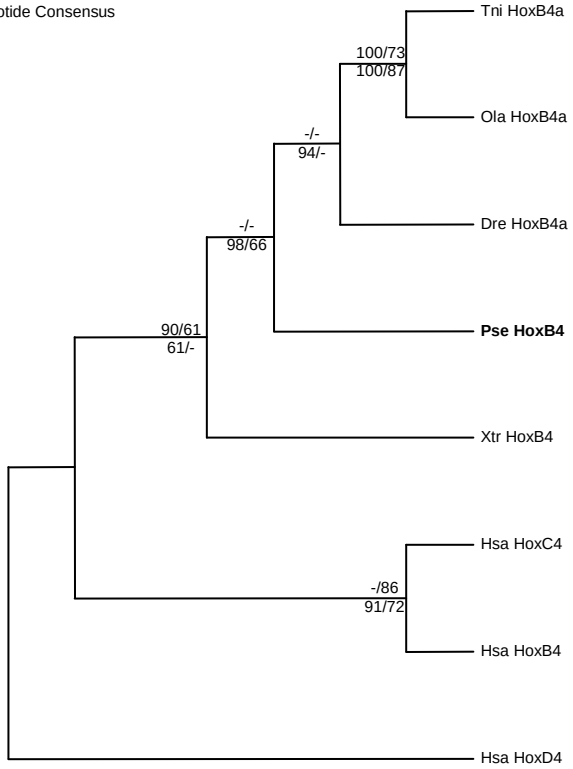
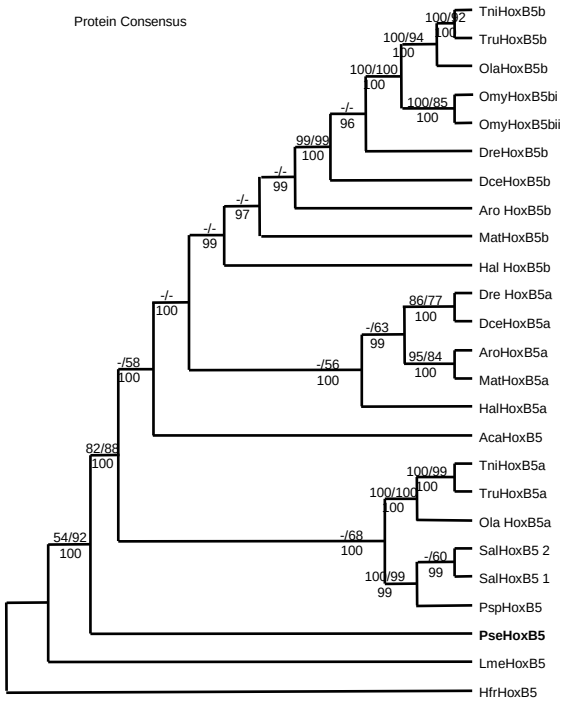


Figure 7. (continued)

Nucleotide Consensus



Protein Consensus



Nucleotide Consensus

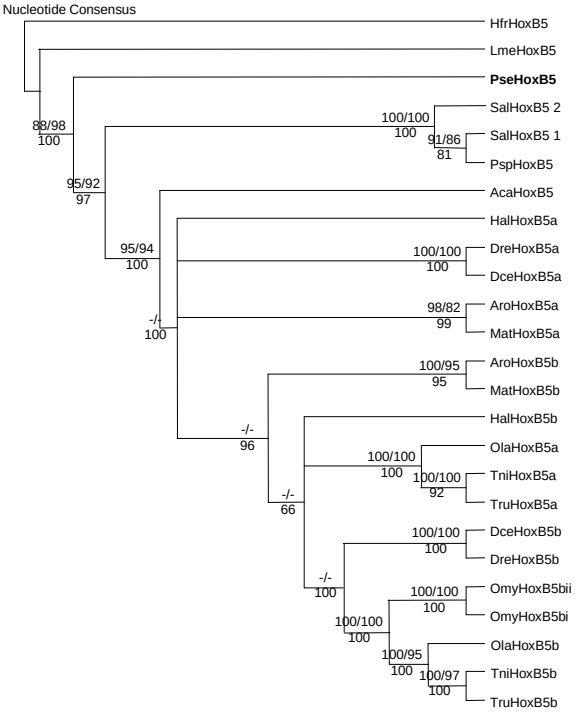


Figure 7. (continued)

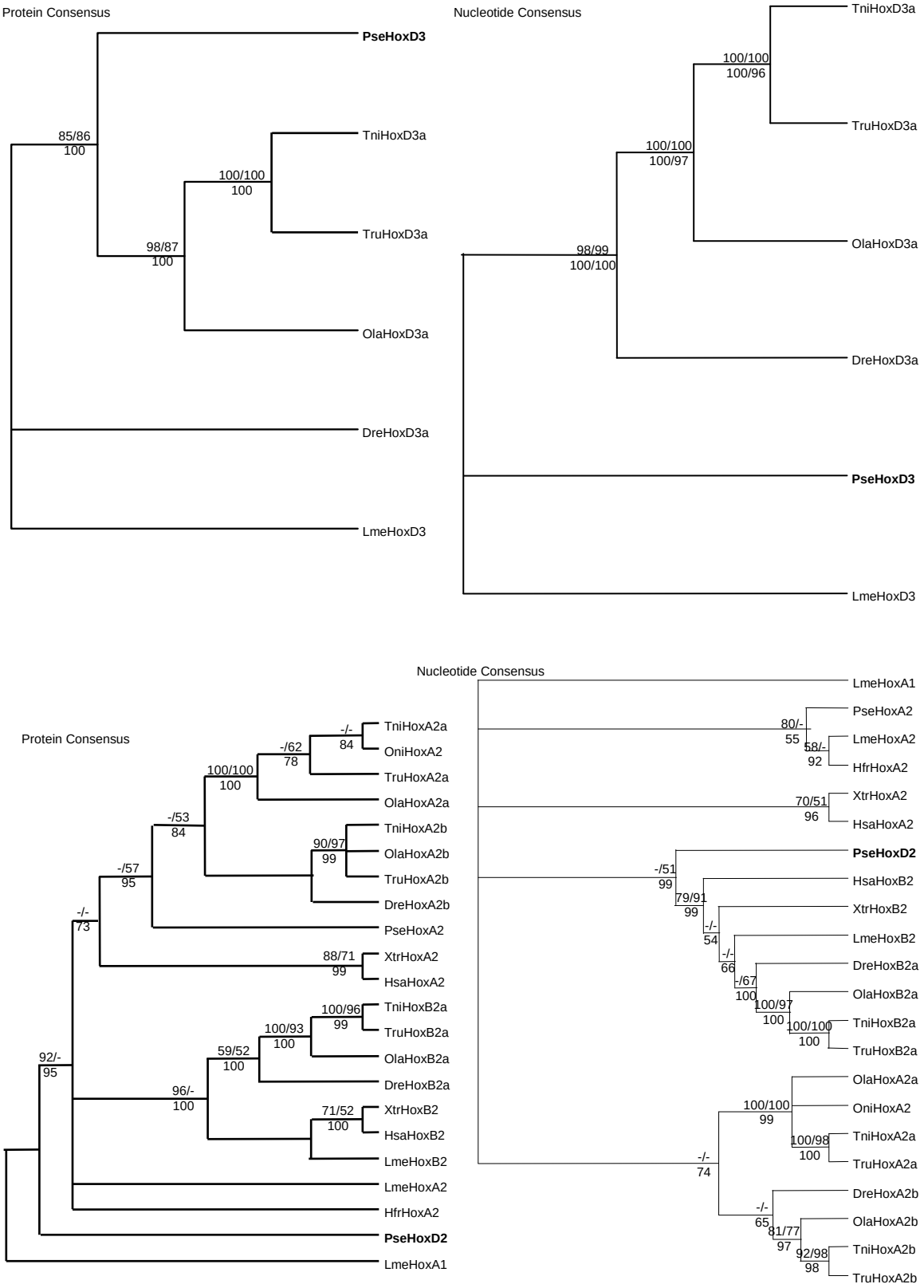


Figure 7. (continued)

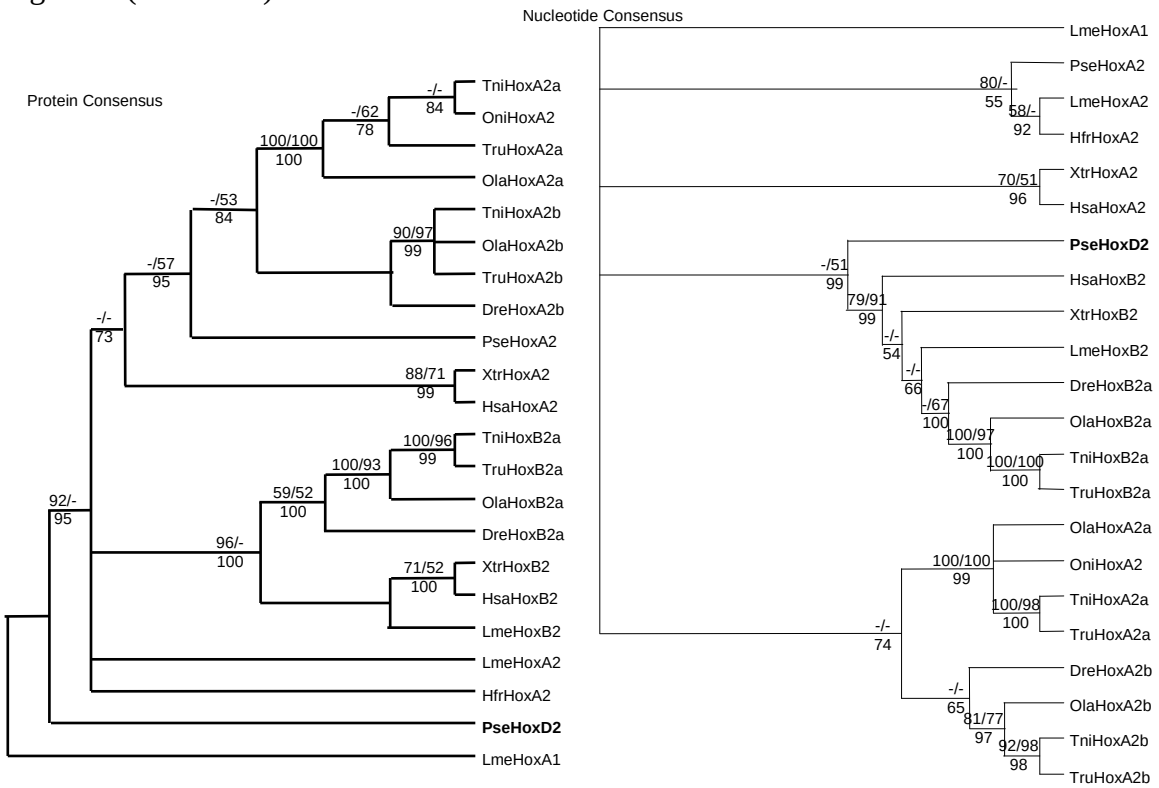


Figure 8.

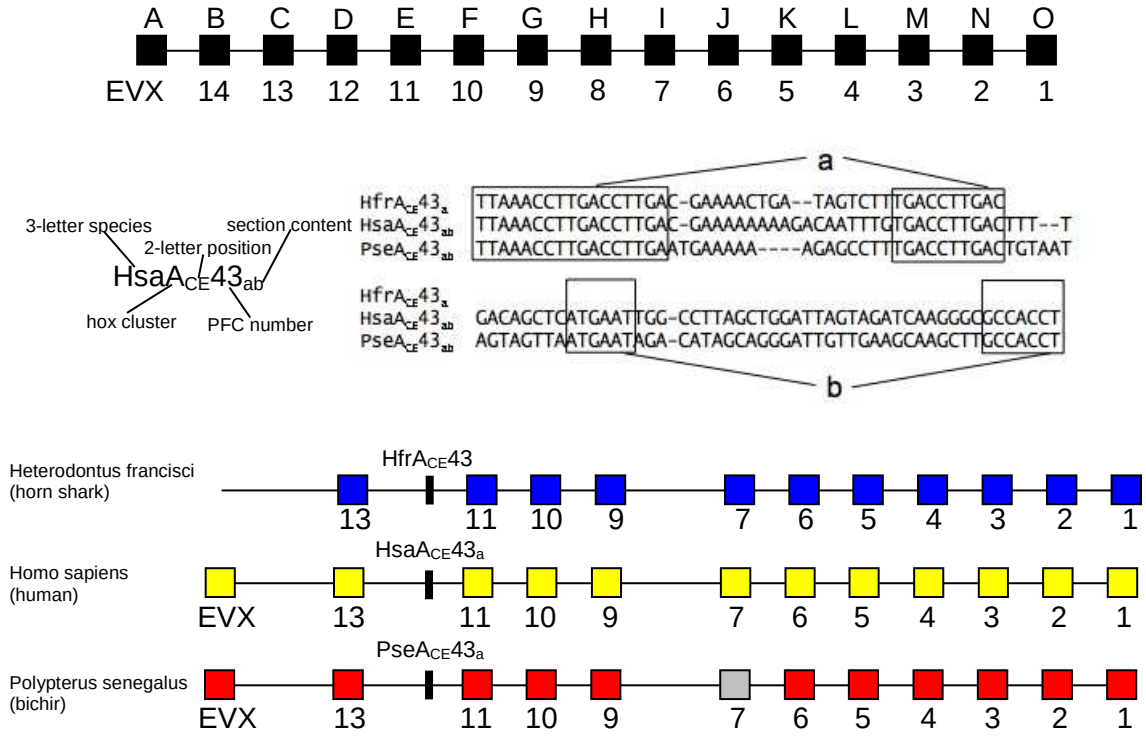


Figure 9.

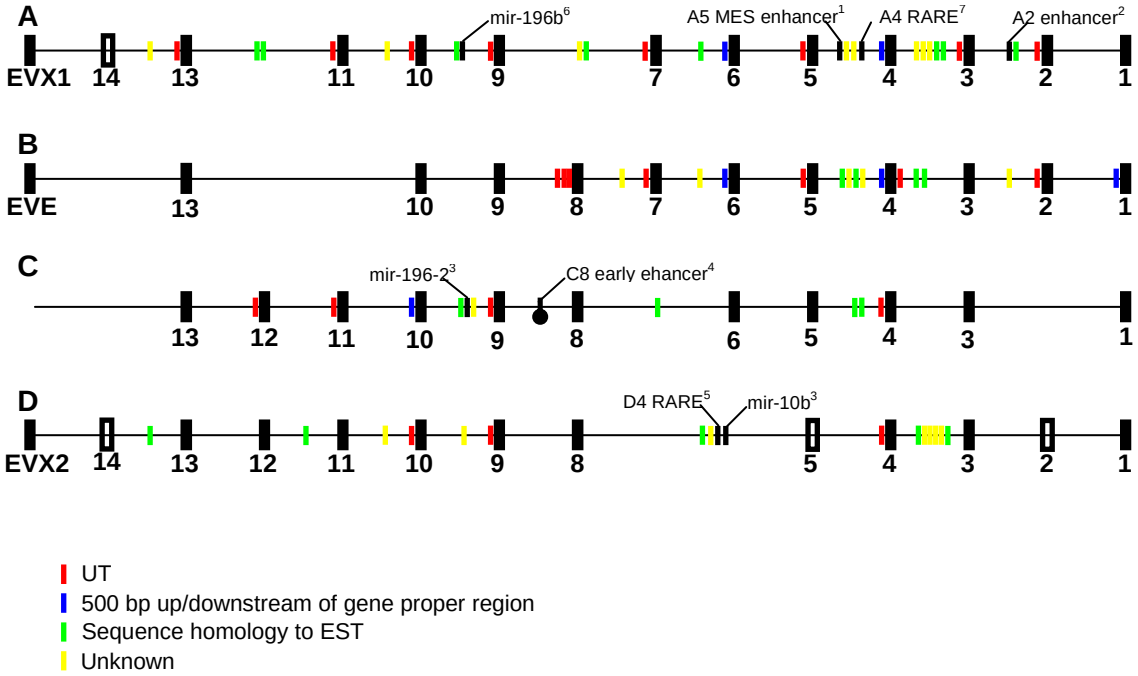


Figure 10.

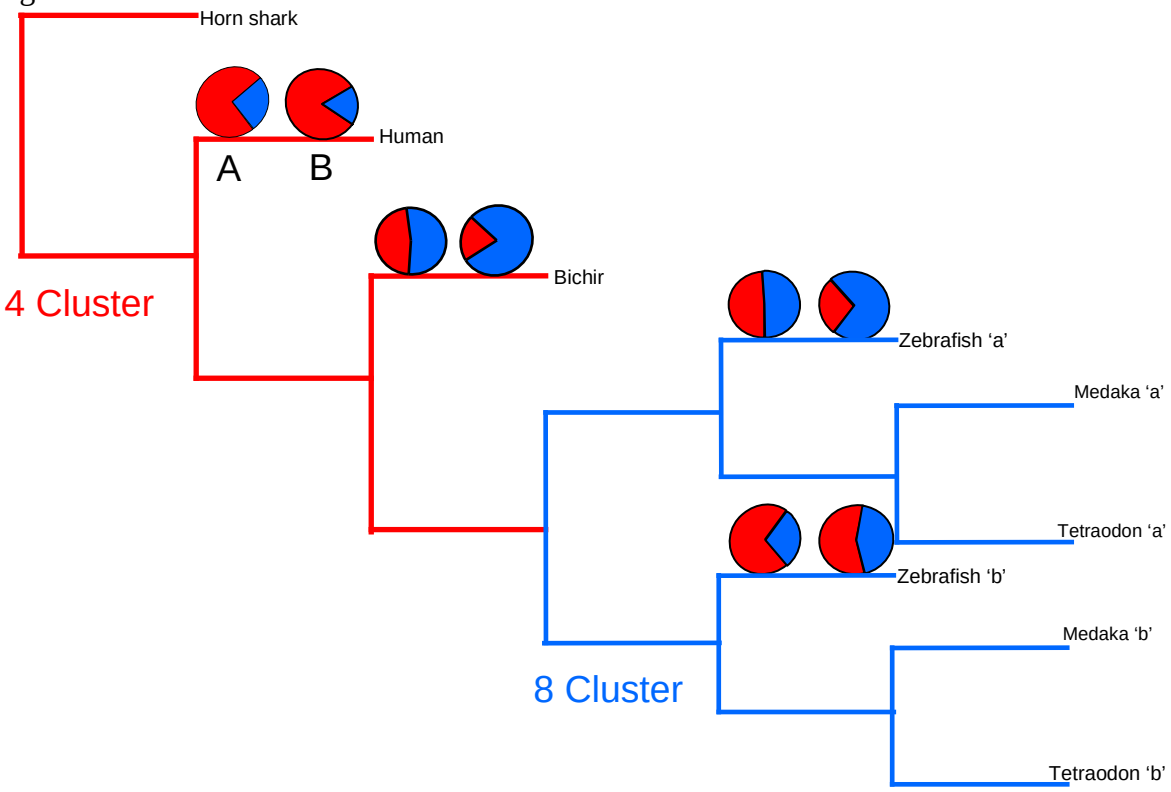


Figure 11.

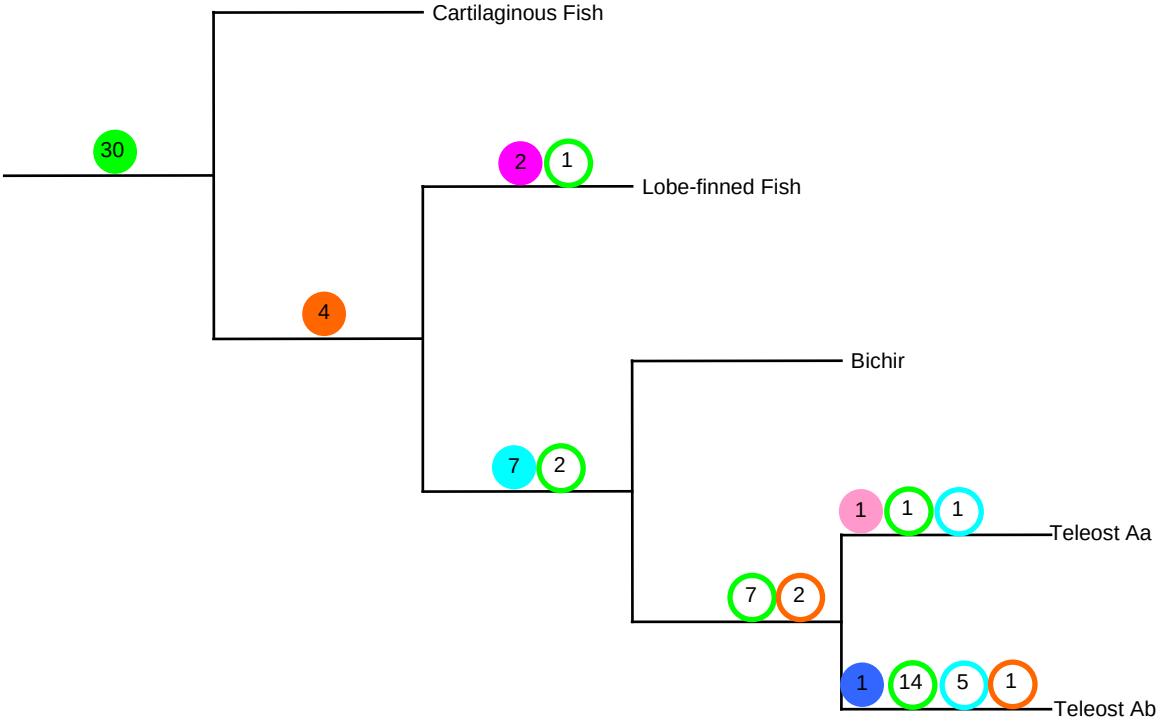


Figure 12.

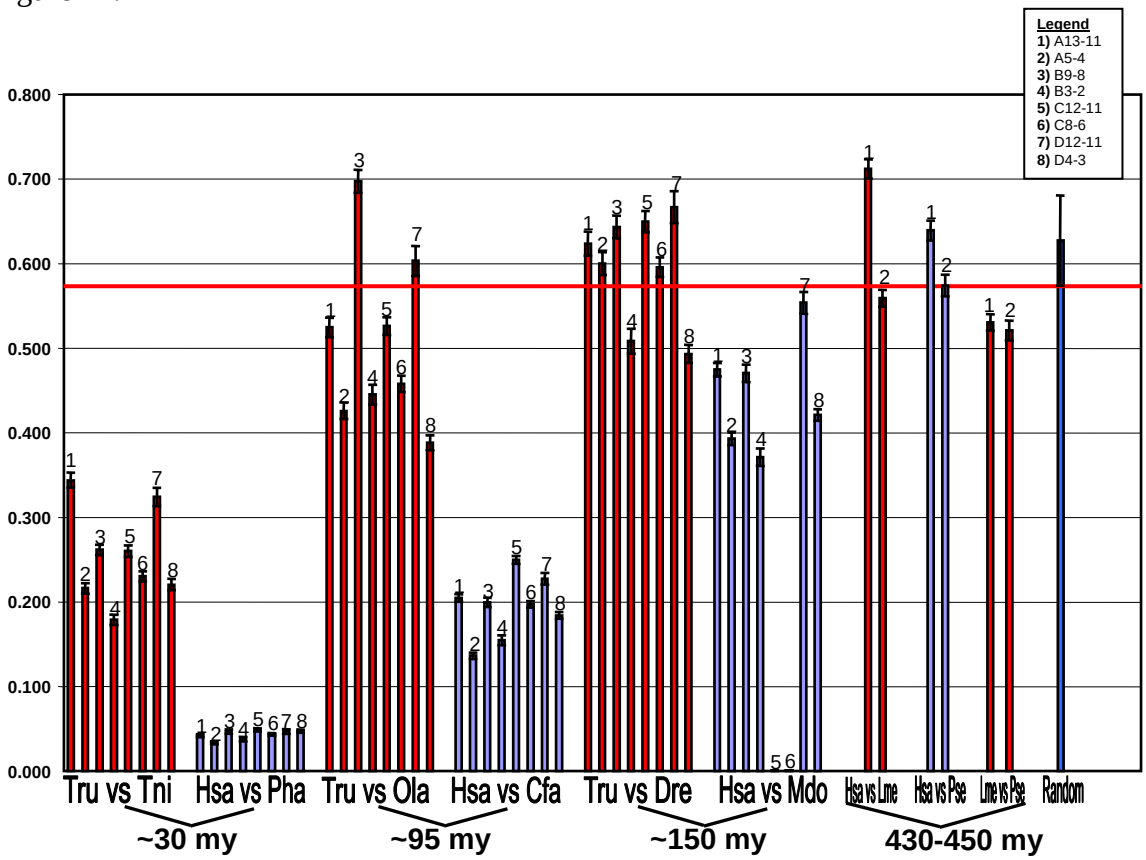


Figure 13.

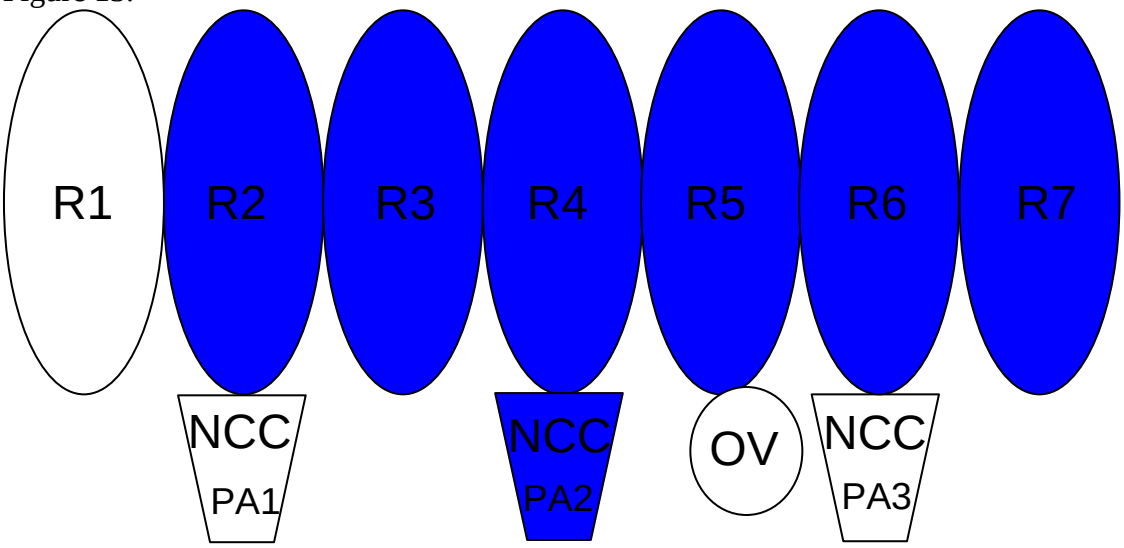


Figure 14.

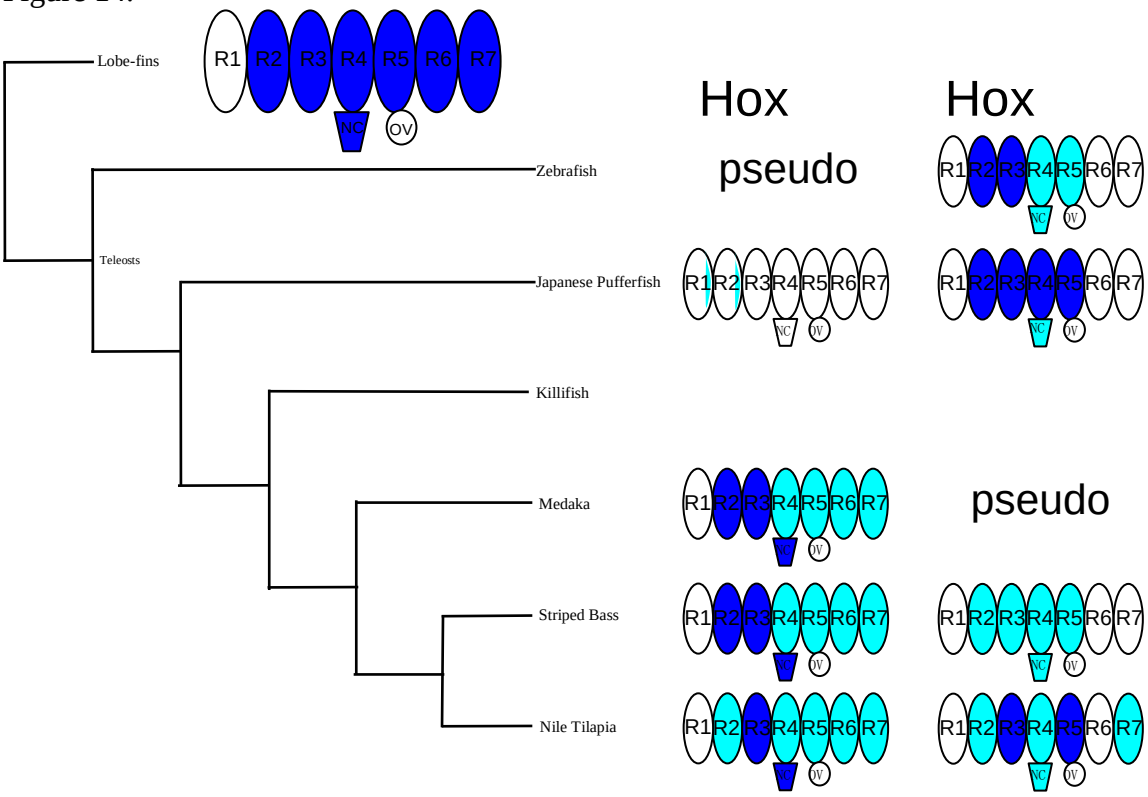


Figure 15.

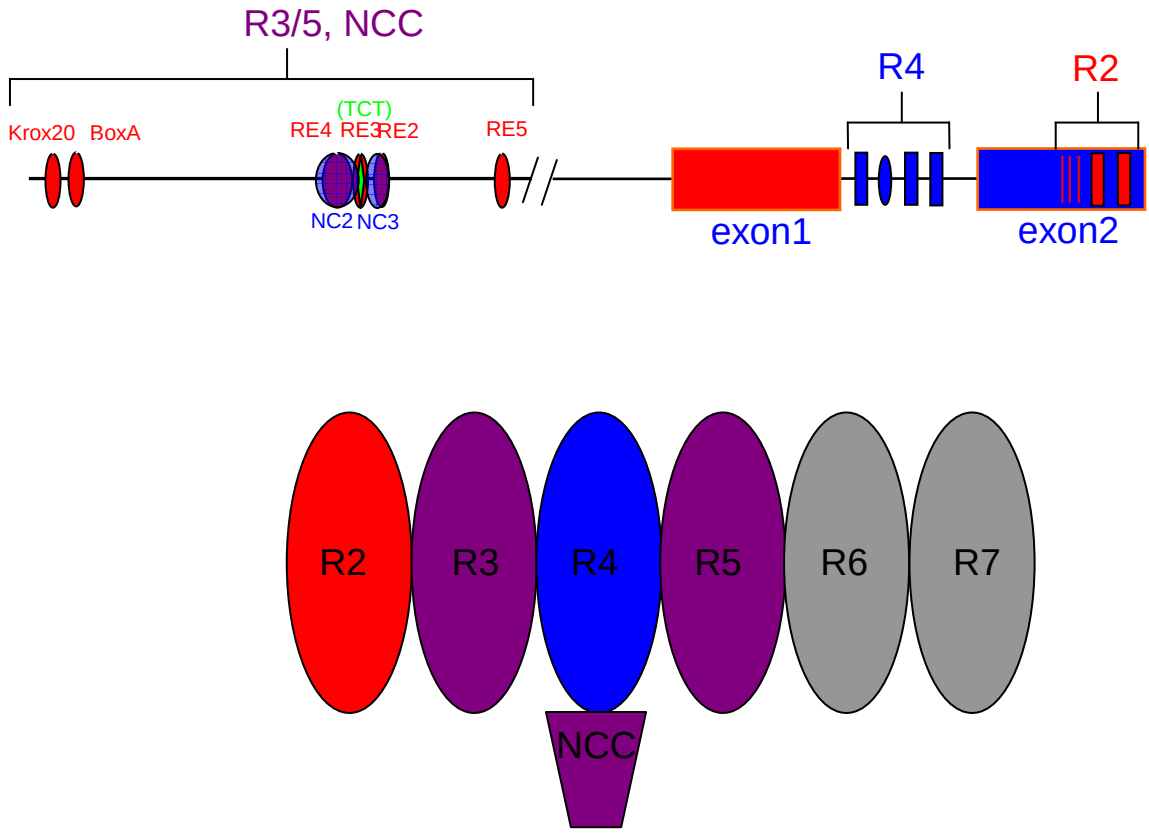

R3/5, NCC
R4
R2
(TCT)
Krox20
BoxA
RE4
RE3
RE2
RE5
NC2
NC3
exon1
exon2
R2
R3
R4
R5
R6
R7
NCC

Figure 16.

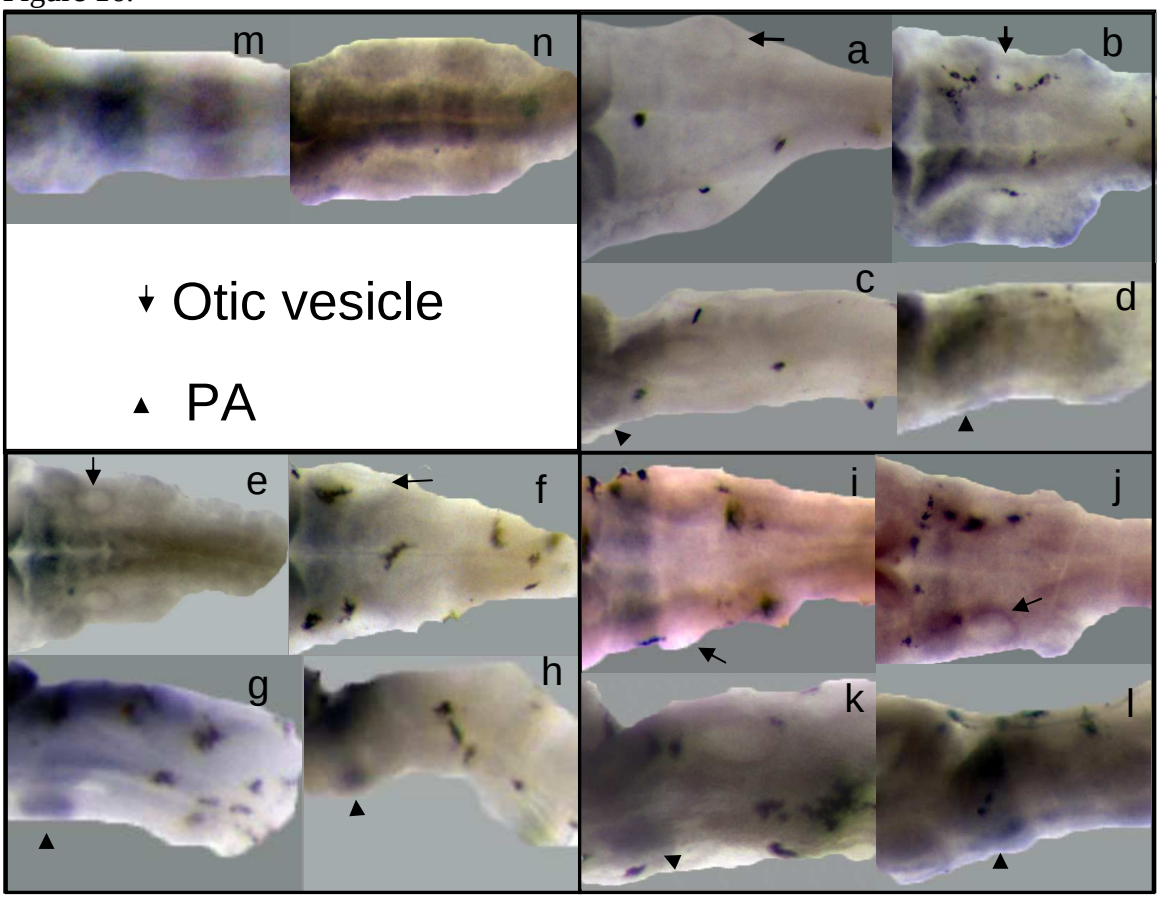


Figure 17.

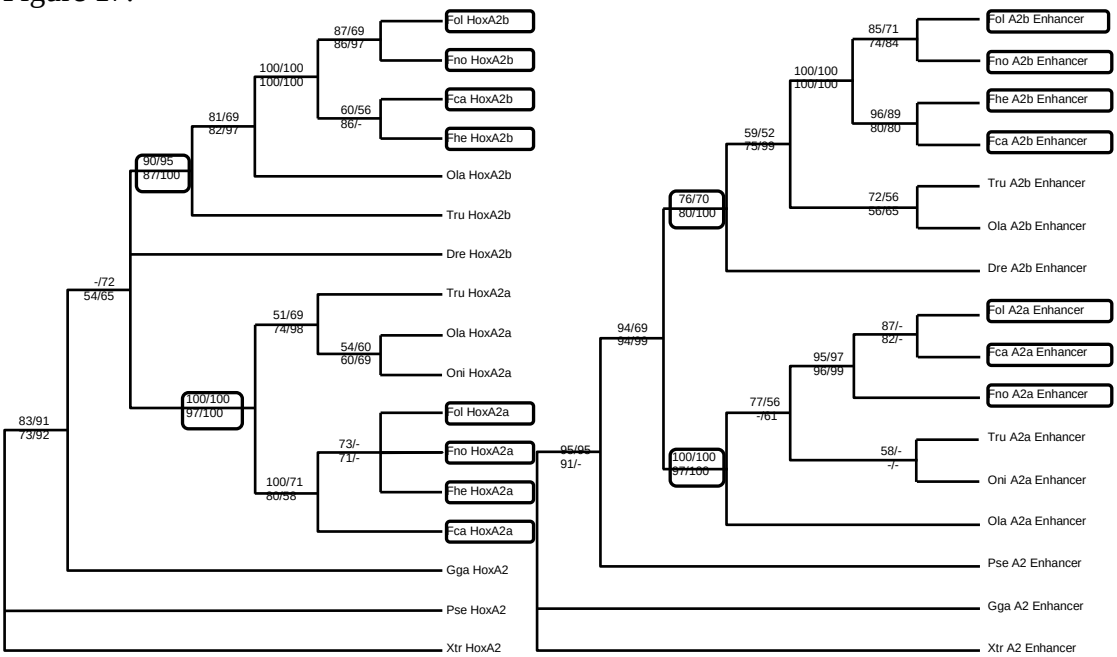


Figure 17. (continued)

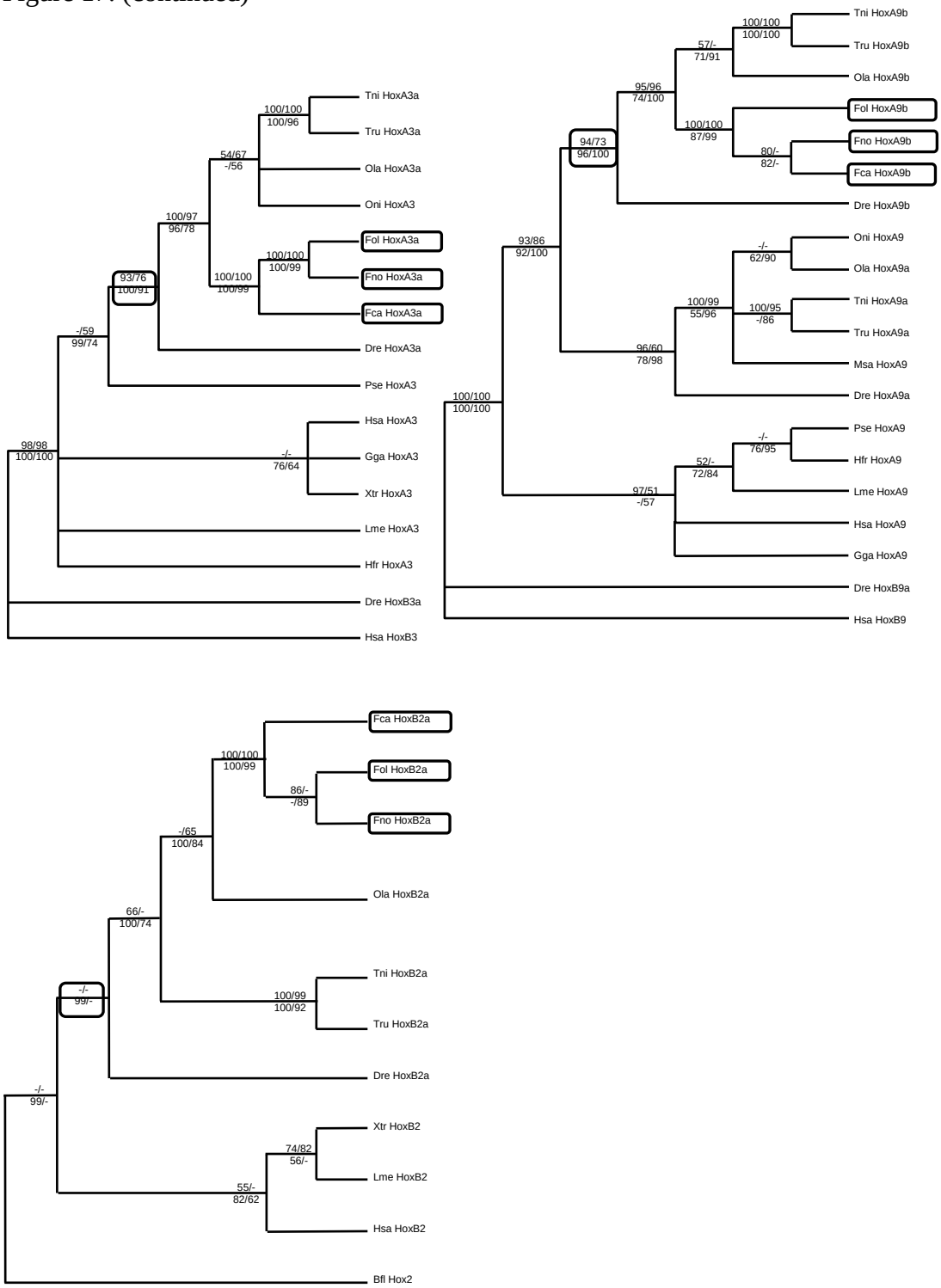


Figure 18.

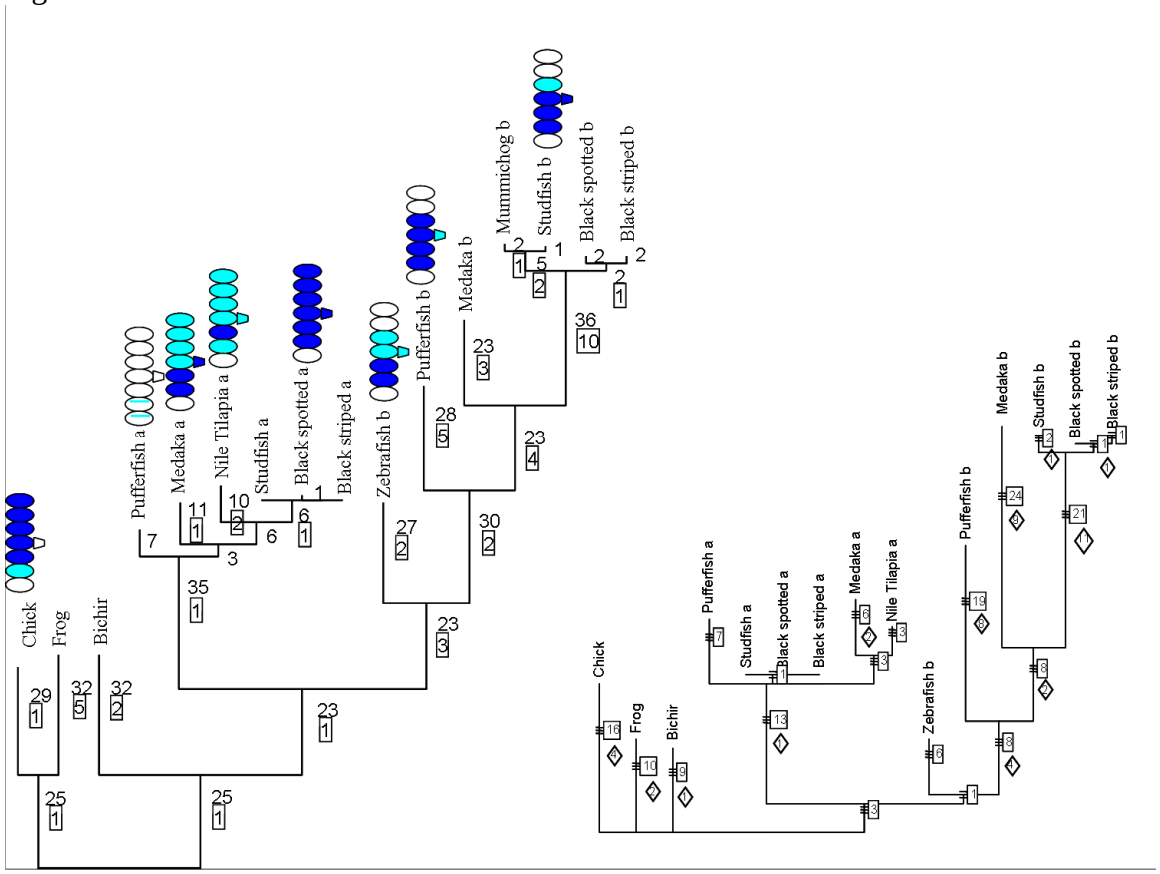
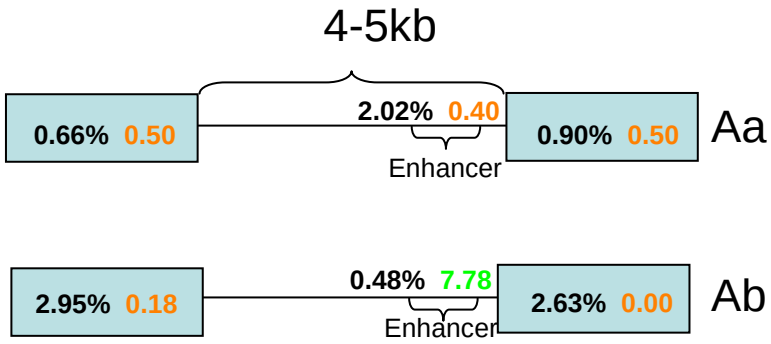


Figure 19.



F. olivaceus
vs.
F. notatus

Figure 20.

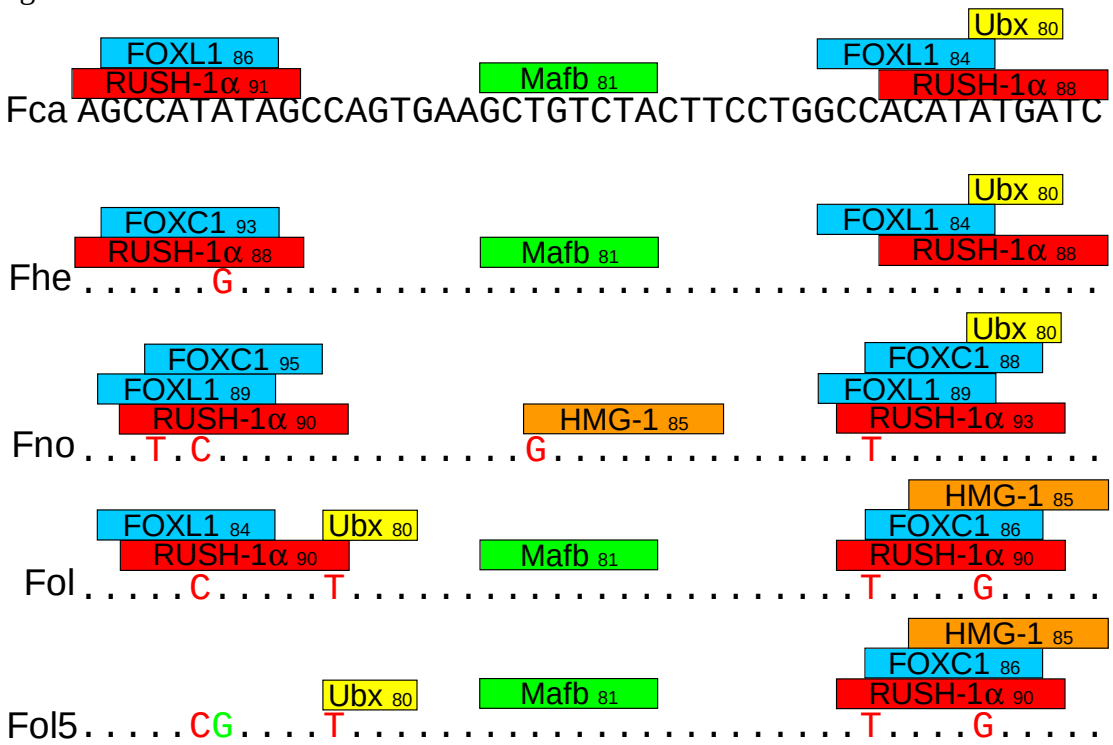


Figure 21.

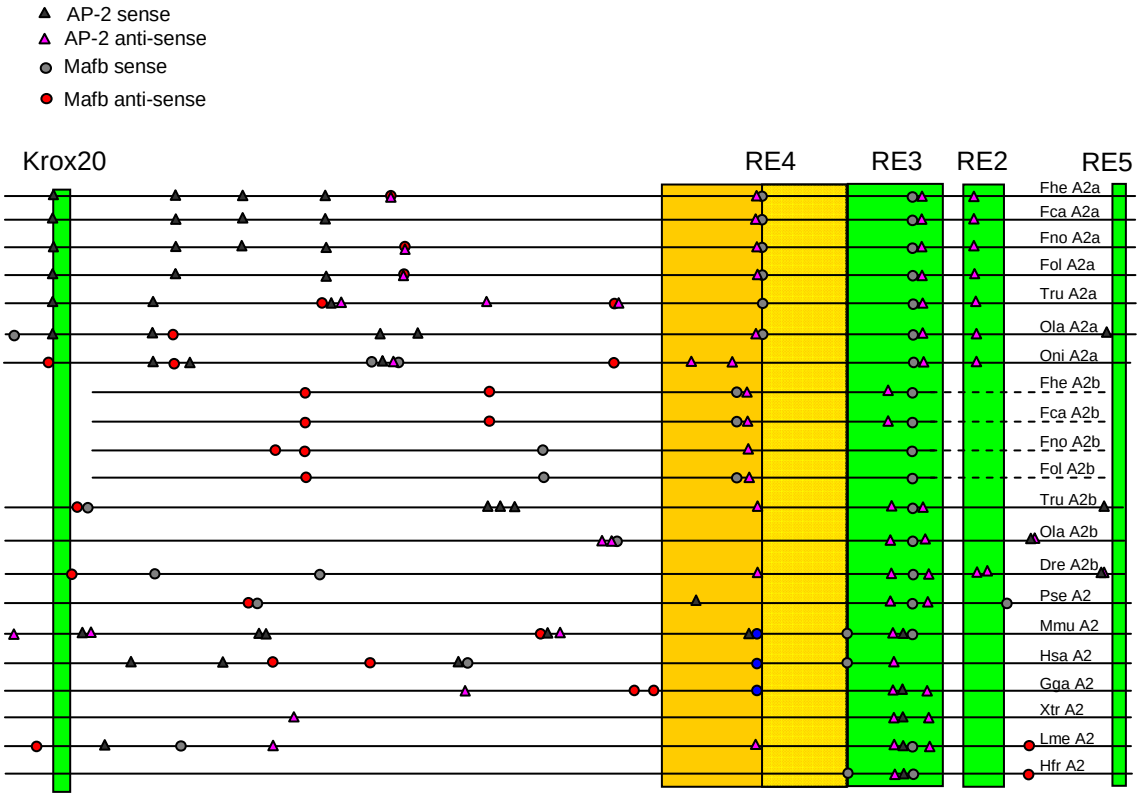


Figure 22.

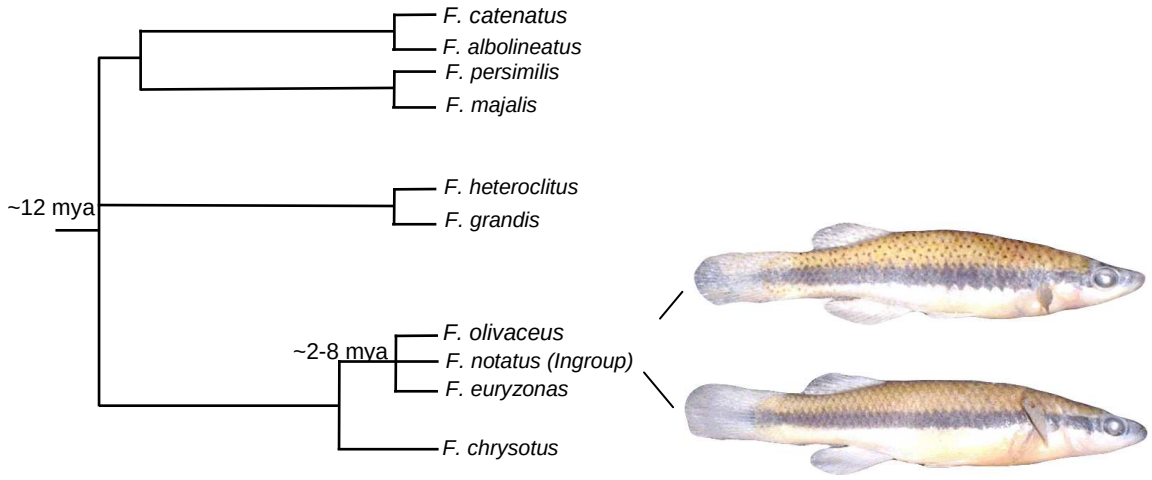


Figure 23.

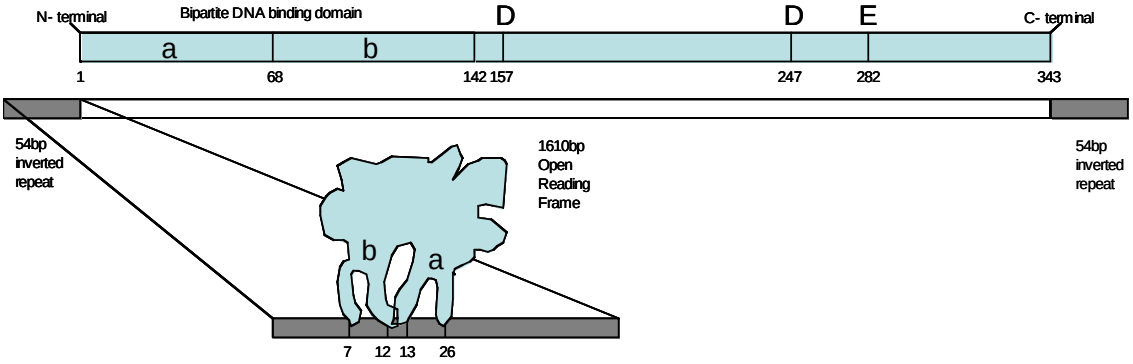


Figure 24.

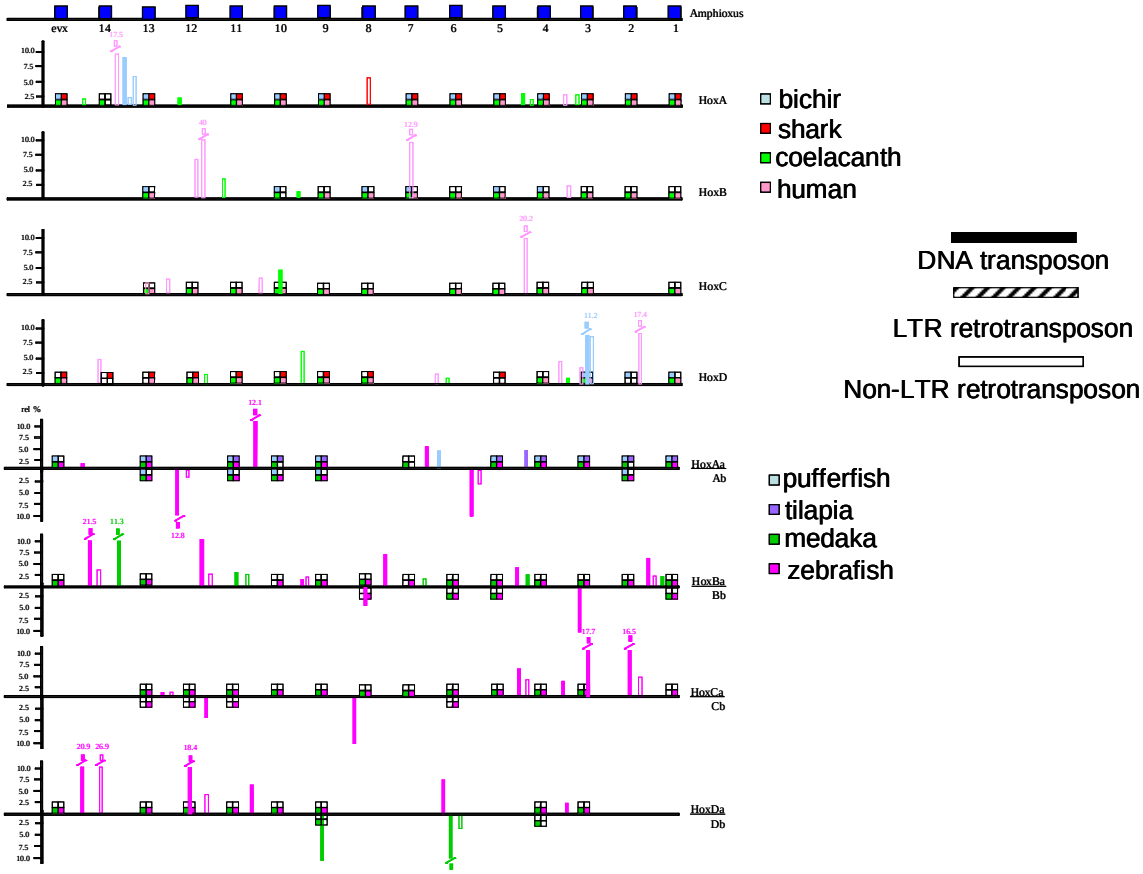


Figure 25.

```
                                                   Paired-like domain with Leucine Zipper                    Homeo-like Domain
                                          helix                helix                helix                                helix
SB10                    1 MGKSKEISQDLRKKIVDLHKSGSSLGAISKRLKVPRSSVQTIVRKYKHHGTTQPSYRSGRRRVLSPRDERTLVRK  75
Salmonid Tc1 ancestor   1 MGKSKEISQDFRKKFVDLHKSGSSLGAISKHLKVPRSSVQTIVRKYKHHGTTQPSYRSGRRRVLSPRDERTLVRK  75
Pse Tzf HoxA            1 M-KSKELSVDLRDRIVSRHKSGE---GYRKALKVPMSTVVSTSRKWKKFETTRTLPRAGRPSKLSDRGRRALVRE  71
Pse Tc1 HoxD            1 MRQSCG-----------------------------------RKCLV-DA-RGQRRMGR---LIQADRRATLTE  33

                                          helix                helix        Nuclear Localization Signal
SB10                   76 VQINPRTTAKDLVKMLEETGTKVSISTVKRVLYRHNLKGRSARKKPLLQ--NRHKK-ARLRFATAHGDKDRTFWR 147
Salmonid Tc1 ancestor  76 VQINPRTTAKDLVKMLEETGAKVSISTVKRVLYRHNLKGRSARKKPLLQ--NRHKN-ARLRFATAHGDKDRTFWR 147
Pse Tzf HoxA           72 VTKNP--IVTELQRSSVESGEPSRRTTISATIHQSGLYGRVARRKPFLS---KRHMAARLEFAKRHLKDFQAIRN 141
Pse Tc1 HoxD           34 IT-------TRYNRGMQQSICEA--TT-RTTLRQMGYNSRRPHRVPLISTTNRKKGLQFAQA-HQNWTEDW--KN  95

                                                                     Glycine-Rich Box
SB10                  148 NVLWSDETKIELFGHNDHRYVWRKKGEACKPKNTIPTVKHGGGSIMLWGCFAAGGTGALHKIDGIMRKENYVDIL 222
Salmonid Tc1 ancestor 148 NVLWSDETKIELFGHNDHRYVWRKKGEAYKPKNTIPTMKHGGGSIMLWGCFAAGGTGALHKIDGIMRKENYVDIL 222
Pse Tzf HoxA          142 KILWTDETKIKLFGVNARCHVWRKPGTAHHQANTISTVKHGGGSIMLSGFFSAAGTGRLVRTKGKMTAAMYRDIL 216
Pse Tc1 HoxD           96 -VVWSDESRFLLRHSNGRVRIWRKQNENMDPSCLVTTVQAGGGGVMVWGMFSRQTLGSIVPIGHRLNATGYLSIV 169

                                                                     DDE Motif
SB10                  223 KQHLKTSVRKLKLGRKWVFQMDNDP-KHTSKVVAK-WL-KDNKVKVLEWPSQSPDLNPIENLWAELKKRVRARR- 293
Salmonid Tc1 ancestor 223 KQHLKTSVRKLKLGRKWVFQMDNDP-KHTSKVVAK-WL-KDNKVKVLEWPSQSPDLNPIENLWAELKKCVRARR- 293
Pse Tzf HoxA          217 DENLLQSALDLRLGRRFIFQQDNDP-KHTANI-SKEWL-QDNSVNVLEWPSQSSDLNLIEHLWRDLKMAVH-RRF 287
Pse Tc1 HoxD          170 SDHVHPFMTTMYPSSDGYFQQDNAPCHKA-RIISNWFLEHDNEFTVLKWPPQSPDLNPIEHLWDVVERELRALDV 243

SB10                  294 -PTNLTQLHQLCQEEWAKIHPTYCGKLVEGYPKRLTQVKQFKGNATKY 340
Salmonid Tc1 ancestor 294 -PTNLTQLHQLCQEEWAKIHPTYCGKLVEGYPKHLTQVKQFKGNATKY 340
Pse Tzf HoxA          288 -PSNLMELERCCKEEWAKLAKDRCAKLEASYSKRLEAVIAAKGASTKY 334
Pse Tc1 HoxD          244 HPTNLHQLQDAILSIWANISKECFQHLVESMPRRIKAVLKAKGGQTPC 291
```

Figure 26.

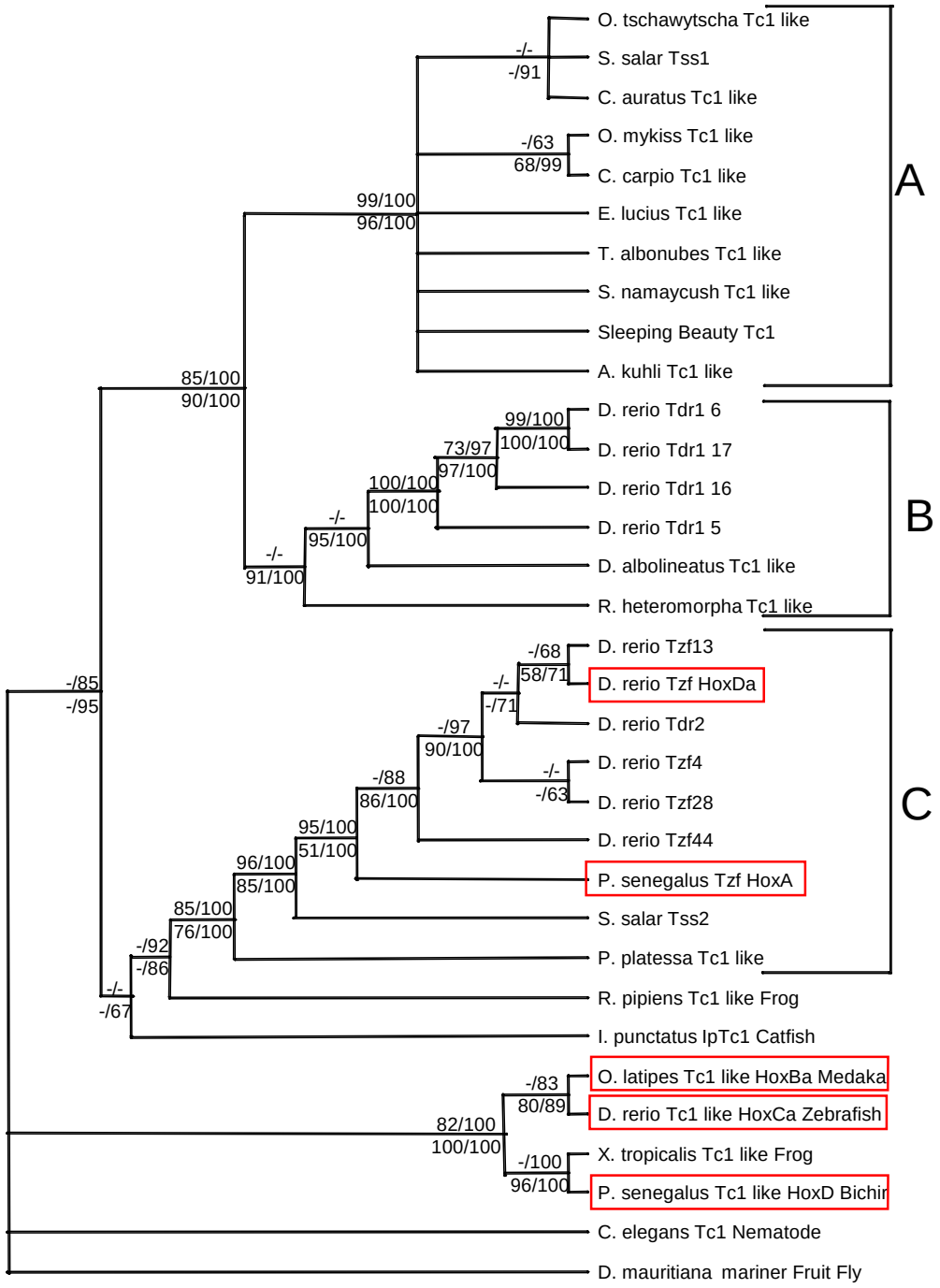

O. tschawytscha Tc1 like
S. salar Tss1
C. auratus Tc1 like
O. mykiss Tc1 like
C. carpio Tc1 like
E. lucius Tc1 like
T. albonubes Tc1 like
S. namaycush Tc1 like
Sleeping Beauty Tc1
A. kuhli Tc1 like
D. rerio Tdr1 6
D. rerio Tdr1 17
D. rerio Tdr1 16
D. rerio Tdr1 5
D. albolineatus Tc1 like
R. heteromorpha Tc1 like
D. rerio Tzf13
D. rerio Tzf HoxDa
D. rerio Tdr2
D. rerio Tzf4
D. rerio Tzf28
D. rerio Tzf44
P. senegalus Tzf HoxA
S. salar Tss2
P. platessa Tc1 like
R. pipiens Tc1 like Frog
I. punctatus IpTc1 Catfish
O. latipes Tc1 like HoxBa Medaka
D. rerio Tc1 like HoxCa Zebrafish
X. tropicalis Tc1 like Frog
P. senegalus Tc1 like HoxD Bichir
C. elegans Tc1 Nematode
D. mauritiana mariner Fruit Fly
A
B
C
-/-
-/91
-/63
68/99
99/100
96/100
85/100
90/100
99/100
100/100
73/97
97/100
100/100
100/100
-/-
95/100
-/-
91/100
-/68
58/71
-/-
-/71
-/97
90/100
-/-
-/63
-/88
86/100
95/100
51/100
96/100
85/100
85/100
76/100
-/92
-/86
-/-
-/67
-/85
-/95
-/83
80/89
82/100
100/100
-/100
96/100

Figure 27.

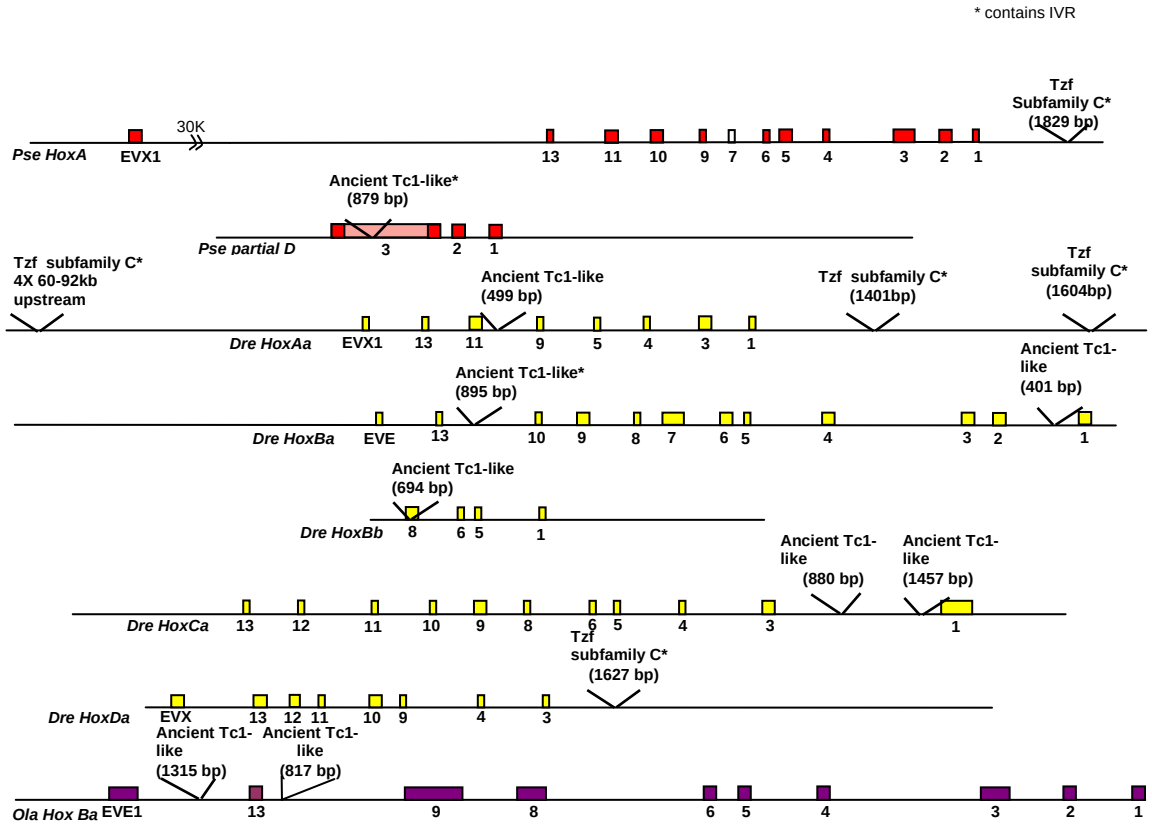


Figure 28.

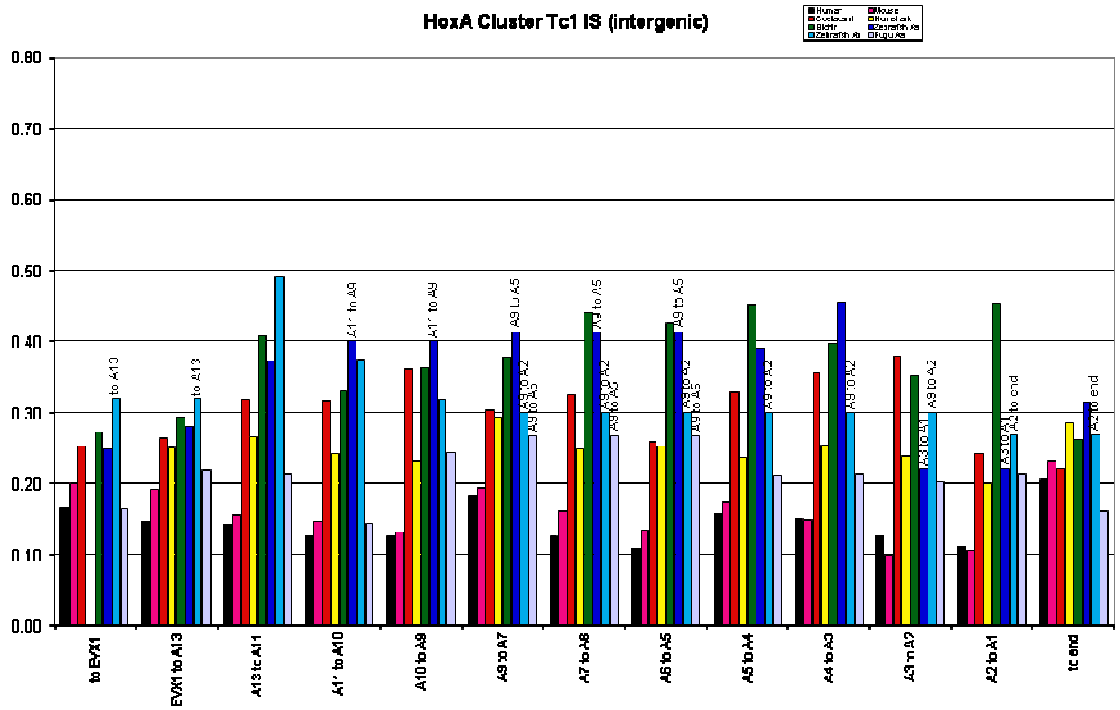
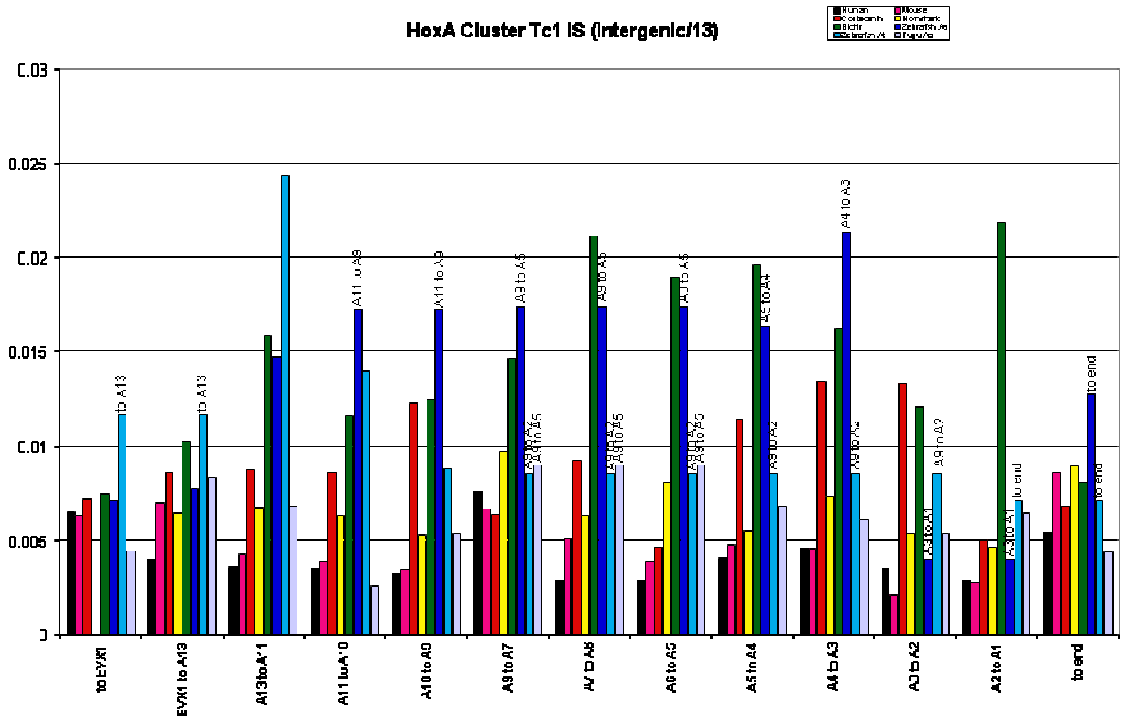


Figure 29.


HoxA Cluster Tc1 IS (Intergenic/13)
0.03
0.025
0.02
0.015
0.01
0.005
0
to EVX1
EVX1 to A13
A13 to A11
A11 to A10
A10 to A9
A9 to A7
A7 to A6
A6 to A5
A5 to A4
A4 to A3
A3 to A2
A2 to A1
to end

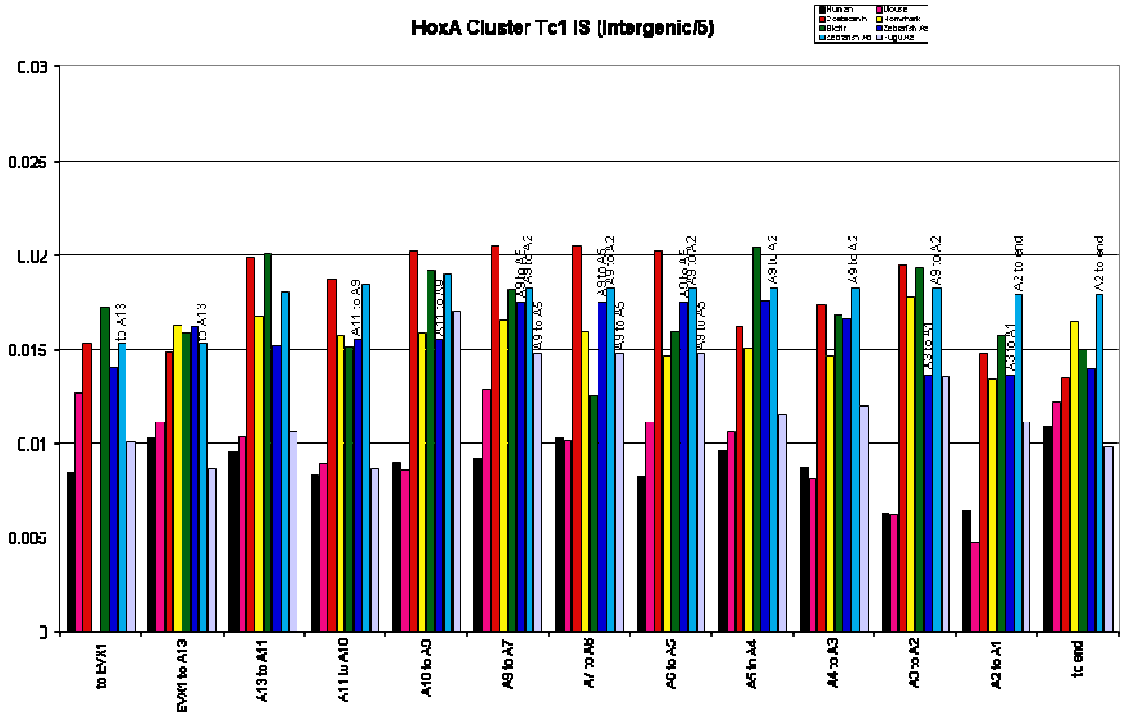
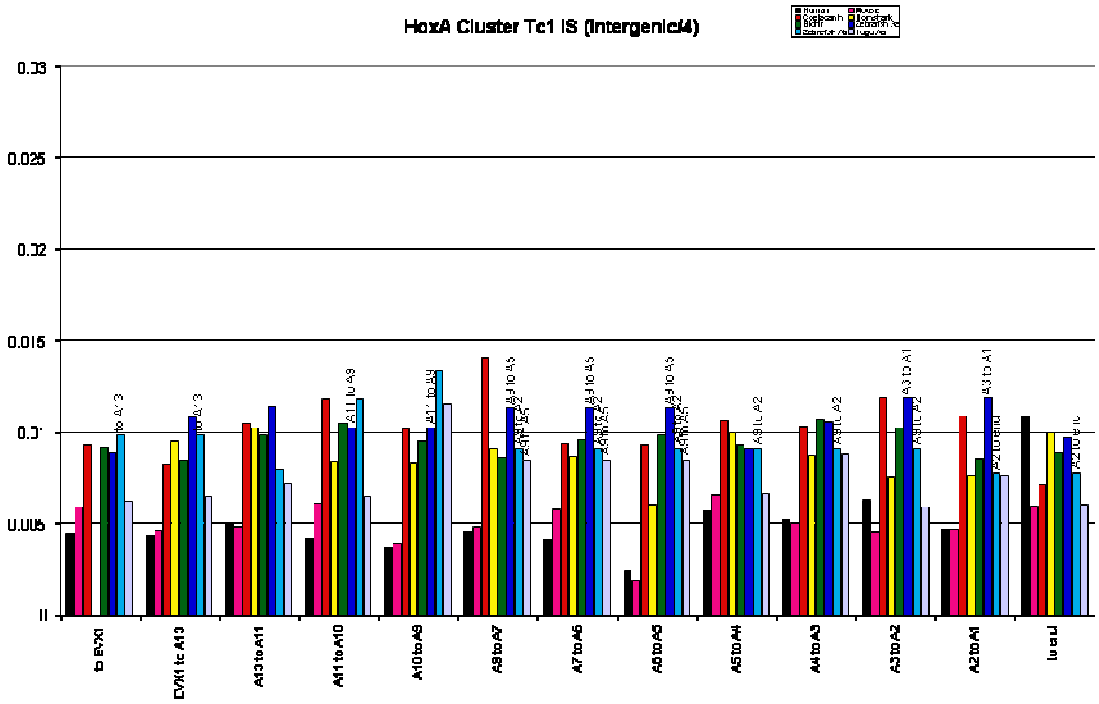

HoxA Cluster Tc1 IS (Intergenic/5)
0.03
0.025
0.02
0.015
0.01
0.005
0
to EVX1
EVX1 to A13
A13 to A11
A11 to A10
A10 to A9
A9 to A7
A7 to A6
A6 to A5
A5 to A4
A4 to A3
A3 to A2
A2 to A1
to end

Figure 29. (continued)



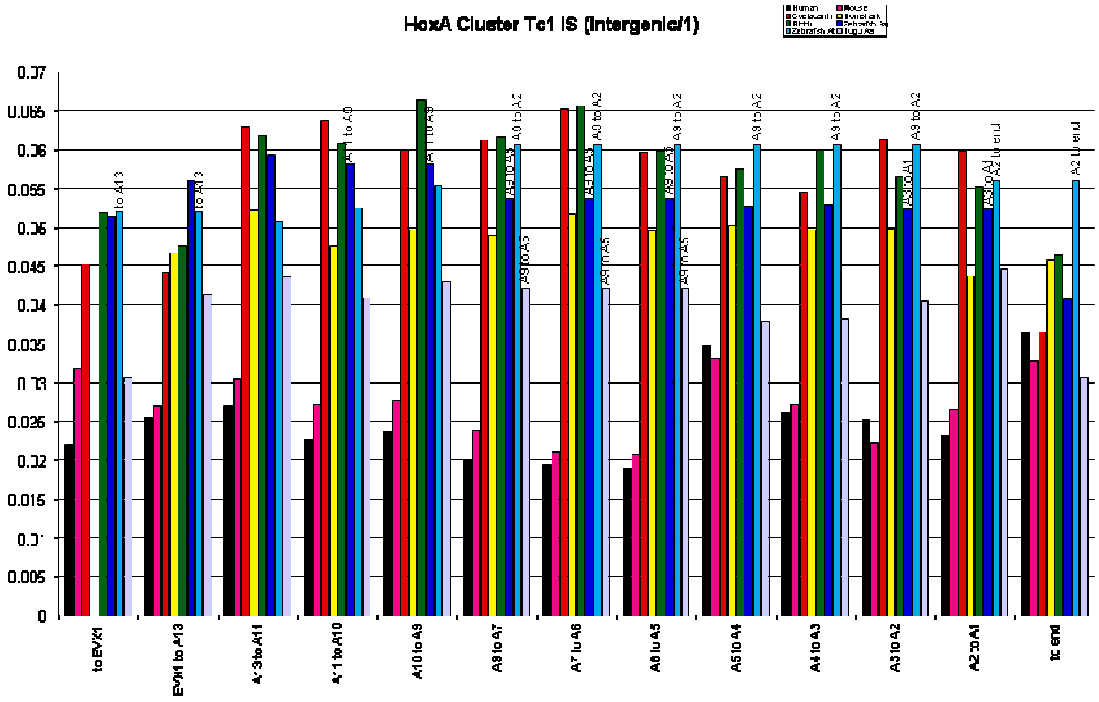
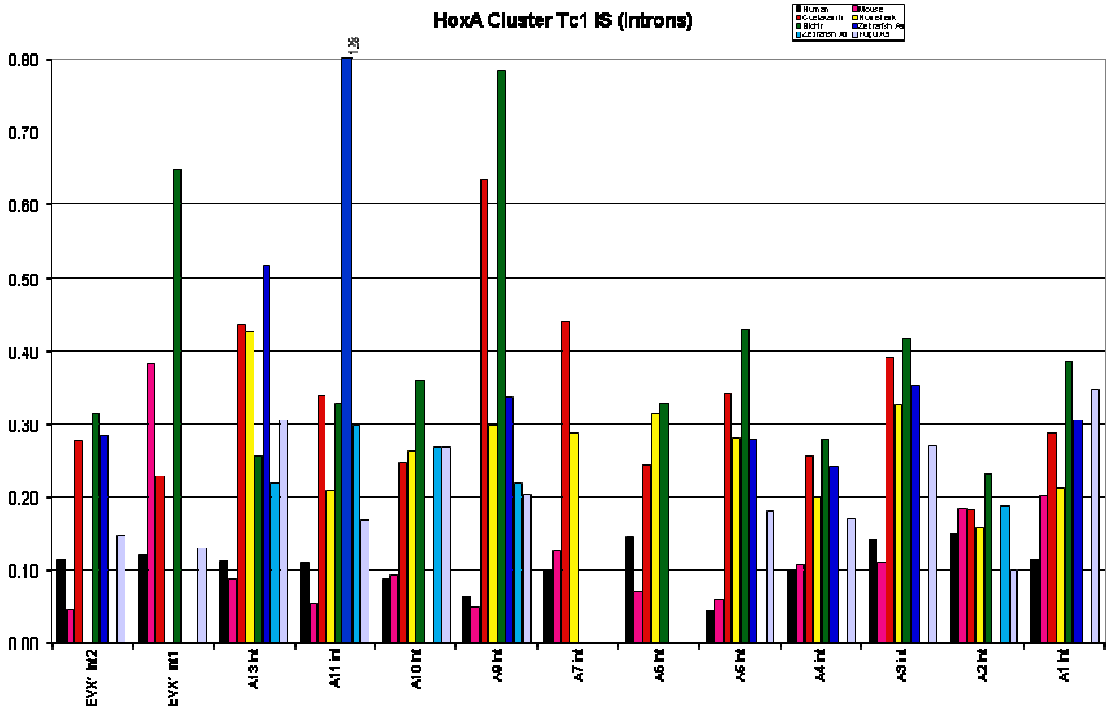


Figure 29. (continued)



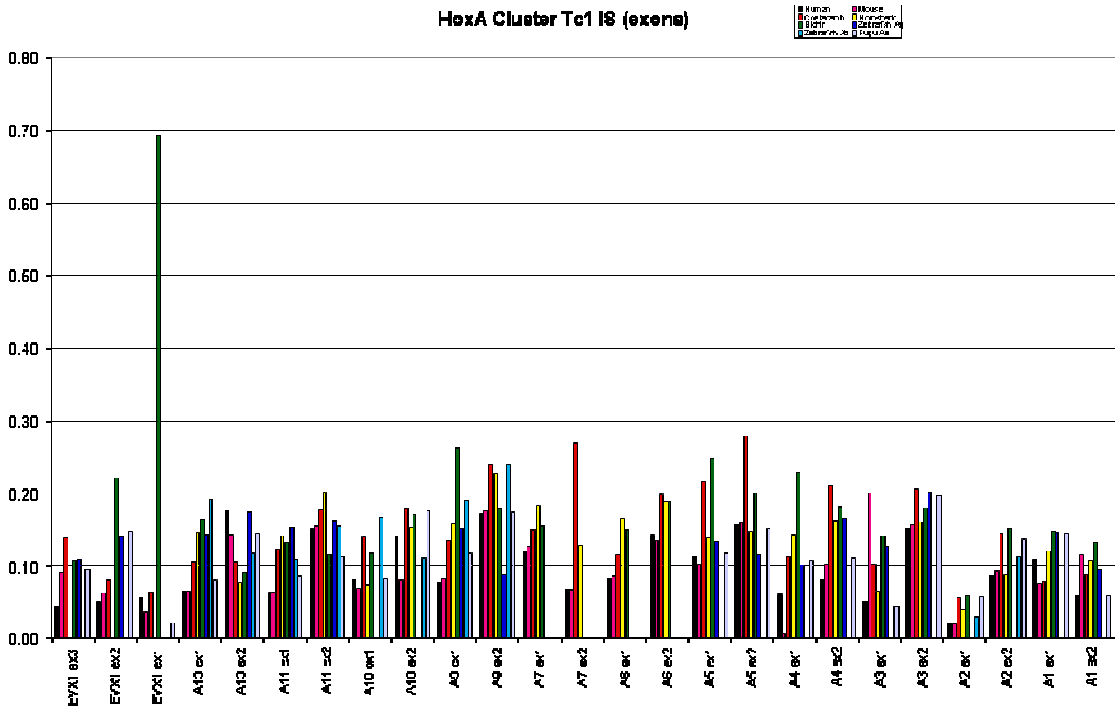


Figure 30.

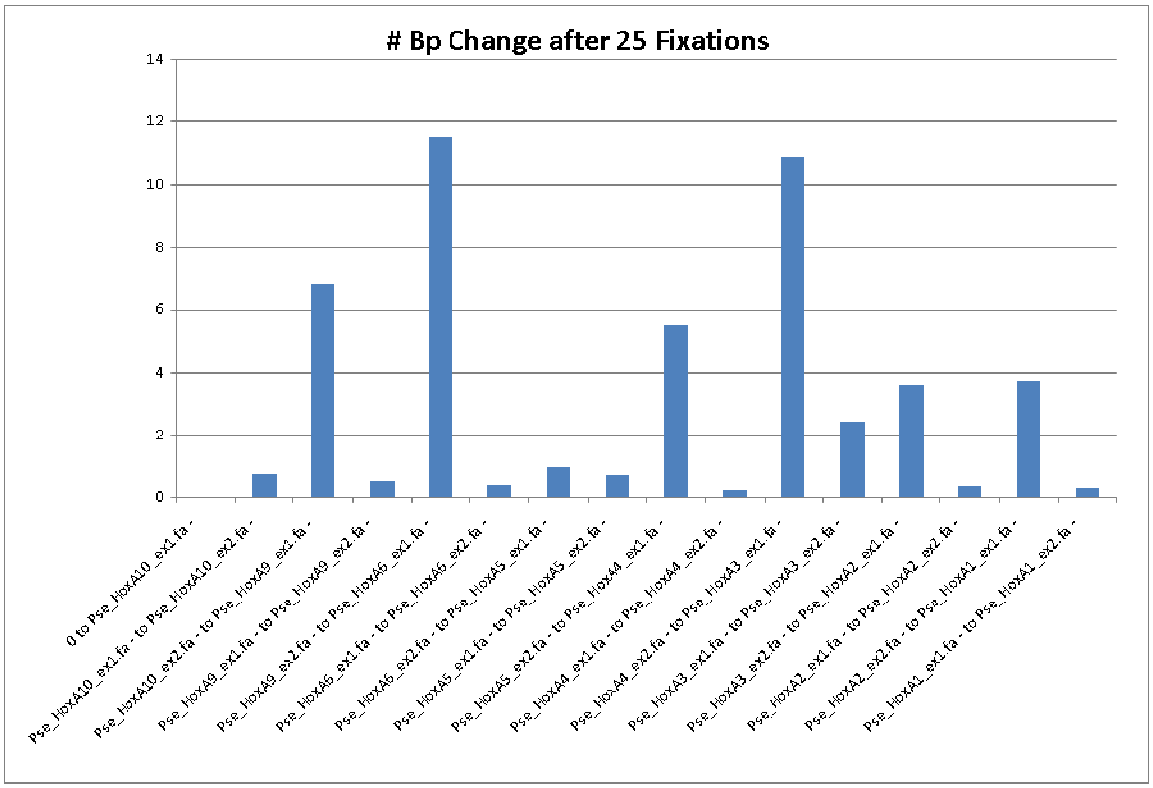
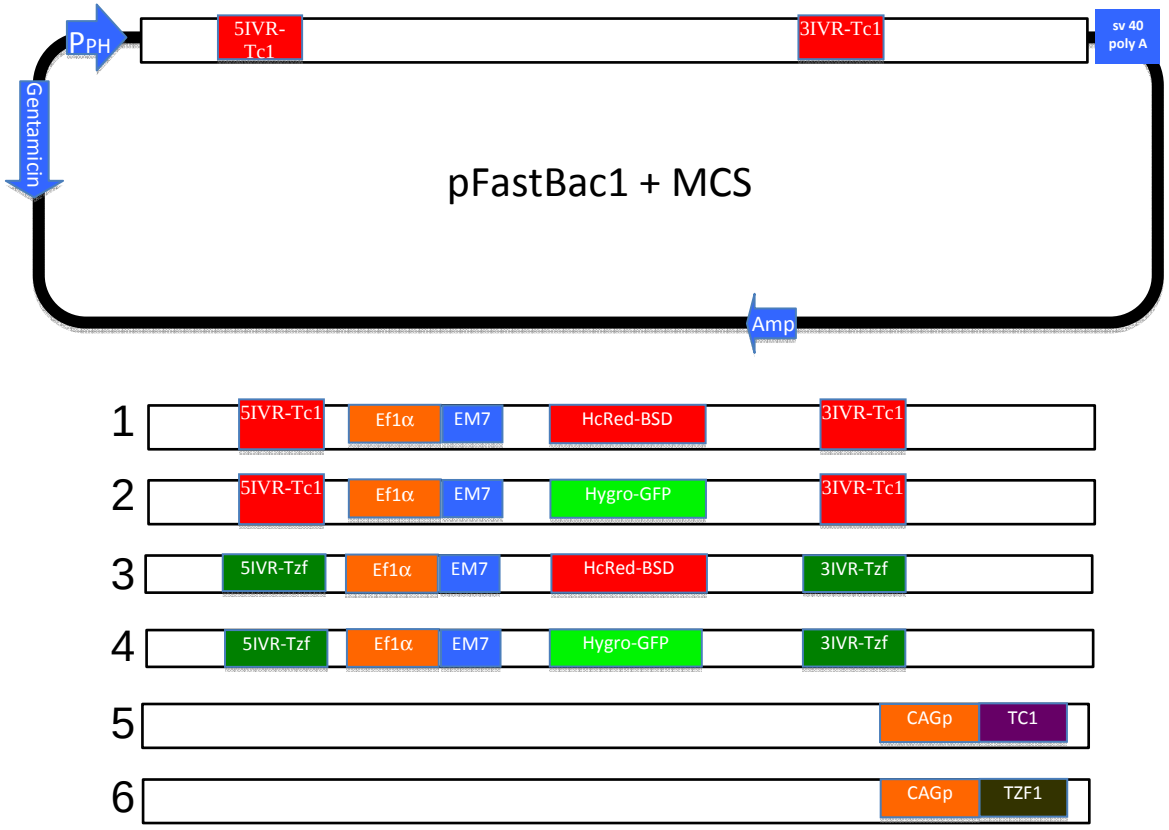


Figure 31.



9. Appendices

Table S1-PFC Database

Each folder contains an “All species” file that contains a list of all PFCs for that particular Hox cluster paralog. For each PFC there is a name in the cell corresponding to the species that contains that particular PFC. For each individual Hox cluster there is a file that contains all PFCs contained within that Hox cluster, the distance to the 3’ Hox gene, the length of the PFC in that Hox cluster, and the sequence of the PFC from that Hox cluster. For the Human Hox cluster files and asterisk has been placed next to the name of all PFCs that also correspond to conserved non-coding sequences found between human (Homo sapiens) and pufferfish (Takifugu rubripes) by Lee et al. (2006).

PFC	Hornshark A	Frog A	Coelacanth A	Human A	Bichir A	Zebrafish AA	Striped Bass A	Nile Tilapia A	Medaka AA	Tetraodon AA	Takifugu AA	Zebrafish AB	Medaka AB	Tetraodon AB	Takifugu AB
1					PseA$_{AC}$1	DreAA$_{AC}$1			OlaAA$_{AC}$1		TruAA$_{AC}$1				
2			LmeA$_{AB}$2	HsaA$_{AC}$2	PseA$_{AC}$2										
3		XtrA$_{AC}$3		HsaA$_{AC}$3											
4		XtrA$_{AC}$4		HsaA$_{AC}$4											
5				HsaA$_{AC}$5							TruAA$_{AC}$5				
6				HsaA$_{AC}$6	PseA$_{AC}$6										
7					PseA$_{AC}$7	DreAA$_{AC}$7									
8				HsaA$_{AC}$8							TruAA$_{AC}$8				
9				HsaA$_{AC}$9		DreAA$_{AC}$9									
10				HsaA$_{AC}$10						TniAA$_{AC}$10					
11				HsaA$_{AC}$11		DreAA$_{AC}$11									
12					PseA$_{AC}$12	DreAA$_{AC}$12									
13					PseA$_{AC}$13	DreAA$_{AC}$13									
14				HsaA$_{AC}$14							TruAA$_{AC}$14				
15		XtrA$_{AC}$15		HsaA$_{AC}$15											
16				HsaA$_{AC}$16						TniAA$_{AC}$16					
17					PseA$_{AC}$17					TniAA$_{AC}$17					
18						DreAA$_{AC}$18				TniAA$_{AC}$18					
19					PseA$_{AC}$19	DreAA$_{AC}$19									
20						DreAA$_{AC}$20			OlaAA$_{AC}$20						
21						DreAA$_{AC}$21					TruAA$_{AC}$21				
22				HsaA$_{AC}$22	PseA$_{AC}$22										
23		XtrA$_{AC}$23		HsaA$_{AC}$23											
24	HfrA$_{AC}$24			HsaA$_{AC}$24											
25	HfrA$_{AC}$25			HsaA$_{AC}$25	PseA$_{AC}$25										
26		XtrA$_{AC}$26		HsaA$_{AC}$26											
27		XtrA$_{AC}$27		HsaA$_{AC}$27											
28		XtrA$_{AC}$28		HsaA$_{AC}$28											
29			LmeA$_{AB}$29	HsaA$_{AC}$29											
30	HfrA$_{AC}$30$_{ab}$	XtrA$_{AC}$30$_{ab}$	LmeA$_{BC}$30$_{b}$	HsaA$_{AC}$30$_{bc}$	PseA$_{AC}$30$_{ab}$	DreAA$_{AC}$30$_{b}$		OniA$_{AC}$30$_{ab}$	OlaAA$_{AC}$30$_{ab}$	TniAA$_{AC}$30$_{abc}$	TruAA$_{AC}$30$_{abc}$	DreAB$_{AC}$30$_{abc}$	OlaAB$_{AC}$30$_{bc}$	TniAB$_{AC}$30$_{bc}$	
31	HfrA$_{AC}$31$_{abcd}$				PseA$_{AC}$31$_{abcdef}$	DreAA$_{AC}$31$_{bdef}$			OlaAA$_{AC}$31$_{cdef}$	TniAA$_{AC}$31$_{cdef}$	TruAA$_{AC}$31$_{abdef}$	DreAB$_{AC}$31$_{ad}$			
32													OlaAB$_{CE}$32	TniAB$_{CE}$32	
33													OlaAB$_{CE}$33	TniAB$_{CE}$33	
34													OlaAB$_{CE}$34	TniAB$_{CE}$34	
35		XtrA$_{CE}$35		HsaA$_{CE}$35											
36				HsaA$_{CE}$36				OniA$_{CE}$36							
37				HsaA$_{CE}$37		DreAA$_{CE}$37									
38			LmeA$_{CE}$38	HsaA$_{CE}$38											
39		XtrA$_{CE}$39		HsaA$_{CE}$39											
40				HsaA$_{CE}$40		DreAA$_{CE}$40									
41						DreAA$_{CE}$41				TniAA$_{CE}$41					
42				HsaA$_{CE}$42				OniA$_{CE}$42							
43	HfrA$_{CE}$43$_{a}$	XtrA$_{CE}$43$_{ab}$		HsaA$_{CE}$43$_{ab}$	PseA$_{CE}$43$_{ab}$										
44					PseA$_{CE}$44							DreAB$_{CE}$44			
45					PseA$_{CE}$45	DreAA$_{CE}$45									
47				HsaA$_{CE}$47					OlaAA$_{CE}$47						
48				HsaA$_{CE}$48							TruAA$_{CE}$48				
49				HsaA$_{CE}$49					OlaAA$_{CE}$49						
50						DreAA$_{CE}$50			OlaAA$_{CE}$50						
51				HsaA$_{CE}$51										TniAB$_{CE}$51	
52	HfrA$_{CE}$52			HsaA$_{CE}$52											
53				HsaA$_{CE}$53							TruAA$_{CE}$53				
54				HsaA$_{CE}$54				OniA$_{CE}$54							
55					PseA$_{CE}$55						TruAA$_{CE}$55				
56					PseA$_{CE}$56							DreAB$_{CE}$56			
57		XtrA$_{CE}$57$_{cd}$	LmeA$_{CE}$57$_{abcd}$	HsaA$_{CE}$57$_{abcd}$	PseA$_{CE}$57$_{abc}$			OniA$_{CE}$57$_{bc}$	OlaAA$_{CE}$57$_{bc}$	TniAA$_{CE}$57$_{bc}$	TruAA$_{CE}$57$_{bc}$				
58												DreAB$_{CE}$58	OlaAB$_{CE}$58		
59				HsaA$_{CE}$59						TniAA$_{CE}$59					
60	HfrA$_{CE}$60$_{ad}$	XtrA$_{CE}$60$_{cdfh}$	LmeA$_{CE}$60$_{abcfh}$	HsaA$_{CE}$60$_{cdefh}$	PseA$_{CE}$60$_{abd}$	DreAA$_{CE}$60$_{abc}$		OniA$_{CE}$60$_{abcd}$	OlaAA$_{CE}$60$_{abcd}$	TniAA$_{CE}$60$_{abcd}$	TruAA$_{CE}$60$_{abcd}$	DreAB$_{CE}$60$_{abc}$			
61													OlaAB$_{EF}$61		TruAB$_{EF}$61
62	HfrA$_{EF}$62		LmeA$_{EF}$62	HsaA$_{EF}$62	PseA$_{EF}$62										
63				HsaA$_{EF}$63	PseA$_{EF}$63										
64					PseA$_{EF}$64							DreAB$_{EF}$64			
65					PseA$_{EF}$65$_{bc}$	DreAA$_{EF}$65$_{bc}$		OniA$_{EF}$65$_{bcd}$	OlaAA$_{EF}$65$_{bcd}$	TniAA$_{EF}$65$_{bc}$	TruAA$_{EF}$65$_{bcd}$	DreAB$_{EF}$65$_{abcde}$	OlaAB$_{EF}$65$_{acde}$	TniAB$_{EF}$65$_{acde}$	TruAB$_{EF}$65$_{acde}$
66					PseA$_{EF}$66							DreAB$_{EF}$66	OlaAB$_{EF}$66	TniAB$_{EF}$66	TruAB$_{EF}$66
67					PseA$_{EF}$67					TniAA$_{EF}$67					
68					PseA$_{EF}$68								OlaAB$_{EF}$68		
69			LmeA$_{EF}$69	HsaA$_{EF}$69											
70				HsaA$_{EF}$70		DreAA$_{EF}$70									
PFC	Hornshark A	Frog A	Coelacanth A	Human A	Bichir A	Zebrafish AA	Striped Bass A	Nile Tilapia A	Medaka AA	Tetraodon AA	Takifugu AA	Zebrafish AB	Medaka AB	Tetraodon AB	Takifugu AB
71				HsaA$_{EF}$71							TruAA$_{EF}$71				
72				HsaA$_{EF}$72				OniA$_{EF}$72							
73	HfrA$_{EF}$73			HsaA$_{EF}$73											
74					PseA$_{EF}$74							DreAB$_{EF}$74			
75	HfrA$_{EF}$75$_{bce}$	XtrA$_{EF}$75$_{abcde}$	LmeA$_{EF}$75$_{abde}$	HsaA$_{EF}$75$_{acde}$	PseA$_{EF}$75$_{acde}$			OniA$_{EF}$75$_{acde}$	OlaAA$_{EF}$75$_{acde}$	TniAA$_{EF}$75$_{ade}$	TruAA$_{EF}$75$_{acde}$	DreAB$_{EF}$75$_{abc}$	OlaAB$_{EF}$75$_{abc}$	TniAB$_{EF}$75$_{abcd}$	TruAB$_{EF}$75$_{abcd}$
76					PseA$_{FG}$76		MsaA$_{FG}$76	OniA$_{FG}$76							
77		XtrA$_{FG}$77		HsaA$_{FG}$77											
78	HfrA$_{FG}$78$_{abc}$	XtrA$_{FG}$78$_{ac}$	LmeA$_{FG}$78$_{abc}$	HsaA$_{FG}$78$_{abc}$	PseA$_{FG}$78$_{ab}$										
79	HfrA$_{FG}$79$_{ac}$	XtrA$_{FG}$79$_{bc}$	LmeA$_{FG}$79$_{b}$	HsaA$_{FG}$79$_{bc}$	PseA$_{FG}$79$_{abde}$	DreAA$_{FG}$79$_{ad}$	MsaA$_{FG}$79$_{abc}$	OniA$_{FG}$79$_{abde}$	OlaAA$_{FG}$79$_{abcde}$	TniAA$_{FG}$79$_{abdfg}$	TruAA$_{FG}$79$_{abdfg}$	DreAB$_{FG}$79$_{bd}$		TniAB$_{FG}$79$_{bd}$	
80					PseA$_{FG}$80										TruAB$_{FG}$80
81												DreAB$_{FG}$81$_{ab}$	OlaAB$_{FG}$81$_{a}$	TniAB$_{FG}$81$_{ab}$	TruAB$_{FG}$81$_{ab}$
82						DreAA$_{FG}$82				TniAA$_{FG}$82					
83					PseA$_{FG}$83			OniA$_{FG}$83							
84					PseA$_{FG}$84					TniAA$_{FG}$84					
85			LmeA$_{FG}$85	HsaA$_{FG}$85											
86				HsaA$_{FG}$86								DreAB$_{FG}$86	OlaAB$_{FG}$86	TniAB$_{FG}$86	TruAB$_{FG}$86
87													OlaAB$_{FG}$87		TruAB$_{FG}$87
88				HsaA$_{FG}$88				OniA$_{FG}$88							
89				HsaA$_{FG}$89					OlaAA$_{FG}$89						
90				HsaA$_{FG}$90											TruAB$_{FG}$90
91	HfrA$_{FG}$91			HsaA$_{FG}$91											
92	HfrA$_{FG}$92$_{cdef}$	XtrA$_{FG}$92$_{cdefg}$	LmeA$_{FG}$92$_{defg}$	HsaA$_{FG}$92$_{defgh}$	PseA$_{FG}$92$_{defgh}$							DreAB$_{FG}$92$_{gh}$			
93						DreAA$_{FG}$93$_{acn}$	MsaA$_{FG}$93$_{abcdef}$	OniA$_{FG}$93$_{abdef}$	OlaAA$_{FG}$93$_{abcdef}$	TniAA$_{FG}$93$_{abcdef}$	TruAA$_{FG}$93$_{abcdef}$				
94					PseA$_{CE}$94			OniA$_{CE}$94	OlaAA$_{CE}$94						
95						DreAA$_{CE}$95					TruAA$_{CE}$95				
96				HsaA$_{CE}$96				OniA$_{CE}$96							
97	HfrA$_{CE}$97	XtrA$_{CE}$97	LmeA$_{CE}$97	HsaA$_{CE}$97											
98	HfrA$_{CE}$98$_{b}$			HsaA$_{CE}$98$_{b}$		DreAA$_{CE}$98$_{abcde}$	MsaA$_{CE}$98$_{abcde}$	OniA$_{CE}$98$_{abcd}$	OlaAA$_{CE}$98$_{abcd}$	TniAA$_{CE}$98$_{abcde}$	TruAA$_{CE}$98$_{abcde}$	DreAB$_{CE}$98$_{bc}$			
99						DreAA$_{CE}$99					TruAA$_{CE}$99				
100					PseA$_{CE}$100	DreAA$_{CE}$100									
101					PseA$_{CE}$101									TniAB$_{CE}$101	
102					PseA$_{CE}$102							DreAB$_{CE}$102			
103	HfrA$_{CE}$103$_{ac}$	XtrA$_{CE}$103$_{acd}$	LmeA$_{CE}$103$_{acn}$	HsaA$_{CE}$103$_{acde}$	PseA$_{CE}$103$_{abcde}$		MsaA$_{CE}$103$_{bcdef}$	OniA$_{CE}$103$_{bcdef}$	OlaAA$_{CE}$103$_{bcdef}$	TniAA$_{CE}$103$_{bcdef}$	TruAA$_{CE}$103$_{bcdef}$				
104				HsaA$_{CE}$104									OlaAB$_{CE}$104		
105		XtrA$_{CE}$105		HsaA$_{CE}$105											
106				HsaA$_{CE}$106										TniAB$_{CE}$106	
107				HsaA$_{CE}$107			MsaA$_{CE}$107								
108					PseA$_{CE}$108		MsaA$_{CE}$108								
109					PseA$_{CE}$109										TruAB$_{CE}$109
110		XtrA$_{CE}$110	LmeA$_{CE}$110	HsaA$_{CE}$110											
111	HfrA$_{CE}$111$_{ab}$	XtrA$_{CE}$111$_{ab}$	LmeA$_{CE}$111$_{ab}$	HsaA$_{CE}$111$_{ab}$			MsaA$_{CE}$111$_{b}$	OniA$_{CE}$111$_{b}$	OlaAA$_{CE}$111$_{b}$	TniAA$_{CE}$111$_{b}$	TruAA$_{CE}$111$_{b}$				
112					PseA$_{CE}$112										TruAB$_{CE}$112
113					PseA$_{CE}$113					TniAA$_{CE}$113					
114				HsaA$_{CE}$114		DreAA$_{CE}$114									
115						DreAA$_{CE}$115	MsaA$_{CE}$115								
116						DreAA$_{CE}$116					TruAA$_{CE}$116				
117				HsaA$_{CE}$117			MsaA$_{CE}$117								
118		XtrA$_{CE}$118		HsaA$_{CE}$118											
119	HfrA$_{CE}$119$_{abcd}$	XtrA$_{CE}$119$_{abcd}$	LmeA$_{CE}$119$_{abcd}$	HsaA$_{CE}$119$_{abcd}$	PseA$_{CE}$119$_{abcd}$	DreAA$_{CE}$119$_{abd}$	MsaA$_{CE}$119$_{bcd}$	OniA$_{CE}$119$_{abc}$	OlaAA$_{CE}$119$_{abc}$	TniAA$_{CE}$119$_{abc}$	TruAA$_{CE}$119$_{abc}$				
120						DreAA$_{CE}$120			OlaAA$_{CE}$120						
121					PseA$_{CE}$121									TniAB$_{CE}$121	
122		XtrA$_{CE}$122	LmeA$_{CE}$122	HsaA$_{CE}$122											
123				HsaA$_{CE}$123			MsaA$_{CE}$123								
124	HfrA$_{CE}$124			HsaA$_{CE}$124											
125				HsaA$_{CE}$125							TruAA$_{CE}$125				
126				HsaA$_{CE}$126					OlaAA$_{CE}$126						
127				HsaA$_{CE}$127		DreAA$_{CE}$127									
128					PseA$_{CE}$128		MsaA$_{CE}$128								

129				$Hsa A_{G}129$						$TniA A_{GK}129$					
130	$HfrA_{J}130_{ab}$	$XtrA_{J}130_{ab}$	$LmeA_{J}130_{ab}$		$PseA_{J}130_{a}$										
131					$PseA_{CN}131$										$TruAB_{CN}131$
132	$HfrA_{JB}132_{ac}$	$XtrA_{JB}132_{ac}$	$LmeA_{JB}132_{ac}$	$HsaA_{JB}132_{ac}$	$PseA_{JB}132_{ac}$	$DreAA_{GK}132_{abcd}$	$MsaA_{JB}132_{abcd}$	$OniA_{GK}132_{abcd}$	$OlaA A_{JB}132_{abcd}$	$TniA A_{GK}132_{abcd}$	$TruAA_{GK}132_{abcd}$				
133			$LmeA_{KL}133$	$HsaA_{KL}133$	$PseA_{KL}133$										
134			$LmeA_{KL}134$	$HsaA_{KL}134$	$PseA_{KL}134$										
135					$PseA_{CN}135$							$DreAB_{CN}135$			
136	$HfrA_{KL}136_{abc}$	$XtrA_{KL}136_{abc}$	$LmeA_{KL}136_{abc}$	$HsaA_{KL}136_{ab}$	$PseA_{KL}136_{bc}$	$DreAA_{KL}136_{bc}$	$MsaA_{KL}136_{abc}$	$OniA_{KL}136_{abc}$	$OlaA A_{KL}136_{ab}$	$TniA A_{KL}136_{ab}$	$TruAA_{KL}136_{abc}$				
137			$LmeA_{KL}137$	$HsaA_{KL}137$											
138	$HfrA_{KL}138_{b}$	$XtrA_{KL}138_{abc}$		$HsaA_{KL}138_{abc}$											
139				$HsaA_{KL}139$			$MsaA_{KL}139$								
140	$HfrA_{JG}140_{cde}$	$XtrA_{JG}140_{bcde}$	$LmeA_{JG}140_{bcde}$	$HsaA_{JG}140_{bcde}$	$PseA_{JG}140_{bcde}$	$DreAA_{JG}140_{cde}$	$MsaA_{JG}140_{abcde}$	$OniA_{JG}140_{abcde}$	$OlaA A_{JG}140_{abcde}$	$TniA A_{JG}140_{abcde}$	$TruAA_{JG}140_{abcd}$				
PFC	Hornshark A	Frog A	Coelacanth A	Human A	Bichir A	Zebrafish AA	Striped Bass A	Nile Tilapia A	Medaka AA	Tetraodon AA	Takifugu AA	Zebrafish AB	Medaka AB	Tetraodon AB	Takifugu AB
141												$DreAB_{CN}141$		$TniAB_{CN}141$	
142					$PseA_{KL}142$							$DreAB_{CN}142$			
143	$HfrA_{JG}143_{abc}$	$XtrA_{JG}143_{abc}$	$LmeA_{JG}143_{ab}$	$HsaA_{JG}143_{abc}$	$PseA_{JG}143_{b}$	$DreAA_{JG}143_{bcd}$	$MsaA_{JG}143_{abd}$	$OniA_{JG}143_{abcd}$	$OlaA A_{JG}143_{bcd}$	$TniA A_{JG}143_{bcd}$					
144					$PseA_{KL}144$	$DreAA_{KL}144$									
145				$HsaA_{KL}145$				$OniA_{KL}145$							
146				$HsaA_{KL}146$					$OlaA A_{KL}146$						
147	$HfrA_{KL}147$			$HsaA_{KL}147$											
148	$HfrA_{KL}148$			$HsaA_{KL}148$											
149	$HfrA_{KL}149_{abc}$	$XtrA_{KL}149_{abc}$	$LmeA_{KL}149_{abc}$	$HsaA_{KL}149_{abc}$	$PseA_{KL}149_{bc}$	$DreAA_{KL}149_{cde}$	$MsaA_{KL}149_{bcde}$	$OniA_{KL}149_{bcde}$	$OlaA A_{KL}149_{bcde}$	$TniA A_{KL}149_{bcd}$	$TruAA_{KL}149_{bcd}$				
150			$LmeA_{KL}150$	$HsaA_{KL}150$											
151				$HsaA_{KL}151$				$OniA_{KL}151$							
152	$HfrA_{KL}152$			$HsaA_{KL}152$											
153				$HsaA_{KL}153$								$DreAB_{CN}153$			
154				$HsaA_{KL}154$				$OniA_{KL}154$							
155	$HfrA_{KL}155$			$HsaA_{KL}155$											
156			$LmeA_{KL}156$	$HsaA_{KL}156$											
157	$HfrA_{JG}157_{bc}$	$XtrA_{JG}157_{bc}$	$LmeA_{JG}157_{bc}$	$HsaA_{JG}157_{bc}$	$PseA_{JG}157_{abc}$	$DreAA_{JG}157_{abc}$	$MsaA_{JG}157_{abc}$	$OniA_{JG}157_{abc}$	$OlaA A_{JG}157_{abc}$	$TniA A_{JG}157_{abc}$	$TruAA_{JG}157_{abc}$				
158						$DreAA_{LM}158$				$TniA A_{LM}158$					
159					$PseA_{LM}159$							$DreAB_{CN}159$			
160			$LmeA_{LM}160$	$HsaA_{LM}160$											
161	$HfrA_{LM}161$			$HsaA_{LM}161$											
162			$LmeA_{LM}162$	$HsaA_{LM}162$											
163		$XtrA_{LM}163$		$HsaA_{LM}163$											
164				$HsaA_{LM}164$		$DreAA_{LM}164$									
165						$DreAA_{LM}165$			$OlaA A_{LM}165$						
166	$HfrA_{LM}166$			$HsaA_{LM}166$											
167	$HfrA_{LM}167$			$HsaA_{LM}167$											
168	$HfrA_{LM}168$			$HsaA_{LM}168$											
169				$HsaA_{LM}169$		$DreAA_{LM}169$									
170				$HsaA_{LM}170$	$PseA_{LM}170$										
171					$PseA_{LM}171$									$TniAB_{CN}171$	
172					$PseA_{LM}172$			$OniA_{LM}172$	$OlaA A_{LM}172$	$TniA A_{LM}172$	$TruAA_{LM}172$				
173				$HsaA_{LM}173$				$OniA_{LM}173$							
174		$XtrA_{LM}174$		$HsaA_{LM}174$											
175				$HsaA_{LM}175$				$OniA_{LM}175$							
176				$HsaA_{LM}176$	$PseA_{LM}176$										
177	$HfrA_{LM}177$			$HsaA_{LM}177$											
178				$HsaA_{LM}178$		$DreAA_{LM}178$									
179	$HfrA_{LM}179_{abc}$	$XtrA_{LM}179_{abc}$	$LmeA_{LM}179_{abc}$	$HsaA_{LM}179_{abc}$	$PseA_{LM}179_{abc}$			$OniA_{LM}179_{ab}$	$OlaA A_{LM}179_{ab}$	$TniA A_{LM}179_{a}$	$TruAA_{LM}179_{ab}$				
180	$HfrA_{LM}180$				$PseA_{LM}180$										
181	$HfrA_{LM}181_{bcde}$	$XtrA_{LM}181_{bcde}$	$LmeA_{LM}181_{bcde}$	$HsaA_{LM}181_{cd}$	$PseA_{LM}181_{bd}$	$DreAA_{LM}181_{abc}$		$OniA_{LM}181_{abcde}$	$OlaA A_{LM}181_{abcd}$	$TniA A_{LM}181_{abcde}$	$TruAA_{LM}181_{abcde}$				
182				$HsaA_{LM}182$					$OlaA A_{LM}182$						
183						$DreAA_{LM}183$					$TruAA_{LM}183$				
184			$LmeA_{LM}184$	$HsaA_{LM}184$											
185	$HfrA_{LM}185$			$HsaA_{LM}185$											
186				$HsaA_{LM}186$							$TruAA_{LM}186$				
187				$HsaA_{LM}187$							$TruAA_{LM}187$				
188	$HfrA_{LM}188$	$XtrA_{LM}188$	$LmeA_{LM}188$	$HsaA_{LM}188$											
189					$PseA_{LM}189$				$OlaA A_{LM}189$						
190				$HsaA_{LM}190$				$OniA_{LM}190$							
191				$HsaA_{LM}191$				$OniA_{LM}191$							
192				$HsaA_{LM}192$				$OniA_{LM}192$							
193	$HfrA_{LM}193_{cd}$	$XtrA_{LM}193_{cd}$		$HsaA_{LM}193_{cd}$	$PseA_{LM}193_{cd}$	$DreAA_{LM}193_{cd}$		$OniA_{LM}193_{cd}$	$OlaA A_{LM}193_{cd}$	$TniA A_{LM}193_{cd}$	$TruAA_{LM}193_{cd}$				
194	$HfrA_{LM}194_{abc}$	$XtrA_{LM}194_{bcde}$	$LmeA_{LM}194_{bcde}$	$HsaA_{LM}194_{bcde}$	$PseA_{LM}194_{bcde}$	$DreAA_{LM}194_{abcd}$		$OniA_{LM}194_{acd}$	$OlaA A_{LM}194_{acd}$	$TniA A_{LM}194_{abcd}$	$TruAA_{LM}194_{abcd}$				
195					$PseA_{LM}195$							$DreAB_{CN}195$			
196					$PseA_{LM}196$				$OlaA A_{LM}196$						
197						$DreAA_{LM}197$			$OlaA A_{LM}197$						
198	$HfrA_{LM}198_{abc}$	$XtrA_{LM}198_{abc}$	$LmeA_{LM}198_{abc}$	$HsaA_{LM}198_{abc}$	$PseA_{LM}198_{ab}$										
199	$HfrA_{LM}199_{cd}$	$XtrA_{LM}199_{cd}$	$LmeA_{LM}199_{cd}$	$HsaA_{LM}199_{cd}$	$PseA_{LM}199_{cd}$	$DreAA_{LM}199_{cd}$		$OniA_{LM}199_{cd}$	$OlaA A_{LM}199_{cd}$	$TniA A_{LM}199_{cd}$	$TruAA_{LM}199_{cd}$				
200					$PseA_{MN}200$						$TruAA_{MN}200$				
201	$HfrA_{MN}201$	$XtrA_{MN}201$	$LmeA_{MN}201$	$HsaA_{MN}201$											
202	$HfrA_{MN}202$			$HsaA_{MN}202$											
203				$HsaA_{MN}203$				$OniA_{MN}203$							
204		$XtrA_{MN}204_{bcd}$		$HsaA_{MN}204_{acd}$	$PseA_{MN}204_{acd}$			$OniA_{MN}204_{bcd}$	$OlaA A_{MN}204_{acd}$	$TniA A_{MN}204_{acd}$	$TruAA_{MN}204_{acd}$	$DreAB_{CN}204_{cd}$	$OlaAB_{CN}204_{c}$	$TniAB_{CN}204_{cd}$	$TruAB_{CN}204_{cd}$
205					$PseA_{MN}205$							$DreAB_{CN}205$			
206					$PseA_{MN}206$									$TniAB_{CN}206$	
207					$PseA_{MN}207$					$TniA A_{MN}207$					
208				$HsaA_{MN}208$											$TruAB_{CN}208$
209				$HsaA_{MN}209$		$DreAA_{MN}209$									
210	$HfrA_{MN}210_{b}$	$XtrA_{MN}210_{abc}$	$LmeA_{MN}210_{abc}$	$HsaA_{MN}210_{abc}$	$PseA_{MN}210_{ab}$										
PFC	Hornshark A	Frog A	Coelacanth A	Human A	Bichir A	Zebrafish AA	Striped Bass A	Nile Tilapia A	Medaka AA	Tetraodon AA	Takifugu AA	Zebrafish AB	Medaka AB	Tetraodon AB	Takifugu AB
211	$HfrA_{MN}211_{bcdef}$	$XtrA_{MN}211_{bcdef}$	$LmeA_{MN}211_{abcdef}$	$HsaA_{MN}211_{abcdef}$	$PseA_{MN}211_{cdef}$			$OniA_{MN}211_{acdef}$	$OlaA A_{MN}211_{abcdef}$	$TniA A_{MN}211_{abcdef}$	$TruAA_{MN}211_{abcdef}$	$DreAB_{CN}211_{cdef}$	$OlaAB_{CN}211_{c}$	$TniAB_{CN}211_{acd}$	$TruAB_{CN}211_{acd}$
212					$PseA_{NO}212$			$OniA_{NO}212$							
213					$PseA_{NO}213$			$OniA_{NO}213$							
214					$PseA_{NO}214$						$TruAA_{NO}214$				
215	$HfrA_{NO}215$			$HsaA_{NO}215$											
216						$DreAA_{NO}216$		$OniA_{NO}216$	$OlaA A_{NO}216$	$TniA A_{NO}216$	$TruAA_{NO}216$				
217						$DreAA_{NO}217$			$OlaA A_{NO}217$						
218	$HfrA_{NO}218$			$HsaA_{NO}218$											
219	$HfrA_{NO}219$		$LmeA_{NO}219$	$HsaA_{NO}219$				$OniA_{NO}219$	$OlaA A_{NO}219$	$TniA A_{NO}219$					
220				$HsaA_{NO}220$					$OlaA A_{NO}220$						
221				$HsaA_{NO}221$				$OniA_{NO}221$							
222					$PseA_{NO}222$			$OniA_{NO}222$	$OlaA A_{NO}222$	$TniA A_{NO}222$	$TruAA_{NO}222$				
223						$DreAA_{NO}223$			$OlaA A_{NO}223$						

Heterodontus francisci horn shark Hox A

PFC	Dist. To 3' gene	Length	Sequence
$HfrA_{AC}24$	9,775	54	GATTAAAGTTTAACCAGAGGCGTGTTTACGGTTTTACCTTGCTATTTAAACACAT
$HfrA_{AC}25$	9,493	13	CATCTAGTGGGAA
$HfrA_{AC}30_{ab}$	3,064	34	GGCTAGACTGCAACTTTCAACTTGACCTTGGCCT
$HfrA_{AC}31_{abcdf}$	150	141	CTTTGCATATCATGTGATCATCTCGCAGCCAATGAGGAGGCAAGTTTCAGCGACACTGGCGCGTCAATGCTGCTAAGTCTGAGAGACTCAGTATCTCTTTTCAGTCTTAGGGGCTTTTAAAACGGCCACTAGTCTCAATGT
$HfrA_{CE}43_{a}$	10,846	44	TTAAACCTTGACCTTGACGAAAACTGATAGTCTTTGACCTTGAC
$HfrA_{CE}52$	8,880	24	GATTAACCTGGTTTATTGTAAAAA
$HfrA_{CE}60_{ad}$	191	117	TGTCAATTTCTCCAGGCATCTGGTCACATGACTGGCTTGTCCCTGCGATGTATGGAGATACATTTCCACGTCAGCTTACGTTTCCAAATTTTTACTCCGCGGATCAGCTTCAAAGAG
$HfrA_{EF}62$	9,010	40	TTGAACTTCTGTTGCGCCATGGACGTCACCTCTAACCTTG
$HfrA_{EF}73$	1,534	20	GGGAGGGTGTGGGGGGAGGG
$HfrA_{EF}75_{bce}$	337	105	ATTGGCTCGTCCCGTCTGTCGTGCTGCGGAGGTATCTCTAATCATACTCAGCGTGTTTGCACAAGAAATGTCAGCCTGAACTGGATATCTTCTGCTTTCGCCAAA
$HfrA_{FG}78_{abc}$	6,925	170	ATCTATTTAAATATTACCTAGACGGTCGTAATTTGTCTGGGCTATATAGCAATGGTGCTGTAACGTTTGTCGGCTTTTGTTGAGTTTTATGACTTGCTAGTATATCTGGGTTGTTGCTGTCTTGGACGAGTGGTTTCAGTTGAGCGAACAGCCACGAAGACAGAGGAAGC
$HfrA_{FG}79_{bc}$	5,761	54	AGGTAGTTTCATGTTGTTGGGGCTCAAGTCTATCTCTACAACACGAAACTGCCT
$HfrA_{FG}91$	416	79	TTTTATGTGAAGGTTAACGGTTTTAAGGTGTTCGTTAAAAAAAACGAGCATCTGATGGGAAGATTGATTATGATATAAAA
$HfrA_{FG}92_{acdef}$	237	203	CATAAAAACATACGGAGTTGCTATATAAAATTATGTCTATAAAAGATTGACTCTGTTACAAGCTGCGGATTGATTTATATTTTATTGTTGTTTTTTGAATCACGTGAGTAACTTAGCCAATGGCATGGAAGTGCTTCTGCACCATATTAGTAGACTGTACTTGTCGCGGCGGTAAGTTGTTACATTGAAATCTGCTCTTTCAT
$HfrA_{GI}97$	6,493	23	TACCTCGCGTCATATTGACCAAA
$HfrA_{GI}98_{b}$	5,339	37	ATAAAAGAAAATGACTGCTGTAAGCGTTTATGGGGTG
$HfrA_{GI}103_{ac}$	5,105	99	ACGCCCTCTGTTGTTGCAGATGGAACTTCAAAGACTAAACTGGGAGACTTGCAATAAAATGAAGGGTCTGGCAAGCCGTCTGGTATTCAGGTTTTAATG
$HfrA_{GI}111_{ab}$	184	171	CACCTGTGATGTGGAATGCTATTGGGGGACGGGATCACATGGTCGGGTCACGTGAATAGAAATAAGTCGGACTGAAGGAAAATGGGTGTTTGGTGTAAATCGGGGGTGTAATGCTGCCATATATCACGCTATCTCGTAAAACCGACACTGAAGAGCCTTGGCCAACAAATC
$HfrA_{IJ}119_{abcd}$	5,493	187	ATGGCGTCTAGGACCTGCCTCCTAATAGTCATGGGGAAATGTCATAAATCCGTTGTTGTTTTATGAAAATTTACAACTTTGCAATAGAACTTTACGAGTTGTTCGGTTTTTCCATTGGCCGCTGAAAGTCATGTGGAGTGTAACCGTGAACATGAACTTTTTATAATTTCCCTTGCGACTATAGAGC
$HfrA_{IJ}124$	3,635	46	TTCCCCACAATTAGTTATTGGATCCTGATTGATCGTTAGTAAATAG
$HfrA_{IJ}130_{ab}$	136	111	GATCAGCTGGTTATATTTGCTGCTGTCGCTTTGGCGCTCGGCCATCCTGAAACAAACCACCTCGATGACTACTAATAGCTAAACCACATTGTGTACTAATACATAACAAAT
$HfrA_{JK}132_{ac}$	324	297	TGGCGCGCAAGGACAACCCAAAGCTACTACAGGGAGCCGTCGGCGCGCCCCCGTGCGCGTGCCCCACGTTCTTCCTTCCACGCAACTTTCCAATTTGTATCACAGTGCACGGATTTACCTCTAGAGGTCATCAGCGAGAATTTACGACTGGACAACAAAAGCACGTGATTCCAAGTCGTACCCCATATTTGGGCGCCTACGTAGGAGGGAACGAAGTACATGTCCCAGTCATTTCCATAATTCATCATAAATTGT
$HfrA_{KL}136_{abc}$	10,279	210	AAACTTTTATACACCTCAGTTCCGGCTTTATGACATTTGGGTGCCAAATGAATAGGGTTTTGTCTATGAATTAGATCGTAAAATCATCCATAGCACAGACTGATAGCCTCACTGGCTATAAAAAATCACGTGGTGTCATTAAAGTAAGTTTTATGGTTTTGGGGAGTTGACAACCAACAGTATATTCTACATAACATATAATCTCACTGA
$HfrA_{KL}138_{b}$	9,028	12	AACTTTATGGCC
$HfrA_{KL}140_{cde}$	5,906	49	ATCTGCACACTCTGCGACTCACCAGCAGAGCTCGCTTTAGACCAAGTTC
$HfrA_{KL}143_{abc}$	5,397	162	AGGTCAAAAAGTTGAGGGTCAAAAGTTTACGTTTTGCACGACTTGTAATCATCTGCTCAGACAGAGTTTGATGTCAATGTTATAGCGCGATCTTGACTATCAGCACAAAAGATAAAATAGCTTTGAGTTACTCGTGTATCATACTATGGACTCCAGGTGAAC
$HfrA_{KL}147$	4,073	21	GGATCAATAAATAAAAAAACA
$HfrA_{KL}148$	3,281	14	AATTAATTTCAAAT
$HfrA_{KL}149_{abc}$	2,989	153	CTTTTAGAAATGGCTAAGTCCAAGGCCGAGGTGAAATTCAGGTCACCGCGTCTAACAAATATGAAAATGTCGCCTCACGAATGGCACGCCTTGTTACTTAACAAAGACTGTCAATGTATAAGATTAATAAGAAACAAAACGCACACGGTGTCA
$HfrA_{KL}152$	2,336	163	TTTGATATCAGCTCTTAAAACATACACGTTTTGGAAATGTCGTAAAAAGCGCTATCTTTGCCCTTCGCTTTAAAATGGCTGTAAATCACCCTTGCAGTAAAAATGTCAAATGTAAAAGCAGGCGAGGTATCACTTGGGGGTCCTGGCTCCATCTCTGTTTAT
$HfrA_{KL}155$	1,459	50	AGAAGGCATATTGCATATGACTACTTTGAAGAAATAACGTGCAGTGCACC
$HfrA_{KL}157_{bc}$	66	66	GGAGGGTTTTATGGAACAGAAAAACGACAACGCGAGAAAAATTAGTATTTTGCACTTCAGAAATTA
$HfrA_{LM}161$	15,251	79	GAATCATTAATCACCTCATTCAATAAATACTTCTTTGTATACTTCAACAGCAAGTCTGAAGGGCGTTCTTTTGGAAAAGT
$HfrA_{LM}166$	13,538	124	GATAAAGGGTCACGCGTCTCCTCTGTGTATGTATTGAAATTTTAATCTCCCTAGTCACGTGACTCATTAAATAATTAATGCAGCCGGTCCGGAATAGATATGTAGTCGTCAGTATTCGCACTAG
$HfrA_{LM}167$	13,097	189	ATGTTCTAGTTCACAGTATGTATGCCTGGATACCACGGTGTAGCTGTGCTGCTTTGTTCAATTTTCCTGTCGTTAATAAATCACCATCTCTTACATGCGCAGGGCTGAGAGGCAACTGTAGGCAGCAGCGTTTGGAAATAAAGAGGCATAAGCTAAATTCCTCTATGATTCATTAATTTGAGCGATTTA
$HfrA_{LM}168$	12,281	18	CAAGAGCAACACAAATAC
$HfrA_{LM}177$	8,118	267	AATTCCATGCATCACTCAATTGGCAGAAAACGTGTAAGCACACATTTCGCATGAGGGTCCCTATAGACACGATTTACGATTCCACTGGAAAAAAAATCACAATCATTACATCGCTGTCAAGTACACTCCTCACCCTGTGCTATACTCACTAAAATCACCTAAGCTTTCTATGACACTGCTTTGCCTTTGTCAGGTAATTATAAACCTTTTCTGGGTTCAATAACTAGCATGGTGTAACGTGTGGGTGTTAAGCAA
$HfrA_{LM}179_{abc}$	7,472	70	AAATAAAAGGATTTAAGTTGACGTATGACCACGTGAGCACATAATACAGCGCATTATTGTGGCATTTGGA
$HfrA_{LM}181_{bcde}$	7,253	85	ATGGATTTTATTTTGAGGTATTTCAAATACTTCCGTTGATTGTTCATTAATATGGTGAGTTATTGCTGTTCAAGGCAAAGGTCAG
$HfrA_{LM}180$	6,306	68	ATATAAAATTGCACCCACCAGTGAGTCCTGCTCAGCAAAGGTCAGGGTGTTGCGATTTGATTTATGTT
$HfrA_{LM}185$	5,757	95	TTTAATTGTAGTTCGGTTACACGTCACTGTTTCTTTCTGATTTATTGTCCAGACTTGGTGCAAATTTGACAAGGCTAAAACAGACATAAATAATA
$HfrA_{LM}188$	4,340	110	GCAATAAGCGATAGGAGCAAGTGTTTTCTCTTTAGTGCTCTGTTTACAGCCTTAGTGCATGAAGTCATAAATCTTACGTAGCCATAAACGACAGAAGCATTGGTATGCAT
$HfrA_{LM}193_{cd}$	3,664	76	AATATTCATAAGAAACATACCCCATTTCTTTGGTATTAGATAAGAGCAGCCAGGAAAGCTGATAGCAGTCTTATCT
$HfrA_{LM}194_{bde}$	3,295	95	GAGTTTATGTCCCAGTGATTTATGACCATATGACTTAAATGTCGGTTCAAGAAGAGTTCACATGCTATAGCTTCCTTTCACATCGTGACCTGATAC
$HfrA_{LM}198_{abc}$	1,285	153	TGACCCCCATACATCAAATTAAGAGCAGCTTCAGAGACAGAGCCTATTATCGTTCGGTTGGAAGAAAGTTCACACCGTGGTCATAAAGAAGAACGCCAGTTTTTCAGGTTCCTGGTGCTTTCCGATCCATCATGCTTTTAGCAGGTAGAATT
$HfrA_{LM}199_{abcd}$	219	213	AATATTGCTGCACACTTTTTATGGCTTTTCATGTCAGTTGCTATTGAGAGTAAGATGGATTTGTACCATTTCGCCATTATGGTTTTGTTTCTTGTAACCCCAGGTCTACCCGAAGAGGCCATTGGAAGAAAAGAGTCACGTGACACAGGGCGCCAATGTTATTCTATAAGGGTGTCAAGACCCTGTCAGTTTGCGAAATAAATATTGGGAAAC
$HfrA_{MN}201$	4,347	288	AACAGGGTATATGAACAAATTTTCTAAAAATTACTTAAACGTGAATTTGTTCGTGCTTTATTTTTCCTTGCGAGGAACAGTAAAATGGTTTATATTGCACTTCTTTACATTGAATGGTTAAACAAGATCAAAGCGCTTTTACAGGGTCTCCCTGGGCAAATAAGATTCTTGTCCATCTTTCAGTATGTGTGTATGCTGGTGAACATTCAAATTTATTTATTTGCTCTTGCAAAGGAGAATATTCTAAGTGTTTA
$HfrA_{MN}202$	3,536	53	TGCTGCATTTATTGAGGCCGAAAAAAGGACCCAGTTCTGCCCCACGCAAGAAA
$HfrA_{MN}210_{b}$	559	78	GTGTCAAAACTTTGAAGATTAATGGATTACTTTGTTAATGACTCTAGGCGTCAGATTTAGGTGCAAGATGATTTGTGA
$HfrA_{MN}211_{bcdef}$	279	178	GAGGCGTTTCTCTCCGACTTTTTGGATCAATCACATAGACAGTGCCTTCTTTTGATTAAACCCCAAATTGTCATTGGGCAGACGCAATCATGTGACAGCGAATTCGGTCCAATTTCAACCTTGTCTCCATGAATTCAATAGTTTAATAGTAGCACGGTCCCCATACGGCTGTAATCAG
$HfrA_{NO}215$	4,296	12	TGCCTTCTTTTA
$HfrA_{NO}218$	3,924	59	AATAAAATTGAGCAGTCTGGTCAGTCTCGTTCTTTACTTTTTTCATATCAGATTTATT
$HfrA_{NO}219$	2,804	31	CTCATAAATCAAATCCCTTTTATGAATGAGA

Xenopus tropicalis western clawed frog

PFC	Dist. To 3' gene	Length	Sequence
$XtrA_{AC}3$	55,355	20	CATTATTGCCATGACAACTG
$XtrA_{AC}4$	54,155	25	TTCAAAAATAAGCACCTGTTTTGAG
$XtrA_{AC}15$	37,794	31	AAATGTTAAACGCATCAAGAATTGTTCTGAG
$XtrA_{AC}23$	23,047	147	TCTCTGTAATAATAAAACAGTACCTTGTAATTGATCCCAACTAAGATATAATTAATCCTTATTGGAGGGAAAACAATCCTACTGGGTTTAATTAATGTTTTATTGATTTTTAGTAGACTTAAGGTATGGTGATCCAAATTATGGAAA
$XtrA_{AC}26$	15,294	20	CAGGTACATGTATGGACTCC
$XtrA_{AC}27$	13,284	24	TGTAAACACTTCTGAGCTTTTAAT
$XtrA_{AC}28$	12,711	44	TAGTTTTTGCCACTGATGTTTAAAAATTTCATTTTCACTTTCTA
$XtrA_{AC}30_{ab}$	1,773	35	TGGCTAGACTGCAGCTTTCAACTTGAACTTGGCCT
$XtrA_{CE}35$	10,652	49	CTGCTCAGGTAAATTATACATATAACTATCTACATTCTACATACATTTT
$XtrA_{CE}39$	10,217	205	TATTAAAATTTGCAGTAATCTTGTTCATTTTTATTCGGTTGAATTCGAAAAACAGCCAGTGGTGTCTAGCATTTTTAATCAAAATAGATATTTGAACAAATCATTATTTTCCTGGCTTATTCACTCAAAAATGGTATCTTGTAAATAATAAATTAGAGCACCGTGAACGTGTTTTTGTGCACAATTAATTCCTATTTTTACTAAA
$XtrA_{CE}43_{ab}$	9,669	99	TTAAACCTTGACCTTGACGAAAAGACACAATTTTTGACCTTGACACTGTCAACTCAAGTATAGGGCATCATTTCTATTAATGGAGCAATATCGCCACCT
$XtrA_{CE}57_{cd}$	2,671	57	AGAATTTGTTTTCCTTACCAGAGGCGGGGCGAGTCCGGCCAATACTGGCGCCTTTGT
$XtrA_{CE}60_{acde}$	175	169	TGTCAATTTCCCCGTGTGATCACATGACCAGCACCTCCCTGCTAAGGATGGGGATAGATTTCCACGTCAGCTTACGTCTTCCAATTTCTACTTCACGGATCTGCTTCAAAGAGGCAGCTGCATTAGGGAATGATGTTAAGCTCAGCTAATGCGGACAGCCCGAGGTAGC
$XtrA_{EF}75_{abcde}$	150	137	GGGAGGGGTTGCTGGGAGCTAGCGAGCTTCCATTGGCTGCGTGCATTTGCCGTGGTGCAGGGGTATCTCTAATCATATTCAGCATGTTTTGCACAAGAAATGTCAGCCAGAAAGGAATATCTGCTCTCTTCGCCAAA
$XtrA_{FG}77$	5,446	67	GACAAGCACACCGCAAGCTCCACTCTTTCCTTGTCTCTTTAAATGTAATCTTTTGCCTAATGTTTTT
$XtrA_{FG}78_{ac}$	4,865	170	ATCTATTTAAATATTACCTAGACAGTCGTAATTTGTCTGGGCCTTATAGCAATGGAGCTGGAAGGGTGATCGGCTTTTGTGTAGTTTTATGACTTGCTCGCATATCTGCTTTGCTGCAACCCTAGACGAGTGGTTTCAGTTGTACAAGGAACTGTTTAGACAGAGGAAGC
$XtrA_{FG}79_{bc}$	4,255	55	AGGTAGTTTTATGTTGTTGGGCATTCACCTTTCTCTCTACAACAAGAAACTGCCT
$XtrA_{FG}92_{abcdef}$	230	204	CATAAAAACATATGGCTCTGCTATAAAAAATGATGACTGCAAAACAGTGACCCATTAATAGCCTGCGGACTGATTTATACCTTATTGTTTCTGCTGCACAGTCACATGGCCCCGGCAGCCAATAGGAATGAAGGCTGCCTTCAACTTATTAGGTGACTGTACTTCTTTGTAGGACCAAGTTGTTACATGAAATCTGCAGTTTCAT
$XtrA_{GI}97$	4,224	23	TACCTCATGGCCTAATGACCAAA
$XtrA_{GI}103_{acd}$	3,038	129	ACGCCCTCTGTTGTTCCAGACAGAAAGAGACTTCAAAGAATGGGCACTAAAAGTGTGTAATAAATGGCTGGTCTGCAAACTGTCTGGAATTCGCTCCTTAATGATTTTATGGCTGTCCTGCAGCAATTA
$XtrA_{GI}105$	2,004	12	TTGAAATTTTAT
$XtrA_{GI}110$	1,237	33	GTTTATGGTGTGCAATATACTGGGGTTGTAAAA
$XtrA_{GI}111_{ab}$	193	179	CACCTGTACGGTGCAATCAGATTGGACGGGGCGGTCAGATGGTGGCAGATCACGTGGCCCAGGCAGGCAGCTCAGTGCAAAGGAAAAGATGGGGTTTTGTGTAAATGTGGGGGTTTAGTGCTGCCATATATCACTGCCGCCTCGTAAAACCGACACCGGAGCCTACCGGACTACAAATC
$XtrA_{IJ}118$	5,451	28	CATAATATATTTTTTTGTTAAACATAAT
$XtrA_{IJ}119_{abcd}$	5,009	192	ATGGCGGCTAGGAAGGACCTGCCTCCTGTAAGTGATGAGGCAATGGCTATAAATCCGTTGTTGTTTATGAAAATTTACAACTTTGCAATACAAGTTTATGAGTTGCTGGGATATTCCATTGGCCGCTGCTGGTCATGTGAATGGGAACCGAGAACATGAAGTTTTTTATCATTCCCCCTGCGAGGATAGAGC
$XtrA_{IJ}122$	4,612	77	CCTGGCTCATTTCCTGCTCTTCCTGGCACCAAACCGGGCTCTATACTTCTCTATGCAACCAGCCACACAGCTGTCAG
$XtrA_{IJ}130_{ab}$	134	110	GATCAGCTGCTCAGATTTGCCTCTGTCGCTTTTGGCTCTCGGCCATCCAGAAACAAACCAGTTGGATGAGTGCTAATAGTTATAGGCAGATGTACTAATACACAACAAAT
$XtrA_{3K}132_{ac}$	307	287	TGGCGCAAGAGGCTGCTCAAAAGCTTCCAGCTCTACACGGGGCCGCGCACAAACGCTCGCTCCCTTAGCAGCCTCCCTGCTCCCCCACTAGTACACCAGTTTACCTCTAGAGGTCATCAGGCAGGATTTACGACTGGACAACAAAAGCACGTGATTCCAAGTCGTACCCCATATTTGGGTGCCTACGTAGGAGGGAACCAAGTACATGTCCCAGTCATTTCCATAATTCATCATAAATTGTGCAAGGGTGCTATA
$XtrA_{KL}136_{abc}$	8,170	209	AAACTTTATAAGCTCTAGTTCCGGCTATGTGACATTTGCTTGCCAAATGAATAGGGTTTTGTCTATGAATTACATCGTAAAATCCTCCATAGCACAGACAGATAGGCTCACTGGCTATAAAAAGTCACGTGGGGCTATTAAAGTAAGTTTTATGGTTTTAGGGAGTAGACGACCAACATTATATAGTACATATCATATAATCTCACTGA
$XtrA_{KL}138_{abc}$	7,296	88	CTGTTTACTGCAACATTCTTTTAAGTCAATAGTGAGTTCACACTGCAGTGTGATTAGTCAGCAATGAACTTTATGGCCAGAATTAACT
$XtrA_{KL}140_{bcde}$	5,144	56	AAGATAAATCTGCAAACCCCAGGAATCACCAGCAGAGCTCGCTTTATACCAAGTTC
$XtrA_{KL}143_{abc}$	4,712	154	AGGTCAAAAAGTTAGAGGTCAAAAGTTTACTTCTCCAAGTCATCTATCCAGGCAAAGTTTGATGTCAATGTTATAGTAGAAATTTTGCCTATCAGCACAAAAGATAAAACGTGTTTGAGTTAGGCAAGCACCATAGACTGGACTCCTGGTGAAC
$XtrA_{KL}149_{abc}$	3,549	153	CTTTTAGAAAAGCTTAAGTCCAAGGGCGAGGTGAACTTCAGGTCACCGAGTCTAACAAATATGAAAATGTCGCCTGGTGAACTGCGAGGCTACTTATTTAACAAAGACTGTCAATGGGGAAGATTAATATGAAACAAAATGTGCGCAGTGTCA
$XtrA_{KL}157_{bc}$	74	74	GGAGGGTTCTGTGCAGAACAGAAAAACGACAACGCGAGAAAAATTAGTATTTTTTTGTTGCACTTGCACAAATTA
$XtrA_{LM}163$	13,031	106	TGCCCTTTGCTAAGGCAACCAGGGTTCACAGAGAGGACACGTTTTCTGTCCCATTGATTGTTTTTTCTGGAAAGATGCAAGCTGACCCTCAAACCCTTGACCTTTA
$XtrA_{LM}174$	7,764	17	CAGTGTAAATCTGGTGA
$XtrA_{LM}179_{ac}$	6,224	70	AAATAAAAGGATTTAAGTTGACGTAGGACCACGTGAGCCCATAACACAACGCATTATTGTTGTATTTGGA
$XtrA_{LM}181_{bcde}$	6,016	76	ATGGATTATATTTTAAGGTATATCCGTTGATTGTTCCTAAAAGGGGTGAGTTATTGCTGTGCACGTCAGAGGTCAG
$XtrA_{LM}188$	4,053	113	GCAATAAGCAGTAAGTAGCAAATGTGTCCTGTTTTAATGCAACGTTTACAGCATTCGTACATAGGGCTATAAATCTTTCCTAGCCATAAATGACAAAACCCATTGGTATGCAT
$XtrA_{LM}193_{cd}$	3,035	76	AATATTTGTAAGAAACATACCCATGTTGTGTGTTGGACTACACTTGGAAAGCCCTGTGAGTTGATAGCTCTTATCT
$XtrA_{LM}194_{bcde}$	2,715	148	TCGTCAGAAACTAAGGTAAGCAGGCCAGAAATCTTCTATCACTTCTCAATGGTGGAGTTTATGTCCCAATGATTTATGGCCATAGTCTTAAATCTCGGTTCAAGCAGAGTTCACAAGCTCAAGCTTCCTTTCAAGAAGTGACTGATAC
$XtrA_{LM}198_{ac}$	1,474	155	TGACCCCCATACATCAAACGACAGAGCAGTTGCAAAGACAGAACCTATTATCGTTGGGACTGCACAGAAAGTTCATGCCGTGGTCATGGAGATGAACGCGGGTTTTCTGGTTCCTGGTGTTATTTCAGATCCATCATGTTTTAACAGGTAGAATT
$XtrA_{LM}199_{acd}$	225	219	AATATTGCTGAAGACTTTTTCCTTTCTGGCTTTCATGTCACTTAGCCATTGTAAGTAAGATGGATTTGTCCTACTTCTTCACTGTACTGTGCTTCTCATCCACCACAGGTCTGACCAGCGAGGCCATTGGAGGAGGAACGCCACGTGACAGAGGGGTGCCAATGTTATTCTTTACGGGTGTCAAGACCCTGTCAGTTTGTGAAATAAATATTGGGAAAC
$XtrA_{MN}201$	4,619	308	AACAGGGTATAGCAACAATTTTTATATGGGGAACAAATTCTTCTTACATGTGAATTTGTTTGCACATTTTCTTAACCCACAGGGAACCCATTTAAATTATTATTGCACTTCTATTCAATTGCTAGCCAAATATTAAAAAAAAACGCTTGAACAGGGACATCCCGATCCATAATTATACGTGTAAATCATTTAATGTGTGTATTCTGCTGCATATACTCCGGTTAATTTATTATATTTTCATGGAAAAAAAAGAAT
$XtrA_{MN}204_{abcd}$	1,800	92	AGAAGCTTTAAATGTGATCTTTAGGGCCAGTAGCTGTCAAGCCATTTGGCAAGCAAGATTGATCACACGCAGACTTCCTCCCAGCTTTGTTT
$XtrA_{MN}210_{bc}$	564	114	GTGTCAAAACTTTGAAGATTAATGGATTACTTTGTTAATGACTTAAGGCGTCAGATTTTGGTCCTTAAATGATTTGTGAGGTGTAAAACGCCTTCCTGACAGACAGAAACAATG
$XtrA_{MN}211_{bcdef}$	303	180	GAGGCGTTCCTTCTTGACTTTTTTTGGATCAATCACACAGACAGTGGCTTCTTTTGATTAAACCCCAAATTGTCATTGGGCAGGAGCAATCATGTGACAACCAATTCGGTCCAATTTCAACCTTGTCTCCATGAATTCAATAGTTTAATAGTAGCACAGTCCCCATACGGCTGTAATCAG

Latimeria menadoensis coelacanth Hox A

PFC	Dist. To 3' gene	Length	Sequence
$LmeA_{AB}2$	71,523	139	TCTGGGAGGGATCACCATTGCCCCTTCTTTCTAAACTGTCATTCTTAATGGTGCAGTCGTCATTACAGTACCACTGGTGACGCCACTCATTGATTGCAAGTGGATAAATAGAAACGTCTGCCACCGTAAAGATGAAAGG
$LmeA_{AB}29$	25,959	245	TCCTGGAACCAAAATGGCCTGTACAGTTCGCCAAACACTCGCTGAGCTGCTGGAGCTGCCTGGGGACCTGCGGCTGCAGCGTCCCTGGGTCTTCCAGGATGAGAGGGGGCGGCGCTGGGCCGCCTATATGGAGAGGGTAAGTACCCAAGATGTCTGCTGTGTGTGTGTGTGTGTGCCTGCTGTGTGTGTGCCTGCTGTGTGTGTGT
$LmeA_{BC}30_{b}$	2,389	24	CAACTTTCAACTTGACCTTGGCCT
$LmeA_{CE}38$	14,071	24	GAATGAACCACCTTTGTGCTGAAA
$LmeA_{CE}57_{abcd}$	3,121	173	AAGCTGAAGTTGACTGCCAAATGTGCTATAAAACCGCAGGTCAGTCTAGAGCTGGTTTGGTTCTCCTGGACGCACTGCATTCAATTATAGCTTAAAAGCCCTAATACAAGTGCAAGAATTTGTCTGCGTTCAAGCAGAGGCGGGGCAAGCCTCTGTGATGGTGGCGCCCTTTGT
$LmeA_{CE}60_{abcde}$	171	168	TGTCAATTTCAGCATCTGATCATGTGACATGCACCTCCCTGAGATGGCTGTAGATGGATCTCCACGTCAGCTTACGTCTCCAAATTTCTACTTCGGGGATCTGCTTCAAAGAGGCAGCAGCATAGAGAATCATGTTGAGCTCAGCTAAAGTGGAGATCCAAAGGTAGC
$LmeA_{EF}62$	6,797	40	TTGAACTTCTGGATTTATACGGACATCACCTATAACCTTG
$LmeA_{EF}69$	4,536	38	CCTACAACAAGCAGTGTCCTCAGGTCATACATTGTTTA
$LmeA_{EF}75_{bcde}$	118	105	ATTGGCTTTAAACTCTTGCTGTGGTATGGGGTATCTCTAATCATATTCAGCATGTTTTGCACAAGAAATGTCAGCCAGAAAGGGCTATCTTCTCCGTTCGCCAAA
$LmeA_{FG}78_{abc}$	4,812	170	ATCTATTTAAATATTACCTAGACAGTCGTAATTTGTCTGGGCCATATAGCAATGGTGCTCTAAGGTTTGTCGGCTTTTGTTCAGTTTTATGACTTGCTAGTATAGCTGGATTGTAGCTGTCTTGGACGAGAGGTTTCAGTTGAGTGAAGAGCTGGAAAGACAGAGGAAGC
$LmeA_{FG}79_{bc}$	4,124	55	AGGTAGTTTCATGTTGTTGGGGCTCCAAATTTATCTCTATAACAAGAAACTGCCT
$LmeA_{FG}85$	3,504	44	GGGCGCCAGCATTCCATTGAGCCTGAATGCACAACAGCGACCTC
$LmeA_{FG}92_{abcef}$	247	213	CATAAAATAATAATAAAAAATATGGCTTCTACATTGAAATTATGACTGCAAAACATTGACCCTCTAATAGACTGCGGACTGATTTATACTTTATTGTTCTGTTTTACAATCACGTGACACCCCTAACCAATTGAATGGTAGGCAGTACTGAACTTAATAGCTGACTCCACTTGTTCACGCGGCCAAGTTGTTACATGAAATCCGCTGTTTCAT
$LmeA_{Gl}97$	7,064	23	TACCTCGTTGCCTATTGACCAAA
$LmeA_{Gl}103_{ace}$	5,638	207	ACGCCCTCTGTTGTTGCAGATAGAAGGAACTTCAAAGACTTAGCCCAGAGACTGCAATAAATGAAGGGTCTGGCATGTTGTCTGGAATTCAGCTTTTAATTAGTTTACAACAGCCTAGCTTCACTTAAGATTTTCTAGAGAAAGCTATTCATTTGTTTTGTCTGACTGTTTCAGACAAAGGGTCCGCGGAAATAACTGTCTTTATTG
$LmeA_{Gl}110$	1,514	33	GTTTATGGTGTGTAATATAGCACGGCAGTAAA
$LmeA_{Gl}111_{ab}$	181	168	CACCTGAAATGCAAGAAGAGATTGGGTGAAACGGTCATATGGTATATCACGTGTATAAAACGGCTCAGACCAAAAGAAAATGGGGTTTGGTGTAAATCTGGGGTGTAATGCTATCATATATCAAGCTACCTCGTAAAACCGACACTGAAGCGTCCTGGCCAACAAATC
$LmeA_{Lj}119_{abcd}$	4,898	190	ATGGCGCCTCTCAAAGAACGACTCTTCACAGGGAAGAGGAAAACGCCATAAATCCGTTGTTGTTTATGAAAATTTACAACTTTGTAATACAAGTTTATGAGTTGTTCAGTTTCTCCATTGGCCGCCGCTGGTCATGTGGATTGTAACCGTGAACATGAACTTTTTATAATTTCCCTTACGAGAATAGAGC
$LmeA_{Lj}122$	4,510	77	CCTGGCTAATTTCCTGCTTCTCTTGGCATGCAGAGGAAAGGTTTCTTTCTGAGTACAACCAACAGGACAGCTGTCAG
$LmeA_{Lj}130_{ab}$	135	110	GATCAGCTGATTGTATTTGCTGTCGTCGCTTTTGGCGTTACGCCATCTAGAAACAAACCATTTCGATGACTACTAATAGTTATAGCTAGATGTACTAATACACAACAAAT
$LmeA_{JK}132_{ac}$	329	293	TGGCGCAAAAGTATAATTCAAAACTAGAACTTTAAGGCGGGGACGCGCCACTAGTACGCGCTTCTTCCCTCCGGCATCAACCCTGCCTTCTCATTTGTGCACGAGTTTACCTCCAGAGGTCATCAGGCAGGATTTACGACTGGACAACAAACGCACGTGATTCAAAGTCGTACCCCATATTTGGGTGCCTACGTAGGAGGGAACCAAGTACATGTCCCAGTCATTTCCATAATTCATCATAAATTGTGCAAGGGT
$LmeA_{KL}133$	16,559	45	AATAGCTCCAAGCTGTCAGGAAATAATTTACTTTTAAACTACCTA
$LmeA_{KL}134$	16,393	72	GTGCCTTTGTATAGCTCTTTGCACGAACTTCAATAAGTGTCTCTTATAAGCGCAGCTTCAGTGATGTATGTT
$LmeA_{KL}136_{abc}$	15,637	207	AAACTTTATTACCCCGAGTTCCGGCTATATGACATTTGGGTGCCAAATGAATAGGGTTTTGCCTTTGGATAAGATCGTAAAATCATCCATAGCAGAGATAGATAGGCTCACTGGCTATAAAATGTCACGTGGTGCCATTAAAGTAAGTTTTATGGTTTTGGGGAGTTGACATCCAACAGTATATGCCACATAACATATAATCACTGA
$LmeA_{KL}137$	15,105	13	AAGGCAAAGCAAA
$LmeA_{KL}140_{bcde}$	9,436	56	AAGATAAATCTGCACACCCTGGGAGCCACCAGCAGAGCTCGCTTTAGACCAAGTTC
$LmeA_{KL}143_{ab}$	8,950	27	AGGTCAGAGTTGAGGGTCAAAAGTTTA
$LmeA_{KL}149_{abc}$	3,004	151	CTTTTAGAAAAAGCTAAGTCCAAGGCCGAGGTGAATTTCAGGTCACTATGTCTAACAAATATGAAAATGTCGCCTGCTGAACGGCACGCCTTGTTATTTAACAAAGACTGTCAATGTGTAAGATTAATAAGAAACAAAGTGCACAGTGTCA
$LmeA_{KL}150$	6,987	31	ACATTTCTGAAGTGTCTTTCTTTTTTTCTCA
$LmeA_{KL}156$	4,887	13	TTCTACATTAGCT
$LmeA_{KL}157_{bc}$	75	69	GGAGGGTTTTACATAGACCAGAAAAACTACAACGCGAGAAAAATTAGTATTTTTGCACTTCACAAATTA
$LmeA_{LM}160$	20,895	254	AAAGTTAAGTACTTGCTGAAGATAGCAATCACTTACAAAGGAATAACGGAATAATAAACTGCACACTCAGAAAAGTTTTAAAGTATTCACATGCACGAATGCACTGTATTTGTTACAAGTGACTTTTAAAAGTATTTTTTCTCTTTCATAATGACCTTTTGGACTTCAATGAGCAGCACAATATCCTTGCAAACTCTTTAAAGCAACTCTTTATGATTCTAAAGAAACTATTTTATTTTCTTGTTCCTAACTAC
$LmeA_{LM}162$	19,007	18	CCAACTCCTGCTTTGATT
$LmeA_{LM}184$	8,692	38	TCCTTTTTTTCCTTCTTTATCGTAGTTTCCTAAAAGAT
$LmeA_{LM}179_{abc}$	6,986	70	AAATAAAAAGGGTTTAAGTTGACGTATGACCACGTGAGCACATAATACAGCGCATTATTGTAGCATTTGGA
$LmeA_{LM}181_{bcde}$	6,749	76	ATGGATTTTATTTTGAGGTATTTCCGTTGATTGTTCATTAATGTGGTGAGTTATTGCTGTTCAAGGCAAAGGTCAG
$LmeA_{LM}188$	4,048	114	GCAATATGCCGTAAGGAGCTAGTGTTTTCTGTTTTAGTGCTCTGTTTACAGCTTTGGTGCATGAAGCCATAAATCTTGCGTAGCCATAAATGACAAAAATGCATTGGTATGCAT
$LmeA_{LM}194_{bcde}$	2,874	149	TCGTCAGAAACTAAGGTAAGCAGGCCAGAAATAGGCAATCAGTTGTGAATGGCGGAGTTTATGTCCTAGTGATTTATAGCCATATGACTAAAATCTCGGTTCAAGAAGAGTTCACAAGCTATAGCTTCCTTTCACGCAGTGACTGATAC
$LmeA_{LM}198_{abc}$	1,238	155	TGACCCCCATACATCAAATTACAGAGCAGTTTCAGAGACAGAACCTATTATCGCTAGGTTGAGAGAAAAGTTCAAGCCGTGGTCATGGAGATGATCGCCAGTTTTTTCAGGTTCCTGGTGTTTTCCCCAATCCATCTTGTTTTAACAGGTAGAATT
$LmeA_{LM}199_{acd}$	228	222	AATATTGATGGCCTTTTTTTTCTTTTATGGCTTTCATGTCACTTAGCTATTGAGACTAAGATGGATTTGTAGTTCTATTTCTCCACTATAGTAAGCGTCTTGTAATTCCAGGTCTAGCCGAAGAGGCCATTGGAGGAAAGACGTCACGTGAAAGGGGGGTGCCAATGTTATTCTATAAGGGTGTCAAGACCCTGTCAGCTTGTGAAATAAATATTGGGAAAC
$LmeA_{MN}201$	4,657	284	AACAGGGTATATGAACACATTTTAAAGGTTATATAAAATGTGAATTTCTTCGAACTTTATTTATCCCATGAAACCAATTAATTTGATTTTATTGCACTTCTCTGGTTTTAAATGGACAGAACAAACTGAAAGGCGCTTGAACAGGGTCCCTGGACAATTATGGATATTTGTACGTCTTTCATTGTGTGTATGCTGGTGAACATTCAGATTTATTTATTTGTCCTTGCATAAATAGAATATTCTAAGTGTTTAAAGT
$LmeA_{MN}210_{abc}$	567	161	TTCTCCTGGTGTGTTGGTTCTTATGTGGCTAGGATGTGGGGGTCTGGGTGTCAAAACTTTGAAGATTAATGGATTACTTTGTTAATGACTCAAGGCGTCAGATTTAGGTGCTTAAATGATTTGTGAGGTGTAAAGCGTCTTCCCGACAGTCAGAAACAATG
$LmeA_{MN}211_{abcdef}$	306	209	TGTAAATAAAGCACTGCTGGTCCAGAATGAGGCGTTCCTTCCTGACTTTTTTTTGGATCAATCACACAGACAGTGGCTTCTTTTGATTAAACCCCAAATTGTCATTGGGCAGAGGCAATCATGTGACAACCAATTCGGTCCAATTTCAACCTTGTCTCCATGAATTCAATAGTTTAATAGTAGCACGGTCCCCATACGGCTGTAATCAG
$Lme_{NO}219$	3,061	32	CTCATAAATCAAATTGCTTTTTATGAATGAGA

Homo sapiens human Hox A

PFC	Dist to 3' gene	Length	Sequence
$HsaA_{AC}2$*	42,727	139	TCTGGGAGGAAGCCCCATTGCCCTCTTCTTTCTAAGCTGTCATC CTCCTGCTGCAATCGTCATTACAGTACCGCTGGTGACGCCACT CGGCGAGCGCAAGTGGATAAATAGAAACGTCTGCCACAGCGA AGATGAAAGG
$HsaA_{AC}3$	42,243	14	CATTATTACAACTG
$HsaA_{AC}4$	41,350	21	TTCAAACTGACCCTTTTTGAG
$HsaA_{AC}5$	36,891	227	TTATTGAGAAGAGATTGGAGTGGGGTCAGGGGAGGATATGTTT TCAGGATCTTTTGTTTGTTTTTAATAGATTTTGTCCAGGCACAA ATCCTTTCCAAGCACTGGGCTATCCAAGGGCCAGACAAACCT GACCACCTTGCTCCCCCTACTCCCAGTCCCCTGCCATGGGGCA CCATCTCCACCCTCTTCCACCCCTGCACTGTAAGAGGGAACCT GCCCACTGCAGC
$HsaA_{AC}6$	32,134	26	TTTTGTAAAGAATTCTACAGAATGAA
$HsaA_{AC}8$	30,093	45	GAAAAAGAAAAATCCCATGTGCTCTGGCTCTAAGTCTTATAAC AA
$HsaA_{AC}9$	29,255	20	ATAAATAACTACAGAGCAAA
$HsaA_{AC}10$	29,122	32	TCCAGGATTTTGTGTCCTTTTTGCAAAAGACA
$HsaA_{AC}11$	28,926	20	TACTTAAATTAGACTGCTTG
$HsaA_{AC}14$	28,729	18	ACTGAAAGGAAGAGGAAT
$HsaA_{AC}15$	27,827	28	AAATGTAAGTTTCCCAAGTCTTTCTGAG
$HsaA_{AC}16$	27,563	26	AATAAAACGTTTTTCTGGAGATAGAA
$HsaA_{AC}22$	23,227	16	TATTATTCACTCTTTT
$HsaA_{AC}23$	15,760	106	TCTCTGGCTAATAGAAAACAGAAAGCCAAAGAATGAGCACAA GACTTCTAGGAAAAGATTTAAATGAGGCTCCAATTTTCCTAGG CAGGCAAACCAAATCTGGAAA
$HsaA_{AC}24$	10,752	54	GATTAAAGTTTAATCCGAGGTGTGTGCTCAGACTTGCCATGTTA TTTAAACACAT
$HsaA_{AC}25$	10,451	13	CATCTAGTGGGAA
$HsaA_{AC}26$	8,804	16	CAGGTAGGAAGACTCC
$HsaA_{AC}27$	7,466	25	TGTAAAAGCATCCCAGATATTTAAT
$HsaA_{AC}28$	7,033	35	TAGTTTAATCATTAGAAATTTCTTTTGTGTTTCTA
$HsaA_{AC}29$	5,150	193	TCCTGGGTGGCACGGGCCCCTGCAGTCCAGCAAGCCGGTGTCC CCCGCGGGCTGGCCGCGCACTGCGAGAAGGCGTTTCCTCCTTC CAGCCGAGCCCTTCACCAAGCTCCCTCTGCTTCTAACGCATGC TCCGGTCCTTTCCGCGAAGGCTGTTTCCAGCTCCAGTGAAGAG CGATCAGTGGCACTGTCGGGG
$HsaA_{AC}30_{bc}$*	2,062	31	CAACTTTCAACTTGACCTTGGCCTCCAGCCG
$HsaA_{CE}35$	12,608	59	TCTGCAAAACATTTT
$HsaA_{CE}36$	12,393	26	CCGGGTCGGTAATGTCTTTTTAAGAA
$HsaA_{CE}37$	12,360	20	TAATTGCTTATAACAAGCAT
$HsaA_{CE}38$	12,082	24	GAATGAAAATTCATTCCCCTGAAA
$HsaA_{CE}39$	12,027	218	TATTAAGGGATTAGTATTTTTCCATGTTTATTGTGTTATCAGAGT GCATTAGAAAGATTAGTGATTCATCTTCACAGCACATTTTTAAT CAAGCAGTTATTTCAACCAGCACATTCGTTTTGTTCATATTCAC TATAGAATGATATCTTGTAAATAAAGACATTCAGCACACTGTG AAAATGTATTTGTGCACCTGCTTTTTAAATATTTCTACTAAA
$HsaA_{CE}40$	11,961	31	CATCTTCACAGCACATTTTTAATCAAGCAGT
$HsaA_{CE}42$	11,632	23	ATTATTATTCAGTACTTCTTTTG
$HsaA_{CE}43_{ab}$	11,330	101	TTAAACCTTGACCTTGACGAAAAAAAAGACAATTTGTGACCTT GACTTTTGACAGCTCATGAATTGGCCTTAGCTGGATTAGTAGAT CAAGGGCGCCACCT
$HsaA_{CE}47$	10,585	80	ACCACAGGATTGGATACAGAACGAGAGTTATCCTGGATAACT CAGAGCTGAGTACTGCTCCAGGGTGGTGTGCAATCTTA
$HsaA_{CE}48$	10,227	26	TAATTTTATCTTTACGTTAGAGGAAA
$HsaA_{CE}49$	10,007	21	AGAAATTCAATTATATTTTGC
$HsaA_{CE}51$	9,807	23	TGTGTGTTTAAGAACAGCTCCAA
$HsaA_{CE}52$*	9,292	24	GATTAAAGTATTTTTTTCTAAAAA
$HsaA_{CE}53$	9,037	35	TTAAAACAACCTAGTTGAAATCTTTTCTTAAAAAT
$HsaA_{CE}54$	4,355	40	GCCTGTCAAGTGACGATCCTTGCAGCTCACGTCCCTTGAA
$HsaA_{CE}57_{abcd}$*	2,744	175	AAGCTGAAGTTGACTGCCAAAACTGCTATAAAACCGCAGGTC TGTCTAGAACTGTTTTCGGTTTTCCTAGACGCCCCCGGGTTCAA TTATGGTCTTAAAAGCCCTAATACAAGTGCAAGAATTTGTTTG CCTTCACCCAGGGGCGGGGCGAGCCCCGCAATGGTGGCGCCT TTGT
$HsaA_{CE}59$	2,045	21	GTGGTGGGGGGAGGTGGAGAG
$HsaA_{CE}60_{abcde}$*	175	169	TGTCAATTTCAACATCGGGTCACATGACCAGCACCTCCCTGCT AAGGATGGGGATAGATTTCCACGTCAGCTTACGTCTCCAAATT TCTACTTCACGGATCCGCTTCAAAGAGGCAGCTGCAGTGGAGA ATCATGTTAAGCTCGGCTACTGCGGAGAGCCCAAGGTAGC
$HsaA_{EF}62$	6,367	41	TTGAACTTCTGCCTTAAAATTGGACATCACCCATAACCTTG
$HsaA_{EF}63$	5,802	35	GATTTCTTAGATGTAAAAATGAGATCTCAATAGCA
$HsaA_{EF}69$	4,319	47	CCTACAATGCCAGAGCCGCCTCCAGTTCTAAGGCCAGAGGCT GTTTA
$HsaA_{EF}70$	3,537	16	AGAAATTTAGTAGCTG
$HsaA_{EF}71$	3,041	20	TTTATTACAAGGGGGAAAAA
$HsaA_{EF}72$	2,150	65	CTCTGAAGCTTTTCTGCACTCAGGTTCGTCCGTCTCATGGGCCT AGAGGGTAGAGAATTTGCCTT
$HsaA_{EF}73$	1,158	20	GGGAGGGGAAGAGGGGAGGG
$HsaA_{EF}75_{bcde}$*	119	106	ATTGGCTGCGGGCGCCCGCCGCGGTGCGGGGGGATTGCTAATC GTATTCAGCATGTTTTGCACAAGAAATGTCAGCCAGAAAGGGC TATCTGCTCCCTTCGCCAAA GACAAGCACACCACAATTCTCCCTATCTTGTGAAGTTGTTTTTT
$HsaA_{FG}77$	5,580	70	TAAATCGCCTTGAACAAAAAGTTTTT
$HsaA_{FG}78_{abc}$*	5,037	170	ATCTATTTAAATATTACCTAGACGGTCGTAATTTGTCTGGGCCC TATAGCCCTGGTGCCGTAAGGTTTGTCGGCTTTTGTTCAGTTTT ATGGCTTGCTAGTATATCTGGATTGTGGCTGTCTTGGACCAGTG ATTTCAGTTGAGAGGGGAGCTACATAGACAGAGGAAGC
$HsaA_{FG}79_{bc}$	4,091	54	AGGTAGTTTCCTGTTGTTGGGATCCACCTTTCTCTCGACAGCAC GACACTGCCT
$HsaA_{FG}85$	3,758	58	GGGCGCCCGCTTCCATCCTGAGCCTCCCACTGGAGGCCTGCGC CTGCCCAGGGACCTC
$HsaA_{FG}86$	3,176	14	CAAGGCCTCGGTTA
$HsaA_{FG}88$	2,843	66	ATACAGCACCTTTGTGCAGCTGGGTGCTGGAGAAGGGGACCCT TCCAAGAGCCAGTATGTTTTAAA
$HsaA_{FG}89$	2,226	16	GTAACTCTCTTATTTT
$HsaA_{FG}90$	1,921	28	CAGGACGGGGGCATTTCGGAGTTCATTG
$HsaA_{FG}91$	461	80	TTTTATGTGAGGGGCTCCGCTGGCCGCACTCGCACGCGGGACC CGCGGCCTTCTTGATGGCGTGATTAATTGTGATATAAAA
$HsaA_{FG}92_{abcdef}$*	252	207	CATAAAAACATATGGCTTTTGCTATAAAAATTATGACTGCAAA ACATCGGACCATTAATAGCGTGCGGAGTGATTTACGCGTTATT GTTCTGCTGGACGGGCACGTGACGCGCACGGCCAATGGGGGC GCGGGCGCCGGCAACTTATTAGGTGACTGTACTTCCCCCCCGG TGCCACCAAGTTGTTACATGAAATCTGCAGTTTCAT
$HsaA_{GI}96$	6,293	16	CTATATAGTCCTGTCT
$HsaA_{GI}97$	5,974	23	TACCTCAAGGCCTACTGACCAAA
$HsaA_{GI}98_b$*	5,246	37	ATAAAAGAAAATGAGGGCTGTTACCGTTTATGGGGTG
$HsaA_{GI}103_{acde}$*	4,970	211	ACGCCCTCTGTTGTTGCAGACAGAAGGAACTTCAAAGAATAG GCAGTAAGAGTGTGCCATAAAGGCCGGGTCTGCGAACTGTCTG GAATTCGTCCCTTAATGAGTTTACAACTGTCCAGCCCCAATTA GGATATTCCACCAAAGCCCTTTCATTTGTTCATTTGTTCTGTCT GCCGCCGATAAAGCGGCTGCGGAAACACCTCTTTTATTG
$HsaA_{GI}104$	3,880	32	TTTAATGTGATGTTTAATATAACAATCACACC
$HsaA_{GI}105$	2,428	18	TTGAAAGCCTGATTTTAT
$HsaA_{GI}106$	2,014	43	CCTTGGTCTTTTCTCTTTGCGTCTTGCTGTGTGTCACCCCAGG
$HsaA_{GI}107$	1,896	20	TAACTTTCTTGGGTTTTTAT
$HsaA_{GI}110$	1,448	33	GTTTATGGTGTGTAATCGAGCTGAGCAGTAAAA
$HsaA_{GI}111_{ab}$*	182	169	CACCTGTGAGGACTGCTGAGATTGGCGGAGGCGGTCATGTGGG CGGTCACGTGCTGCGGCGAGCTCCGTCCAAAAGAAAATGGGG TTTGGTGTAAATCTGGGGGTGTAATGTTATCATATATCACTCTA CCTCGTAAAAACCGACACTGAAAGCTGCCGGACAACAAATC
$HsaA_{IJ}114$	6,864	52	CAAAATACTACCTAGCACAGGCCTCTGCTCGAGGCACCCCCA AACTACCTAT
$HsaA_{IJ}117$	5,835	18	TGTCATGTATCTTTTTAA
$HsaA_{IJ}118$	5,711	36	CATAATTGGGTGCATGTGTTTTTGTGTGTCCATAAT
$HsaA_{IJ}119_{abcd}$	5,027	192	ATGGCGCCGGTGTAAGAGCTGCCTCCTCTCAGTGATGGGGAAA GGGTCATAAATCCGTTGTTGTTTATGAAAATTTACAACTTTGCA ATACAACTTTATGAGTTGTTCGGCCCTTCCATTGGCCGCTGTCG GTCATGTGGATGAGAACCGTGAACATGAACTTTTTTATAATTTC CCTTGCGAGAATAGAGC
$HsaA_{IJ}122$	4,651	79	CCTGGCTAATTTCCTGCGTCCTCTGGCACCAGGAAAGAGGACA ATTCTTCTCTCGGGCTGCCCAAGCGACAGCTGTCAG
$HsaA_{IJ}123$	3,943	24	GAAAAACTTACTAATTCAGCCAGA
$HsaA_{IJ}124$	3,822	29	TGTTGGAGGAGAGGATGGGATTAATTGGG
$HsaA_{IJ}125$	3,428	44	TTCCCCAACTTGAGACAATATCTTGGTATGGATTCGATAAATA G
$HsaA_{IJ}126$	2,663	14	CCCTCCAGGCTGGC
$HsaA_{IJ}127$	2,401	17	GGCTAACATGTAAAGAT
$HsaA_{IJ}129$	2,016	55	TTTTAAAAGTCCCTTTTAATTTTTTGACTCGAGTTTTCATTTTCA CCTATTGAAT
$HsaA_{JK}132_{ac}$*	312	292	TGGCGCTCGAGTCCGGCTGAACGGCGGCAACTGGCGGCGGGC ACGCGCCCGGGGGCGCGCGCGCCACCCCCCTCGCCTCCACCCA ACTCCCCTATTAGTGCACGAGTTTACCTCTAGAGGTCATCAGG CAGGATTTACGACTGGACAACAAAAGCACGTGATTCGAAGTC GTACCCCATATTTGGGTGCCTACGTAGGAGGGAACCAAGTACA TGTCCCAGTCATTTCCATAATTCATCATAAATTGTGCAAGGGT
$HsaA_{KL}133$	10,536	71	GTGCCTTTATAGGACCCTTTGCACGAACTCTGGAAGTGGCTCTT ATAAGCGCAGCTTCAGTGATGTATGTT
$HsaA_{KL}134$	10,696	43	AATAGCTCCAAGCTGTTAAAGATATTTTTATTCAAACTACCTA
$HsaA_{KL}136_{ab}$*	9,690	201	AAACTTTATAAGCCCCAGTTCCGGCTATATGACATTTGGGTGC CAAATGAATAGGGTTTTGTCTATGAATTAGATCGTAAAATCAT CCATAGCACAGACAGATCGGCTCACTGGCTATAAAACGTCAC GTGGGGCCATTAAAGTAAGTTTTATGGTTTTGGGGAGTTGACAT CCAACATTATATACCACATAACATATAAT
$HsaA_{KL}137$	9,235	13	AAGGCAGAGCAAA
$HsaA_{KL}138_{abc}$	8,556	90	CTGTTTGCCTGCATGTTCTTTTTATTTCGGCGGGACTGTTCTCAG CCGCTTGTGAGTCTGAAGCAATAAACTTTATGGCCGGATTTAA CT
$HsaA_{KL}139$	7,803	35	CAAATTCTCCTACTGCATTCAGAAATGAAACAAAA
$HsaA_{KL}140_{bcde}$*	5,879	56	AAGATAAATCTGCACACCCTAGGAATCGCCAGCAGAGCACGC TTTAGTACAAGTTC
$HsaA_{KL}143_{abc}$*	5,424	163	AGGTCAAAAAGTTGAGGGTCAAAAGTTTACATTGTGCAGTGCT CTTAGTCATCTGCCCAGACAGAGTTTGATGTCAATGTTAGAGC TGTGATCTTGACTATCAGCACAAAAGATAAAATGGCTCAGAGT GACGCGTGTATCACGGTATGGACTCCAGGTGAAC
$HsaA_{KL}145$	4,340	24	ACTGAAGAGCTAAACAATATTTTA
$HsaA_{KL}146$	4,292	30	TAAGTGTTGGTGCTGTAAGAATCTATTTTC
$HsaA_{KL}147$	4,222	19	GGATCAGTATAGGAAAACA
$HsaA_{KL}148$	3,518	16	AATTAATTGGTCAAAT
$HsaA_{KL}149_{abc}$*	3,214	152	CTTTTAGAAGCGCAAAGTCCAAGGCCGAGGTGAACTTCAGGTC AGTGCGTCTAACAAATATGAAAATGTCGGCTAGTGAAGGGCG CGCCTTGTGATTTAACAAAGACTGTCAATGTGTAAGATTAATA AGAAACAAAATGCACACGGTGTCA

$HsaA_{KL}150$	2,707	28	ACATTTCAGGGCATCCCATCTTTTCTCA
$HsaA_{KL}151$	2,648	46	TTATTAATTGAGTTAGTTTCTTCCTGGCTTTACTTTGGTGTCTTG G
$HsaA_{KL}152$	2,551	159	TTTGATTCAATGGAATGATGCTATGGTTTGCAAAGGGAGATTG AGGAAGGTTCAGAAAAGGTACTGAGATTGTTTATTACAGCCAT AAATCTTGCAGTAAAATGTCAAAATGTCGGTGTTGAGATAACA CTTGGTGGTCCTGGCTCCGTTTGTGTTTAT
$HsaA_{KL}153$	1,927	26	TAAAGATTACACTCAGTGTTGATTTC
$HsaA_{KL}154$	1,756	18	CCTTTTTAACCTCTTTTC
$HsaA_{KL}155$	1,536	46	AGAAGGCTGTTTATACTGCTCTTTGAAAAATCAACTAAGGTGC ACC
$HsaA_{KL}156$	1,026	14	TTCTACCCTTAGCT
$HsaA_{KL}157_{bc}$*	71	71	GGAGGGGCCTCGGCAAGCCCAGAAAAAACGACAACGCGAGA AAAATTAGTATTTTTGCACTTCACAAATTA
$HsaA_{LM}160$	17,551	223	AAAGTTGAAACCATGCAGTGCTGTGGGCCCAGGAGTTTCTGCA GCTGTTCTACACACTGAACTCTGAGCACACTTGGAAAGAAGTG GTCTCTTTGGTATTTGGTATTTGTTGCACCTTCACACTTAATCTG CCTGTTTCCCAGGAGATAAAAAATAGTTTGTGCAAAGGGTTGA GATTAGTTTTTGTTTCCTGGCTGTTCTTAAGGTCCTCTTCACTAA CTAC
$HsaA_{LM}161$	16,478	77	GAATCATTAATCACCTCAAACAATAAATACTTCTTTGTAATTC AACAGCAATTCTGAAAGGCATTCTCTGGGAAAAGT
$HsaA_{LM}162$	16,053	15	CCAACTGTATTGATT
$HsaA_{LM}163$	15,877	113	TGCCCTTTGCTAGGGCAACCATAAGAGTTCACCGAGAGGACA AATTTTCTATCTCATTAATTGTTTTTTTTTTTTAAGCAAACCCTA CTGGCCCCTCAAATCCTTGACCTTTA
$HsaA_{LM}164$	15,767	35	TTAAAGACAGTATTTTGTGTAAGCTCTATAATTTT
$HsaA_{LM}166$	14,125	121	GATAAAGGTGTAGCGCCTGTGCAAGGGAGATATTGAAATTGTA ATCTTTGAGGTCACGTGACTCATTAAATAATTAATGCAGATCG TCAGGAATGGATCGGAGTCGTCAGGGCCTCACTAG
$HsaA_{LM}167$	13,715	220	ATGTTCTGATTCACGCTGGGGAAGGCTGCAGAGATACCACAGG ACGGGCGCGCGGCTTTGTTCAATTTTCCCGGCGTTCATAAATC ACCCGCGCCGGGCGAGCGAGGGAGCAAGCGAGCGCCAAAAA CGCGGAGAGAGAGGCCACGGCGGCGGCGGCAGCCATTGTAAA GTGAGAGACCTGGGCAGCATCTCTCTGTGACTCATTAGTCTGA ACGATTTA
$HsaA_{LM}168$	12,807	20	CAAGAGAGGGTTGCAAATAC
$HsaA_{LM}169$	12,405	19	ATTCACATTCTCCAAGTTG
$HsaA_{LM}170$	11,622	26	CAAACAAACCACTACACTGAACTTCA
$HsaA_{LM}173$	11,280	16	TCCTATTTGAAACAAG
$HsaA_{LM}174$	9,462	19	CAGTGTGAATGATTGGTGA
$HsaA_{LM}175$	8,495	25	CAATTAGATCATTTCCTCTGATATT
$HsaA_{LM}176$	8,462	50	TTAGAAACTTATAATTCCATGTATCACTCAGTTGTGCACTGAG GAAATGC
$HsaA_{LM}177$	8,450	258	AATTCCATGTATCACTCAGTTGTGCACTGAGGAAATGCTTAAG TGCAAGTTTAGTCAGGGCACTAGGTTTCTCTAGATTTTGCAAA ACCCGGACCCACCACTCTATAGCCAAGCACATCTTTCCTAGTG GGCTCCCACTGAACAAAACCAACTGGACTTGGAGGAGATGCT TTAACCTTTTGTCAGGTAATTATGTGCCTTTTTGGTGTTTCTTGC GTCTGTAGTGTATGGTGTGGGTGTTAGCAAACTCCATAC
$HsaA_{LM}178$	8,391	23	GCACTAGGTTTCTCTAGATTTTG
$HsaA_{LM}179_{abc}$*	7,797	70	AAATAAAAGGATTTAAGTTGACGTATGACCACGTGAGCACAT AATACAGCGCATTATTGTGGCATTTGGA
$HsaA_{LM}181_{ce}$*	7,515	68	TATTTTAAGGTATTTCCGCCGATTGTTCATATCTAGAGTGAGTT ATGGCTGTGAGAGACAAAGGTCAG
$HsaA_{LM}182$	7,025	71	GGAAGAGGGAAATCGATGAGGGAAATGTGCAGATGCGCTGCC ATTTTATTGATGATGCTTCCAGGCACTTT
$HsaA_{LM}184$	7,403	40	TCCTTTGTTTCTCTTTTAGAACCGAAACCATTTGAAAGAT
$HsaA_{LM}185$	6,153	93	TTTAATTGCTGTTCAATTACTTGTCAGCATTTCATATGATTTATG ACCTAGTTCTGGCGCATTTTTGACCCAGTTAAAGACTCATAAA TAATA
$HsaA_{LM}186$	5,530	13	CGGGGATTTGGGG
$HsaA_{LM}187$	5,051	34	GTTTTGCCCGGCAGGGTCTTGGCGCCAGTGGAAA
$HsaA_{LM}188$	4,714	112	GCAATATGCCATAAGGAGCAAGTGTTTGCTGTTTTGTGCTCTGT TTACAGCTTTGGGGCGCCGAGTCATAAATCTTGCCCAGCCATA AATGACAAAAACCATTGGTATGCAT
$HsaA_{LM}190$	4,555	13	GGCTTTGTTTTTG
$HsaA_{LM}191$	4,284	50	CTCAGCGTTACTCCATCCCACTAATGAGGAAAATATGTATATA CATATAT
$HsaA_{LM}192$	3,609	28	GCATAATTTTCCTGGACTTTCGTGACGC
$HsaA_{LM}193_{cd}$	3,373	74	AATATTCATAAGAAACATACCCAAGTCGGTGCCACTAGCCCA GGCAGAGCCCGGCGCCCGCACTAGCGCTTATCT
$HsaA_{LM}194_{bcde}$*	2,999	151	TCGTCAGAAGTTAAGGTAAGCAGGGCCGCAACCGGCCGCTCC CGGCGCTGAATGGCGGAGTTTACGTCTCGGTGATTTATGGCTG CAGACTTAAATCTCGGTTCAAGAAGAGTTCACAAGCCGGAGC TTCCTTCCCGGCAGTGACTGATAC
$HsaA_{LM}198_{abc}$*	1,245	166	TGACCCCCATACATCAAATTACATAGCAGTTTCTAAGACAGAA CCTATTATCGTTAAGTTGGAAGGAGAGTTCAAACTGTGGTCAT GGAGATGAACGCTGGTTTTTCAGGTTCCTGGAGGGGTTTTTTTT CCCCCTCTAATCCATCATGTTTTAACAGGTAGAATT
$HsaA_{LM}199_{ac}$*	227	197	AATATTGCTGACCACTTTTTCTTTTATGGCTTTCATGTCACTTAG CTATTGAGAGTAAGATGGATTTGCGCGGAGCCCTCACCGCGCT GCGCTTCTCGCCCCACCAGGTCTGCGCGGGGCGGCCATTGGCG GCGGAGTGTCACGTGACCGCGGGGGCGTGCCAATGTGCGCCCT CACGGGTGTCAAACCCCTGTCAG
$HsaA_{MN}201$*	5,116	304	AACAGGGTATATGAACAAATTTTCTAGTCGAGTTTTCAATGTG AATTTGTTCTTACATTATGGCTCCCGAGGGGAAGCGATTACTTT TTTTAATTTTAAATTTTTTTTTTTAATTGCACTTCTTGTAAAGAGT GAGAAAAAAAATCAAAGGCGCTTTGAAACAGGGGCTCTCTGT GCAAGGATGACTAAGTGTACGTCTTTCCGTGTGTGTATGCTGGT GAACAGTCAGATTTATTTATATTTTTTTTGCAAGCATT
$HsaA_{MN}202$	4,273	56	TGCTGCGGAGGCAGGCTGAGGGCGCAGGGGCTGCCGAGTGCT GTGCACGGAAGAAA
$HsaA_{MN}203$	3,403	40	CAGAGACAGAAGTGGAAAATGTCGCCATTTTGTTTGCAAT
$HsaA_{MN}204_{bcd}$*	2,238	84	AAATGTGTTCTTAAGGGCTAGAAGCTGTCAAGGCTTTTGGTGA GCAAGATTGATCGCGCCCAGACTTCCTTTGGAGCTTTGTTT
$HsaA_{MN}208$	1,377	42	ACAGACTTCCTGGAAATCGGAAATACTCACCGCACCCGAGCC
$HsaA_{MN}209$	1,752	16	TTAAAAATTATTTTTT
$HsaA_{MN}210_{abc}$	726	162	TTCTCCTGCTTTAACAGAACTTATGTGGCTGGGACGCAGGGCC CTCGGGTGTCAAAACTTTGAAGATTAATGGATTACTTTGTTAAT GACTGCAGGCGTCAGACTGAGGTGCTTAAATGATTTGTGAGGT GCGAGGCGTCTTCCCGACAGTCCCAAACAATG
$HsaA_{MN}211_{abcdef}$*	351	207	TGTAAATAAAGCCTCGCTGGCCCCCAATGAGGCGTTCCTTCCC GACTTTTTTGGATCAATCAAACAGACAGTGGCTTCTTTTGATTA AAGCCCAAATTGTCATTGGGCAGAAGCAATCATGTGACAGCC AATTCGGTCCAATTTCAACCTTGTCTCCATGAATTCAATAGTTT AATAGTAGCGCGGTCCCCATACGGCTGTAATCAG
$HsaA_{NO}215$	4,771	12	TGCCTTCTTTTA
$HsaA_{NO}218$	4,385	51	AATAAAATTCGTCACAATTTATCCTCTTTTTTCAATTTTAATAC ATTTATT
$HsaA_{NO}219$	3,195	31	CTCATAAATCAAACGCTTTCTATGAATGAGA
$HsaA_{NO}220$	1,579	22	TAAACAACAATTTAACGACCTC
$HsaA_{NO}221$	826	32	GAAGTTTCTTTTCTTGCCCTCGTCTCCTTCAC

Polypterus senegalus bichir Hox A

PFC	Dist to 3' gene	Length	Sequence
$PseA_{AC}2$	100,256	140	TCTGGGAGGGATCACCATTGCCCTTCTCTTTCTAAGCTGTCAT TCTTAATGGTGCGCTCGTCATTACAGTACTGCTGGTGACGCCA CTCATTGATTTCAAGTGGATAAACAGAAACAGCTACTACCGA GAAGATGAAAGG
$PseA_{AC}1$	97,071	16	CTGTCCTTTTGCGCCC
$PseA_{AC}6$	64,905	27	TTTTGTAAGGAAGCTGTAAGGAATGAA
$PseA_{AC}7$	55,564	39	TTTAAACAAAAAACATTTAAAATGTGCAGGACATATAGA
$PseA_{AC}12$	4,958	41	ATATTTTCTATTGTTTGCGATGTCATGATACTAGAATTTAA
$PseA_{AC}13$	4,536	12	ATGGATAAATGT
$PseA_{AC}17$	3,796	23	TAAATTCAAAGAACAATTCCCAT
$PseA_{AC}19$	2,233	155	CAATAAACAGTTGGTAAAACTTGTGCTTAAATATTGCAAATA CAAATCTCAAACGTATCTTAGCATACTATATTTTTGCTGTAAG CAATTCTTCATGTTGAGCTAACATAAACAGTTATTCCAATTTA ATTTCAAAAGTATAATCACGGCAATGA
$PseA_{AC}22$	55,617	16	TATTATATTATCTTTT
$PseA_{AC}25$	6,478	13	CATCTAGTGGGAA
$PseA_{AC}30_{ab}$	1,561	35	TGGCTAGACTGCAACTTTCAACTTGACCTTGGCCT
$PseA_{AC}31_{abcdef}$	153	147	CTTTGCATATCACGTGGCTTGCCTCAGCCAATGATTTCTTCTC CCTAATGTACACAGGCGCGTCAGTGCTGGTAAGTCCCTGAGA TTCAGTATCTCTTTTTTCAGTCTTGAGGCTTTTTAAAAAGAGC CAGAAGTCTCAATGTTGGA
$PseA_{CE}43_{ab}$	8,641	100	TTAAACCTTGACCTTGAATGAAAAAAGAGCCTTTGACCTTGA CTGTAATAGTAGTTAATGAATAGACATAGCAGGGATTGTTGA AGCAAGCTTGCCACCT
$PseA_{CE}44$	4,335	18	AATAAGAAATGAAATGAA
$PseA_{CE}45$	3,752	26	AGTGTATTGAAGTAACAAAAATGTAA
$PseA_{CE}55$	2,656	24	ATAAACAACTATGAAATTATTAAA
$PseA_{CE}56$	2,400	41	TTTTTAAGAAATGATTAATTAATTGTATTTTATTTCATTTT
$PseA_{CE}57_{abc}$	1,892	144	AAGCTGAAGTTGATTGCCAAAAGTGCAATAAAACCGCAGGT CAGTCCAGTCCTGTTTTTTGGGTTTTTGTTGCAGCCCTGTGTTCAA TTATGGCTTGAAAGCAAGAATACAAGTGCAAGAATGTGTTTG CAGTCACAAAGAGGCGG
$PseA_{CE}60_{abd}$	176	113	TGTCAATTTCTGCATCTGGTCACATGACCCGCACCTCCCTGTA ATGGATGGAGATGGATCTCCACGTCAGCTTACGTCTCCAAAT TTTTACTCCGCGGATCTGCTTCAAAGAG
$PseA_{EF}62$	3,875	40	TTGAACTTCAATATTCCTATGAACATCACCTATAACCTTG
$PseA_{EF}63$	5,604	25	GATTTCTCCAACTACAGAATGAGAT
$PseA_{EF}64$	4,293	31	ATGGAAACACATGTATCAATGTAAAATATTG
$PseA_{EF}65_{bc}$	3,532	116	CCTAATTACAACACATCCTCCGTGTTGCTGTGGCAACTTGGTC ATAAAAACTGTCTGAGTCTGGTGCATTTGTAGAATTGGAGAG CCTGCAATAAAATGTCTGAGACCAAGGTTAT
$PseA_{EF}66$	3,271	31	CCACTGGGTGGCAGTCTTTAGATCAATGTCA
$PseA_{EF}67$	1,257	47	GCACAGTGAAAAGCTGTTCACATGTAAATAATAATAATAATA ATAAT
$PseA_{EF}68$	936	15	AAAGTAACAACAAAA
$PseA_{EF}74$	1,110	19	CACTGAACTCTTCAGTTAT
$PseA_{EF}75_{acde}$	149	136	GGGAGGAGCTGCTGCTTCCGTCCCGGTTTTCATTGGCCCTCCGC TTCTACCGTGCAGTGGGGGTATCTGTAATGATATTCAGCATGT TTTGCACAAGAAATGTCAGCCAGAAAGGGCTATCTTCTTCCT TCGCCAAA
$PseA_{FG}76$	3,390	12	AGTTTTATGACT
$PseA_{FG}78_{ab}$	3,472	144	ATCTATTTAAATATTACCTAGGCAGTCGTAATTTGTCTAGGCC ATATAGCAATGGTGCTGTAAGGTTTGTCTCCTTTTGTTAAGTTT TATGACTTGCTAGTATATCTAGACTATTGCTGTCTTGGGCGAG TGGTTTCTGTTGAG
$PseA_{FG}79_{bde}$	2,848	71	AGGTAGTTTCATGTTGTTGGGGCTCCATTTTCAATCCCTACAA CACGAAAGTGCCTTAATTGCTCCAGTTA
$PseA_{FG}80$	1,633	28	AAGCAGGGTGTAGGGCACCGATATTTTG
$PseA_{FG}83$	1,392	22	AAGAATTGAAATTCCATAATTT
$PseA_{FG}84$	1,260	23	ATATTTTTATATTAAATAGCACT
$PseA_{FG}92_{abcdf}$	254	209	CATAAAATCATATGGCTTTACAATAAAAAATTATGACTGAAA AATGCTGACCCCCATTTTTACTGGTTTGCGGACTGATTTATAT TTTATTGTTCTGCTTGGCGATCACGTGTTCTGCACCAGCCAAT CCGAATGCAGACGTCTTTTAACTTATTAGGTGTCTGTAGTTCT TTGTGAAGCGAACTTGCTACTTGATTGTACAGTTTCAT
$PseA_{GI}94$	3,806	20	ACTACCTTTATGGCTGCACA
$PseA_{GI}100$	3,592	25	GTAAAACGCAAGTTCTGTGAGTATT
$PseA_{GI}101$	2,865	37	AATATTATCTCATACGTAAATCATATTCAGGTTTCAT
$PseA_{GI}102$	1,725	39	TGCTGTATATGCAGTAGCTTTGTTAGATCCACTAAATAA
$PseA_{GI}103_{abcde}$	1,952	207	ACGCCCTCTGTTGCAGATGGAAGAAACATCAAAGACTTGTCC AGGGATTGTGCCATAAAACGAAGGGTCTGACAAACTGTCTGG AATCTGACTTTTAATGAGTTTACAACTGTGCAGCTACAATTAA GTTGTTTTCAGGAAAGACCATCCGTTTGCTGCTTTTTGCGTGTT TAAACAAAAGGCCAGCGGAAATGCGAGTTTTTATTG
$PseA_{GI}108$	944	44	AGTATATTTGCTGCCTAACTGGCTAACAAATTACCACCAATA AT
$PseA_{GI}109$	461	13	TGGCAATTCATCT
$PseA_{IJ}112$	4,935	24	TATATTTTTGTTCGGTTGTTAATG
$PseA_{IJ}113$	4,041	55	TTAATGTTTTAAGTTGTTTTTAGAAATAAACATGCCACTGTAA ATGTGTAATTTT
$PseA_{IJ}119_{abcd}$	3,100	189	ATGGCGCCTATCTGTGACCTGCCCCTCGGAGAGAAGGAAAA GGCCATAAATCCGTTGTTGTTTATGAAAATTTACAACTTTGCA ATACAACTTTACGAGTTGTTCGGTATTTCCATTGGCCACTCTT GGTCATGTGGATTGTAACCATGAACATGAACTTTTTTATAATA TCCCTATCGAGAATAGAGC
$PseA_{IJ}121$	1,266	15	TTTTTATGTTAATTT
$PseA_{IJ}128$	1,680	23	ATTTTTGAGGTGAAAGTTTTAAA
$PseA_{IJ}130_{a}$	140	60	GATCAGCTGGTTATATTTGCTGCTGTCCTTCTTGGCGTTTAGCC ATCTAGAAACAAACCA
$PseA_{JK}131$	1,077	63	CAATATGCCCACTTGATAATTAATTAAGTAGAGCTCCTAAAC AATTAACACAATTTTTGTTTA
$PseA_{JK}132_{ac}$	328	292	TGGCGCACGATGACAAGGCAAATAGAATTTGAAGCAGGGGC GCCTCTCCAGCGCGCGTTCACTTGTTCAGGCATCCACCCAGT TTTTTTCTGTTTGTGCACGAGTTTACCTCTGGAGGTCATCAAG CAGGATTTACGACTGGACAACAAAAGCACGTGATCCTTAGC CGTACCCCATATTTGGGTGCCTACGTAGGAGGGAACCAAGTA CATGTCCCAGTCATTTCCATAATTCATCATAAATTGTGCAGGG GTG
$PseA_{KL}133$	6,532	50	AATAGCTCCAAGCCACCAGAAAGAAACAAACATTGTTTTAA AACTACCTA
$PseA_{KL}134$	6,356	72	GTGCCTTTGTATGACTTTTTGCACGAACTCCAATAAGTTGCTC TTAAAAGCGCAACTTCATTGATGTATGTT
$PseA_{KL}135$	5,900	31	ATTTTCAGCCTTTGTCTAACAATATTTACTT
$PseA_{KL}136_{ac}$	5,599	207	AAACTTTATTAGCCTCAGTTCCGGCTATATGACATTTGGGTGC CAAATGAATAGGGTTTTGTCTATGAATTAGATCGTAAAATCA TCCATAGAAGAAGCAGATAGGCTTACTGGCTATAAAAAGTC ACGTGATGCCATTAAAGTAAGTTTTATGGTTTTGGGGAGTTGA CATCCAACACTATATTCCACATAACATAATAGCACTGA
$PseA_{KL}140_{bcde}$	4,268	56	AAGATAAATCTGCACACACTCGGAGTCACCAGCAGAGCTCG CTTTAGGCCAAGTTC
$PseA_{KL}142$	3,911	30	TTAAGAAAGTAGACAAGATGTTTTAAAAAT
$PseA_{KL}143_{bc}$	3,707	56	ACAAAAGATTAAACAACTTTGCATCATTCAGATATCTCTCTG GACTCGAGGTGAAC
$PseA_{KL}144$	3,034	36	TAGCTATATCCTGGGTAAGAAAACAAACTGATTTGT
$PseA_{KL}149_{bc}$	2,084	135	TCCAAGGCCGAGGTGAACTTCAGGTCACTGCGTCTCACAAAT ATGAAAATGTCGCCTCCTGAATGACACGCCTTGTTATTGAAC AAAGACTGTCAATGCTTAAGATTAATAAGAAACAAAATGCA CAAGGTGTCA
$PseA_{KL}157_{abc}$	165	165	CCATTGGCGCGTATGCTTACACTGTTGGCAGCTTGTTTTCTGT GATTGGTAGAGAAGGGCACGTGATTATACAACTTTCTACATT TGACAGTGAATGGGAGGGCTTTACTTGGAGCAGAAAAAAAG ACAAAGAGAGAAAAATTAGTATTTTGCCTCCACAAATTA
$PseA_{LM}159$	7,636	51	AACATTACCTTATTCTATTTACTGGATGTTAAACAAAATATAA AATAAAAC
$PseA_{LM}170$	10,595	26	CAAACAGAGGACAGTTTTTCACTTCA
$PseA_{LM}171$	5,653	326	GTTTGAGTCTAGAGCACGCCTTAATTCAATAGCGGTTTGAAG AGGCAAGTATAGTGCAGTATAGTAATGAACCTACCGTTAGGG GTTTCCTTGCAAGAATGATCTCTCGTCGGAAAATATGTTACGT TTAAGACTGTGTTACAAATATGCACAGAATAAAACCCAGATT TAATGCCAAATAAACAAACTATTAACTATGAGATGCATGCGT ATATTGCGTATGAAGGTATAGAACATCCAGCTTACCTCCAAA CC
$PseA_{LM}172$	5,260	56	TTAATGATTCACAAATTCAAATAAAAGGATTTAAGTTGACGT ATGACCACGTGAGC
$PseA_{LM}176$	5,876	42	TTAGAAACCCATAATTGCAAGCCGCGAGCCTATAGGAAATG C
$PseA_{LM}179_{abc}$	5,111	70	AAATAAAAAGGATTTAAGTTGACGTATGACCACGTGAGCGCAT AATACAACGCATTATTGTGGCATTTGGA
$PseA_{LM}181_{bd}$	4,986	58	ATGGATTTAATATTTTGAGGTACTTCCTTTGATTTTTCGCTTTT GATGTTAGTTATTG
$PseA_{LM}180$	4,313	68	ATATAAAATTGCATCCCAGAGTGAGTCCAGCTTTGAAAGGGT CATAATGCTGAGATTTGATTTATGTT
$PseA_{LM}189$	3,359	14	GGATTTGTTTTTTC
$PseA_{LM}193_{bcd}$	2,573	178	GCATAATTTTTCTTGTCTACAGTGACGCTGTTTCGGGGGAGTT CACACAGTGCCTGTACAAGACTTCTTGAGTGAGACTGTCTTTT AAACAGAAGCTGTATGTTTAATATTCATACGATACATACCCA TTTGTTTGCCATTAGATAAGTGCGGACCTGAGGGCTGATAGCT CTTATCT
$PseA_{LM}194_{abcde}$	2,216	198	ATCGATTTTGCTGCTGAATGAGAAAATATTATGTCGTGTTCTG CATTTGTCGTCAGAGGCTAAGGTAAGCAGGCCAGAAATAGG CTATCAGTTGTAAATGGCGGAGTTTATGTCCCAGTGATTTATG ACCATATGACTTAAATCTCGGTTCAAGAAGAGTTCACAAGCT AGAGGCTTCTTTCAAACACAGAATGATAC
$PseA_{LM}195$	1,039	26	TGATTTGCTTGCTCAGTAAACATTAA
$PseA_{LM}198_{ab}$	955	105	TGACCCCTATACATCAAATCTGAGCAATTTTAAAGACAGATA CCATTATTGCTAAAGGTTGGAAGAAAAGTTCAAACTGTGGTC ATGGAGATGAAGACTGTTTTT
$PseA_{LM}196$	889	70	AAAGTTCAAACTGTGGTCATGGAGATGAAGACTGTTTTTCTTT CTTTCTTTCTTTCTTTCTTTCTTTCTT
$PseA_{LM}199_{bcd}$	189	183	TGTCAGTTGCCTATTCAAAAGTAAGATGGATCTGTCTGTTGCC TTATAACAGTCTGGTTGTTTTAACCCTAGGTCTGTTCCAGGTA GCCATTGGTGGAGAGGAGTCACGTGACAGAGGGGTGCCAAT GTTATTCCATAAGGGTGTCAAGACCCTGTCAGTTTATGAAAT AAATATTGGGAAAC
$PseA_{MN}200$	3,484	34	CATTCCATGCTGCTCCGGTTTTCAGAAATAAATA
$PseA_{MN}204_{abcd}$	1,532	91	AGAAGCTGTAAATGTGTTCTTAAGGCTAGAAGCTGTCAGACC GTTTGGCAAACAAGATTGATAGCACACATGCTTCCTTGCTGC TTTGTTT
$PseA_{MN}205$	1,273	22	AAAATATTTCACAGAAATGTTT
$PseA_{MN}206$	1,215	33	CCACAATCTAATCACGTACACTTTATCCAAAAT
$PseA_{MN}207$	954	18	GTATTTTTGATTAATCAA
$PseA_{MN}210_{ab}$	594	126	TTCTCCTTGTGTGTAAATCCTTAAGTAGTTGATATCTAGAGTT CCTTGTGTCAAAACTTTGAAGATTAATGGATTACTTTGTTAAT GACTCCAGGCGTCAGATTTAGGTACTGGAATGATTTGTGA
$PseA_{MN}211_{bcdef}$	290	178	GAGGCGTTCCACAGCGAGTTTTTGGATCAATCACGCAGACAG TGCCTTCTTTTGATTAAAGCCCAAATTGTCATTGGGCAGAAGT AATCATGTGACAACCAATTCGGTCCAATTTCAACCTTGTCTC CATGAATTCAATAGTTTAATAGTAGCGCGGTCCCCATACGGC TGTAATCAG
$PseA_{NO}212$	4,226	16	TACAATTTGTTTGCTA
$PseA_{NO}213$	3,553	20	AAAATATGAAATAGAAAGTA
$PseA_{NO}214$	2,881	219	AGAAAACAAATTAAAGTATAAGTTACAGTTTTCGAGATTTAA AAAGTGGTCAACATTTTTCTGAATAATTTAATCAGTTTTGATT GACTGTGAATTAATATGTAAAACAATAAAATTATTATGCTTA ATGGTTCGTAAGATAAATCGAACGGCAATTTCCCAAGCGATT TGTTAATATAATGCAAATCGCGTTCTAAATATGTTATCAGTTT AGCATTT
$PseA_{NO}222$	237	36	GCGTCACCTTGATACAACAGAGCGTGGAATTTAAAT

Danio rerio zebrafish Hox Aa

PFC	Dist to 3' gene	Length	Sequence
DreAA$_{AC}$1	6,297	16	CTGTCCTCGTGCGCCC
DreAA$_{AC}$7	3,144	39	TTTAAACAAAATATATTTTATATATACTGATAATATAGA
DreAA$_{AC}$9	5,842	21	ATAAATATATTTTTGAGCAAA
DreAA$_{AC}$11	5,590	17	TACTTAGTAGCTGCTTG
DreAA$_{AC}$12	4,907	42	ATATTTTTTACAGTGTGCAGATAATTTGAAAATGTAATTTAA
DreAA$_{AC}$13	4,482	18	ATGGATTTTAGTAAATGT
DreAA$_{AC}$18	4,430	48	CTTGTTATTGCAGTGTTTTATAACAATGAGTTTTAGATGTTTCCAGAA
DreAA$_{AC}$19	2,258	158	CAATAATTATTTCGTTTAATACATTTGTTAAAACGCTATTATTACAAAATATTTAAACAACTTGTACGTCATGTTTAGATTTTAAAAACGAAACAAATTTGTAAATGTTAATAAACCGTTATTTACGATAAGTTTTGATAGTTCATCTTTTACAATGA
DreAA$_{AC}$20	2,153	292	TGTTAATAAACCGTTATTTACGATAAGTTTTGATAGTTCATCTTTTACAATGAAAAATAACACGCAAAAAAAAACATATAAAAATGTAAATCTGTGTTTTATTATTAAGCAATTCCAACCAACTTCTTCACTTTCCAAAGCGTTTGGCTTGATGGACCGTCCATTCACACCCAGATATTGAAATAAATTTAGGCTCCTATCATTTTGCTATTTTGGTGTGTTAAGAAATATTTTACTATAATGCTTTTTTTCAGT
DreAA$_{AC}$21	1,044	238	AAAATACAAAATCAGAAGGAAACCATGTTAAGTTAAAACTACATCTGTTTTCACCAAAGATTTCTCAACAAACAACAAGCTTTTAGACTGGCCATCAAAGTCAGAATAGGCCAACAGGCACTGTGTCGCGGCAAGACGGCAACTTTCAACTTGACCTTGGCCTGCAGACGCGTCAGCCTCGCGCGTAAAATACAGACAAATATAGGGGGGTGTTCTCCGTCTCTGCAACTTGTAAACA
DreAA$_{AC}$30$_{b}$	905	24	CAACTTTCAACTTGACCTTGGCCT
DreAA$_{AC}$31$_{bdef}$	113	103	GCCAATCGAGACGTGATTCTGGATGTACGAGGGTGCGTCAGAGTCAACTGTCTCCCCTTTTCAGTGTCTAAGCTTCACAAAAAAGAGCCACACGTCTCAATGT
DreAA$_{CE}$37	6,026	18	TAATTGAAACAAAAGCAT
DreAA$_{CE}$40	5,678	33	CATCTTCTGATGTGCGTTTGTAAGTGAAGCAGT
DreAA$_{CE}$41	3,163	16	ACATTTGGTTTTAAAA
DreAA$_{CE}$45	2,799	19	AGTGTATTTACAAATGTAA
DreAA$_{CE}$50	1,151	18	CTTGAGGCATTTGGCAGA
DreAA$_{CE}$60$_{abc}$	200	89	TGTCAATTTCTTCACATGATCACGTGACCCTGGCCTCAGTTGGAGCGGATAGAGATGGATTTCCACGTCATCTTACGTCTCAAAATTTC
DreAA$_{EG}$65$_{bc}$	10,155	119	CCTAATTACGGGGACATCCTCCTTGTTGCTTCAGCAACACGGCCATAAAAGCTGTCTGTGTCTGGAGCATTTGGGCAATTGCAGCGTGGTGCCATAAACTGTCTGAGAACCAAGGTTAT
DreAA$_{EG}$70	11,248	16	AGAAATTTAGTAGCTG
DreAA$_{EG}$79$_{ad}$	2,565	77	GACTGTCGAGTGGTTTAGGTAGTTTCATGTTGTTGGGATTACATTCAAACTCTGCAACGTGAAACTGTCTTAATTGC
DreAA$_{EG}$82	1,000	26	TAATGTACTCTTCAAGAGTATAATTT
DreAA$_{EG}$93$_{ace}$	184	108	CACGTGGTCACTTCACCCAATCAAATGGCACTTTGCTCTCAACTTATTATCTGACTGGGGCTTTGAGAGGGGGAAGTTGCCTCTCCATTTTCACATATTGCTATTTTG
DreAA$_{GK}$95	7,678	30	TTTATTGAATAATAAAAGCATTTCATATCT
DreAA$_{GK}$98$_{abcde}$	6,861	113	TGTTAGCCGCATCCAGCCATAAAAGACAATTACTGCTATAACCGTTTATGGTGTGCAAAGCGCTGCGAGGCGAGAAAACGCAACACAACAGATATTATGAGAGTCGACAGCTG
DreAA$_{GK}$99	6,587	20	CCTTTGTTTTATAAGTCCTA
DreAA$_{GK}$100	6,226	25	GTAAAAAGGAAATTATTTAAGTATT
DreAA$_{GK}$114	7,953	50	CAAAATGTGGCCTAAGAACTGGCTTTTGTAATGTCATTACCACTACCTAT
DreAA$_{GK}$115	6,173	13	CGTCCAAACACAA
DreAA$_{GK}$116	3,947	35	TGTTTTAGAGCAAAAACATAATTTAGAAAAAAAGT
DreAA$_{GK}$119$_{abd}$	3,603	205	ATGGCGCCTGACTGCAATCCAGCCCCCTTCTCAGCCAGCGTCTCATGAGAAGCTATAAATCTGTTATTGTTTATGAAAATTTACTACTTTGCCATACAACTTTACGAGCTGCCGCTGGTTCTCATTGGTTACAAGTCAGTCACGTGTTTTAAGACCACATAAACGTTAACTTTTTATGCGTGTGCACCCCAACCCAAAATAGAGC
DreAA$_{GK}$120	2,489	21	TTAAATAGAAAGAAAAACCTT
DreAA$_{GK}$127	2,230	23	AACATGCTAATTCGTGTCTTCAA
DreAA$_{GK}$132$_{abcd}$	378	376	TGGCGCACGGCTACCATCTACAAACAAGCGATTTAACTGCTAGTTCGCCAAAACTGTTGTTGCGCGCTCCCGTGTCCTAACAATTCCACCCTGATTTCCTGCTTACGTCTGAGTTTACCGCTGGAGGTCACAGAATAGGATTTACGACTGGTCAACAAAAAACACGTGATTCTAAGCCATAGCCCATATTTGGAAGCCTACGTAAGGGTGCATTAAGTCTGTGTCTCACTCATTTCCATAATTCATCATAAATAG
DreAA$_{KL}$136$_{ac}$	5,701	216	AAACTTTACTAGACCCTTTACCATTCTGGATCCACGACATTGGCTGGTAAGATAAATAATGAGTGTTTGGTCCACGAATTAGATCGTAAAATTCACTCGGACTGAGACTGCAGATAGGCTCACTTGCCATAAAACAGTCACGTGGTGGCAATTAAAGTCAGTTTTATGGTTTTGGGGAGTCGACACTATATGGCATATAAAGTTAAATCGCACTGA
DreAA$_{KL}$140$_{ade}$	4,105	84	GGTTCACCGTGAGGTCAGTGGCGGCAGGAAGATGAATGTGCACACGGCATAAACCACCAGCAGAGCTCGCTTTTGGCCAAGTTC
DreAA$_{KL}$143$_{bcd}$	3,640	287	GGTCAAAAGTTTAAAGCAGAACAAGTAGCTGTCATCTGTTCTTGTAGAGCGCAGTGCCACTTTTATAATGGCACAAAAGATGGAATAGCTGTGAATCCCGTGAGTTCTGGTCTAACCTCGCGATGAACCCTGCAAGCTGGGTGACCCTAGGTTAGCATCCCAAGAGACTACCTACTCCGCCAATTCCAAAGATCGAGTATTCCTTTTTTATTAAGGGATTTGATTTGGTGAATTTCCTCTGAATAAATAGTATTTG
DreAA$_{KL}$144	2,968	35	TAGCTAAAACCTTCCGAGAAAAGGTATCAATTTGT
DreAA$_{KL}$149$_{bde}$	1,821	165	TCCAAGGCCGGGGTAAATGCAAGGTCACCTCGTCTAACAAATATGAAAATGTCACATTTTCAGAGGGCACGCCTTGTTGTTTAACAAAGACTGTCAATGAACAAGATTAATAAGCGGCGAAATGGACACTGAGTCACTAAGAGGTCAAACCGAAGTTATCTCTGT
DreAA$_{KL}$157$_{abc}$	163	163	CCATTGGTGCTTGTTTACACGATGCCCACAGGAGGCGCTGATTGGTCGCATTTCTCACGTGACCGAGCAACTTTGTACATTTGACAGAGGAGTAGGTGGGTTTTGGAGAGATCAGAATACGACAGCACCATAAAAATTAGAATTGTTGCACTTTTTCAAATTA
DreAA$_{LM}$158	7,072	62	AAATTTTAATTGACTGGCTCACATGACCTTATTAAATAATAGATACCCCTTGGTTAGAAATA
DreAA$_{LM}$164	6,728	31	TTAAAGTTTAATTGTGAATCACAGTAATTTT
DreAA$_{LM}$165	5,813	15	TGACGGAAAACAAGC
DreAA$_{LM}$169	4,278	14	ATTCACTCAAGTTG
DreAA$_{LM}$178	1,141	15	GCACTATTGATTTTG
DreAA$_{LM}$181$_{abc}$	4,699	96	CTGTCATGGACGGATATGTTTTTCCACACTGAAAAAGACAAAGTCCTGCGCACCATGAGGATGGATTTTATTTTGAGGTATTTCCGCTGATTGTTC
DreAA$_{LM}$183	3,833	166	AGCAGTGGCGGCCCGTTTGTTGCGGAGCCTCGCTTGGGAAACGAGGATATAATATAAGGAGTATCATCCAGAAGCACGCTGCCGCACGGGGTCACGACGCCCGAGTCTTGATTTATGGGCTGGGGACTTAGGGCAGCCGCTGCTCGAGAGCTAATGACTATTCGAT
DreAA$_{LM}$193$_{ab}$	2,058	62	AAGGCGCTTAATTTGCATAATTTTTCTTGTCTACGGTGACGCTCTGTTTCGAGGGAGTTCAC
DreAA$_{LM}$194$_{abcd}$	1,624	161	ATCGATTTCTGCCGCTGAATGAGAGAAGCTCGCGGCGTGCTGCGCGCTGCTCGTCAGAGAGGAAGGTAAGCTGGACAGAGGCAGGGTCTCCAATGCTGTTTAAATGGCGAGTTTGTGATTTATGACCGCGCGGCTGAAATCTCGGTTCAGGGAGAGTTCA
DreAA$_{LM}$197	766	17	GGTTTTTCAGGTTTTCT
DreAA$_{LM}$199$_{bcd}$	191	185	TGTCAGCCGAGCTTTGAAAAGTAAGATGGATAGGCCTTATTTTCACCCGCTCACCGTGTCTTCTCTCACCGTAGGTTTACCCGTAGAGCCCATTGGCGAAGAGGTGTCACGTGACTACGTTGAGCCAATGTTCTTCTACAAAGGTGTCAGGACCTTGTCAGAAAGTGAAATAAACATTGGAAAC
DreAA$_{MO}$209	2,217	17	TTAAAAGTTCATTTTTT
DreAA$_{MO}$216	5,075	25	GTCAGACCTTTTGGCAAGTAAGATT
DreAA$_{MO}$217	1,922	28	TTTATGTAATTACGGTTGATGGAAAAAA
DreAA$_{MO}$223	243	78	ACTGGCTGACCCAGCCCACGTGACTGTCACTGGTCATTCATATCACCTAGCGCCCGAACACCTCCAATCCACGACTGAT

Morone saxatilis striped bass Hox Aa

PFC	Dist to 3' gene	Length	Sequence
$MsaA_{FG}76$	3,451	12	AGTTTTATGACT
$MsaA_{FG}79_{bde}$	3,064	69	AGGTAGTTTCATGTTGTTGGGGTCCATTTCTAACTCTGC AACATGAAACTGTCTTAATTGCCCCAGTTA
$MsaA_{FG}93_{abcdef}$	222	127	CACGTGTGTCTGCTACCCAATAGCATGGCAGCCTGACT CCCCATTACTAGCCCACTGTAGTTCTCTGTGGGGCCAA GTTGCTACTTGATTTCTCCACATTGTTATTTTGCGAGGC TGGGTTTACTGC
$MsaA_{Gi}99_{abcde}$	2,714	116	TGTTAGCTGTATACAGCCATAAAAGACAATTACCGCTA TAACCTTTTATGGGGTGCAAAGCGCTGCGAGGCGAGA GGACACAAAACAAAAAAAGACATCGAGGGCTTCGAC AGCTG
$MsaA_{Gi}103_{bcdef}$	2,111	288	GCAGATGGACACAAACTTCAAAGGCTCGACCAGAGAC AGTGCAATAAAACGCCTGGTCTGCTATACTGTCTGGCA TTCCAGTTTTAATGGCTTTATGGCCGTCCAGACACAAT TAGGCCGTTTCCAGAATGGCACCCATTTGTTTTTTCTTC TCTTTCTGTGAGACTGCGCTCTGGACAAAAGGCCGAGC GGAAATGATCAGCTTTATTGGATCCTCCCCGACGGGGA CGCACAGGACTCACGGTCATTTGGAGC
$MsaA_{Gi}107$	1,620	22	TAACTTTTAGCTGGTATTTTAT
$MsaA_{Gi}108$	943	45	AGTATATTTATTGTCATTATTATTTTTAGTATTATTAGTA ATAAT
$MsaA_{Gi}111_{b}$	97	84	GTTTGGTGTAAATCTAGGGTGTATTGCTGTCATATATCA CACTACCTCGTAAAAACGACACTGAGGATTCTGGGCC AACAAATC
$MsaA_{Gi}115$	2,012	13	CGTCCAGACACAA
$MsaA_{IK}117$	4,578	18	TGTCATACGTGTTTTTAA
$MsaA_{IK}119_{bcd}$	3,891	155	ATAAATCCGTTGTTCGTTTAGTGAAAATTTACAACAGT GGTCAATACAACTCTTAACGACCCGGCTCGACATCAT ATTGGCCCAGCTGTGTCACGTGACACCGAGCCGTGAA CATGAACTTTTCTATATATAATTTCCAATAGAGGAGAT AGAGC
$MsaA_{IK}123$	3,033	26	GAAAAATTCATTGGATGGTGGCCAGA
$MsaA_{IK}128$	1,229	23	ATTTTTACATTAAACATTTTAAA
$MsaA_{IK}132_{abcd}$	391	389	TGGCGCACCACGACAATTCAAAACACCCAAGCCCTGG CTCTGAGTTTGCCCTAGCCACGAGCGCATGGAGGCGC GCACCCTTCACATCCCACCCAGTATTTCCTCTGTGCAC GAGTTTACCTCTGGAGGTCACCAAGCAGGATTTACGAC TGGTCAACAAAAGCACGTGATTCTCCGCCGTACCCCAT ATTTGGGTGCCTACGTAAGAGAGAATCAAGTCCATGTC CCACTCATTTCCATAATTCATCATAAATT
$MsaA_{KL}136_{abc}$	7,562	206	AAACTTTATTAGGGCCGATTCTGGCTCTCTGACATTTG GGTGCTAAATGAATGGGGGGTTTTGTCTATGAATTAAA TCGTAAAAATCATCCGGAGCGCGGCCAGATAGGCTCA CTGGCCATAAACGGTCACGTGGTAGCCATTAAAGTAA GTTTTATGGTTTTGGGGAGTTGACAGTATATTGCACATA ACATATAATCGCACTGA
$MsaA_{KL}139$	5,859	46	CAAATTCCGCTCATTGTCATGCAAAGATGTCTCCCTCT TAACAAAA
$MsaA_{KL}140_{abcde}$	5,621	84	GGTTCATCTGGAGGTCAGCCACAGCACCAAGATAAAT CTGCATCTTCTCAGAGCCACCAGCAGAGCTCGCTTTAG GCCAAGTTC
$MsaA_{KL}143_{bcd}$	4,952	315	GGTCAAAAGTTTACGCGCTGAACAAGTCTCCGTCATCT GTTCACTCGGAGCTTGATGCCAGCCTTATAATAGCGAT CTTGACTCTCCGCACAAAAGACAGAATAGCTTTGAATT ACATATGTTGCACGGTGCACTCCAGGTGAACCCTGAA AGCGCGGTGACCTCGCGTTTGGGACCGGGGGAGGGGA CCCCCCACCCGTCCCCAGCAACAAGATTGAGTGGCTG GCGTTTTATTAGGGCGGTTGATTGGTGAAT
$MsaA_{KL}149_{bcde}$	1,855	170	TCCAAGTCCGGGGTGAACCCAGGTCACTGCGTCTAAC AGATATGAAAATGTCGCCTCTTTGGAAAAAAAGGGCA CGCCTTGTTGTTTAACAAAGACTGTCAATGGGCAAGAT TAATCAGAAACAAAATGGAAACGGTGTCACGTTGGGG TCAGGCAGAAGTTATCTCTGT
$MsaA_{KL}157_{abc}$	164	164	CCATTGGTTCCTGTTTACATGATGCCCACAGGACACGC GGTGATTGGTGGCGCTTCACACGTGACCAGGCAACTTT GTACATTTGACAGGGAGTAGGAGGGTTTTGTGGAGATC AGAAAAACGACAGCGCGATAAAAATTAGTATTGTTGC ACTTCACAAATTA

Oreochromis niloticus Nile tilapia Hox Aa

PFC	Dist to 3' gene	Length	Sequence
$OniA_{AC}30_{ab}$	1,073	35	TGGCTAGACAGCAACATTCAACTTGACCTTGGCCT
$OniA_{CE}36$	10,037	24	CCGGGTCTTATCGACGTGGAAGAA
$OniA_{CE}42$	9,408	23	ATTATTATTATTATTTTCTTTTG
$OniA_{CE}54$	3,540	38	GCCTGTAGTAAAGCTTTGAGCTTTTTCTTTTCCTTGAA
$OniA_{CE}57_{bc}$	2,236	120	GCTATAAAAACCCAGGTCTGTGCAGTGCTGGCTTTGGTTTTCCTGG GCATCCTGTCCTCAATTACAGCTTAAAAGCTTCGGCACAACTCTTA GAATTTGTCTGGATTGACCGAGAGGCGG
$OniA_{CE}60_{abcd}$	206	118	TGTCAATTTCTGCCGTGTGGTCGGTCACGTGACCTCCTCCTCCGTG GAGTGGATGGAGATGACTCTCCACGTCAGCCTACGTCTCCCAATTT CTGCTTAGCGAACCTGCTTCAAAGAG
$OniA_{EF}65_{bcd}$	3,627	128	CCTAATTACGGGACATCCTCCCCGTTGCCGCAGCAACGCGGCCAT AAAAGCTGTCTGAGAGTCTGGAGCATTTGTACAATTGGAGTGCAG TGCAATAAACCGTCTGAGACCCAAGGTTATTAACTGTG
$OniA_{EF}72$	1,161	50	CTCTGAAATTTTATGCACACAGGTTTTATATAAAATTAAGAAAGTG CCTT
$OniA_{EF}75_{acde}$	150	137	GGGAGGGACTGCCCATCTTCCAGCTTTTCCATTGGTTTCACAGTCT GCTCGGCGGAGAGGGGGTGCCTCTAATCATATCCAGCATGTTTTGC ACAAGAAATGTCAGCCAGAAAGGGCTACCTTCTCTCTTCGCCAAA
$OniA_{FG}76$	3,273	12	AGTTTTATGACT
$OniA_{FG}79_{bde}$	2,879	69	AGGTAGTTTCATGTTGTTGGGGTCCATTTCTAACTCTGCAACATGA AACTGTCTTAATTGCCCCAGTTA
$OniA_{FG}83$	1,374	20	AAGAATTATGCATTTAATTT
$OniA_{FG}88$	1,721	58	ATACAGTATATTAAATGTATCCTGGTGCAGGCCATTCCCCGCGCCA TTTCCTTTTAAA
$OniA_{FG}93_{abdef}$	219	127	CACGTGTGTCTGCTACCCAATGGCATGGCAGCCTGTCTCCCCATTA CTTTCCCACTGTAGTTCTCTGTGGAGCGAAGTTGCTACTTGATTTCT CCACATTGTTATTTTGTGAGGCTGGGTTTACTGC
$OniA_{GK}94$	10,308	19	ACTACCTTTATATTGCACA
$OniA_{GK}96$	10,146	19	CTATATTTGTTGTCTGTCT
$OniA_{GK}98_{abcd}$	9,217	73	TGTTAGCTGTATACAGCCATAAAAGACAATTACCGCTATAACCTTT TATGGGGTGCAAAGCGCTGCGAGGCGA
$OniA_{GK}103_{bcdef}$	8,610	289	GCAGATGGACCCAAACTTCAAAGACTCGGCCAGAGACAGTGCAA TAAAACGCCTGGTCTGCTATACTGTCTGGCATTCCAGTTTTAATGG CTTTATGGCCGTCCAGACACAATTAGGCCGCTTCCAGAATGGCAC CCATTTGTTTTTTCTTCTCTTTCTGTGAGACTGCGGTCTGGACAAAA GGCCCGAGCGGAAATGATCAGTTTTATTGGATTCTCCTCGACGGGG ACGCGCAGGACTCACGGTCATTTGGAG
$OniA_{GK}111_{b}$	6,657	84	GTTTGGTGTAAATCTAGGGTGTATTGCTGTCATCTATCACACTACCT CGTAAAAACGACACTGAGGATTCTGGGCCAACAAATC
$OniA_{GK}119_{abc}$	3,871	200	ATGGCGCCTGACGCTGACCTGCCCTCCTCCCTCCCTCCTTTTTCTGT CTCTCTCTCCCCCCGTCCCTCTCCCTCCAGCTCTGGGCCATAAA TCCGTTGTTGTTTATGAAAATTTACAACATAGCCATACAAGTTTAC GACGCGGCCCGGGTTCCCATTGGCCGCAGCTTGTCACGTGGCTGG GAGCCATGAACATGAA
$OniA_{GK}132_{abcd}$	484	392	TGGCGCACCACGACAATCCAAAACACCCAAAGCCCCGGCTCAGA GTTTGCCCTAGCCACGAGCGCAGGGAAAGCGCGCACCCCTTATAT CCCACCCAGTATTTCCTGCGTGCACGAGTTTACCTCTGGAGGTCAC CGAGCAGGATTTACGACTGGTCAACAAAAGCACGTGATTCTTCGC CATACCCCATATTTGGGTGCCTACGTAAGAGAGAATCAAGTCCAT GTCCCACTCATTTCCATAATTCATCATAAA
$OniA_{KL}136_{abc}$	7,537	205	AAACTTTATTAGGGCCGTTTCTGGCTCTCTGACATTTGGGTGCTAA ATGAATGGGGGGTTTGTCTATGAATTAGATCGTAAAAATCATCCGG AGAGCGGCCAGATAGGCTCACTGGCCATAAACGGTCACGTGGTAG CCATTAAAGTAAGTTTTATGGTTTTGGGGAGTTGACAGTATATTGC ACATAACATATAATCGCACTGA
$OniA_{KL}140_{abcde}$	5,596	84	GGTTCATCTGGAGGTCAGCCACAGCACCAAGATAAATCTGCATCC TCTCTGAGCCACCAGCAGAGCTCGCTTTAGGCCAAGTTC
$OniA_{KL}143_{bcd}$	4,958	311	GGTCAAAAGTTTACGCGCTGAACAACTCTCCGTCATCTGTTCACTC GGAGCCTGATGCCAGCCTTATAATAGCGATCTTGACTCTCCACAC AAAAGACAGAATAGCTTTGCATTACATATGTTGCATGGTGCACTCC AGGTGAACCCTGGATGCGCAGTGACCTCGGCTTTGAAACCGGGGG AGGAGACCCCCCACCCGTCCCCAGCAACAAGATTGAGTGGTTGGC GTTTTATTAGGGCGGCTGATTGGTGAAT
$OniA_{KL}145$	2,825	21	ACTGAATACTTTGACATTTTA
$OniA_{KL}149_{bcde}$	1,758	169	TCCAAGTCCGGGGTGAACCCCGGGTCACTGCGTCTAACAGATATG AAAATGTCGCCTCTTTGGAAAAAGGGCACGCCTTGTTGTTTAACAA AGACTGTCAATGGGCAAGATTAATCAGAAACAAAATGGAAACGG TGTCACTATGGGGTCAGGCAGAAGTTATCTCTGT
$OniA_{KL}151$	1,254	45	TTATTATTACTATTATTATTATTATTATTATAGGTTTCTTGG
$OniA_{KL}154$	497	17	CCTTTTCTTTTCTTTTC
$OniA_{KL}157_{abc}$	164	164	CCATTGGTTCCTGTTTACATGATGCCCACAGGACACGCCGTGATTG GTGGCTCTTCACACGTGACCAGGCAACTTTGTACATTTGACAGGGA GTAGGAGGGTTTTGTGGAGATCAGAAAAACGACAGCGCGATAAA AATTAGTATTGTTGCACTTCACAAATTA
$OniA_{LM}172$	5,729	57	TTAATGATTCACGGGCTCAAATAAAAGGGATTTAAGTTGACGTTGC GTCACGTGAGC
$OniA_{LM}173$	8,368	16	TCCTATGGACAACAAG
$OniA_{LM}175$	6,224	24	CAATTAGGACTTGGGGGCGATATT
$OniA_{LM}179_{ab}$	5,711	53	AAATAAAAGGGATTTAAGTTGACGTTGCGTCACGTGAGCGCGGCG CATAATAC
$OniA_{LM}181_{abcde}$	5,469	135	CTGTCACGGACGGATATGTTTTTCCACGCGGACGGAGTCAGAGCT CGAGCTTTGGGAGGATGGATTTTATTTTTGAGGTATTTCCGCTGATTG TTCATTACTGTCGTGAGTTATTGCTGCGCGAGGCAGAGGTCAG
$OniA_{LM}190$	3,178	13	GGCTTTCTTTTTG
$OniA_{LM}191$	2,978	37	CTCAGCTATTTGAAAGCACAACTATATATACATATAT
$OniA_{LM}192$	2,438	28	GCATAATTTTTCTTGTCTATGGTGACGC
$OniA_{LM}193_{abcd}$	2,455	201	AAGGCGCTACTTCATTTGCATAATTTTTCTTGTCTATGGTGACGCTC TGGTTCGGGGCAGTTCACTGGTGACTGCATGGTACTAGTGAAGGAT CACATACACCAGCATAGCCTTGGCCGACTTTAATATTGCTATGAA GCCCTGGTCATTGCTCTATAGCAGGACGATACATTCAATCTTGGGG CCAGATAAGTCTTATCT
$OniA_{LM}194_{acd}$	2,028	169	ATCGATTTCGTCGCTGAATGAGAAAATATTGCTTGGTGCTCTGCAT TCGTCGTAAGAGGATAAGGTAAGCAGGCCAGGAATAGGCTCCCTC GGTTGTAAATGGCTGAGTTTGTATGTCGCGCGGTGATTTATCACCG TATGACTTAGATCTCGGTTCAGGAAGAGTTCA
$OniA_{LM}199_{bcd}$	190	184	TGTCAGGGCTGCATTGAAAAGTAAGATGGATCGCCACCATTTCTTC TCCTCACAGTGCTTCTTGTAACCCTAGGTTCACCCAAGGAGGCCAT TGGCGGAGAGGCGTCACGTGACCACGGGGTGCCAATGTTATTCTA CAAGGGTGTCAAGACCCTGTCAGTTTCTGAAATAAATATTGGGAA AC
$OniA_{MN}203$	2,658	45	CAGAGACACAGCTGGAGCTAAAAGACTACACTTTGGGTGTGCAAT
$OniA_{MN}204_{abcd}$	1,720	91	AGAAGCCTTAAATGTGTTGTGAGGGCACCGAGCTGTCAGACCTTTT GGCGAGTAAGATTGATCGCGCGCAGGCTTCCAGGACTCTTTGTTT
$OniA_{MN}216$	1,686	25	GTCAGACCTTTTGGCGAGTAAGATT
$OniA_{MN}211_{abcdef}$	344	213	TGTAAATATAAACAGTCGTCCCCCCCCCAGCTGAGCGAGGCGTTCT TCACCAGAGTTTTGGATCAATCAGGCAGACAGTGGCTTCTTTTGAT TAAACCCCAAATTGTCATTGGGCAGAGGTAATCATGTGACAGGCA ATTCGGTCCAATTTCAACCTTGTCTCCATGAATTCAATAGTTTAAT AGCAGCTCGGTCCCCATACGGCCGTAATCAG
$OniA_{NO}212$	3,601	16	TACAATTAACTTGCTA
$OniA_{NO}213$	3,472	20	AAAATAATTTTTAAAAAGTA
$OniA_{NO}219$	2,919	30	CTCATAAATCACTCCGTGGCATGAATGAGA
$OniA_{NO}221$	784	30	GAAGTTTTCTGTCTTTTGCTTCTCCTTCAC
$OniA_{NO}222$	282	36	GCGTCACCTTGATCGACGAGTGCTTGGAATTTAAAT

Oryzias latipes medaka Hox Aa

PFC	Dist to 3' gene	Length	Sequence
$OlaAA_{AC}1$	5,762	16	CTGTCCTTGTGCGCCC
$OlaAA_{AC}20$	1,772	287	TGTTAATTATCTTGATTTCAAAACAATCTGAATCAAATCATAC CATACAAAAATCAAAAGCAAAAACGCAAGACTTAGCATAA AAACAACGTTCTAGCTTTTATAATTCTAATCCTAAAGTTGGTT TACTTAATTTAAAATTATGAGATTGTCTATTGTTTATATGTTTT CCTGAAACCGCAATAAATATGTTTTTTTTTACTTGTGCTGTCTT CCAATTTTTTCTGTAACAATATGTTTCTAATTTTATGGGGA
$OlaAA_{AC}30_{ab}$	921	42	TGGCTAGACAGCAACTTTCAACTTGACCTTGGCCTCCAGCCG
$OlaAA_{AC}31_{acdef}$	167	157	CTTTGCATAGTGCGGTCACGTGCCGGCGCCCTGACCAATCAG AGCGCGAGCTGCGCGCTCTTGTGCGCGTCAGAGTGGCTTTAC TCTCTGAGAAGCAGTGCGTCCTTTTTTCAGTCAGTGGGCTTTT TAAAAAAGAGCCAGAAGCTTCAATGTTGGA
$OlaAA_{CE}47$	6,877	66	ACCACAAACTCCAGACAGAATCTTTGCTTGTGTTTTGTTCATT ATTTACCACGGTGGTTAATCTTA
$OlaAA_{CE}49$	6,475	17	AGAAATAAAGATTTTGC
$OlaAA_{CE}57_{bc}$	2,146	120	GCTATAAAAACGCAGTTCAGTGCAGAGCTGGCTTTGGTTTTC CTGAGCATCCTGTCCTCAATTACAGCTTAAAAGCCTCAGTAC AACTCGTAGAATTTGCCTGGATTGACTAAGAGGCGG
$OlaAA_{CE}50$	1,266	23	CTTGAGACCATTTTTTAGGCAGA
$OlaAA_{CE}60_{abcd}$	204	115	TGTCAATTTCTGCCGCGTGATCACGTGACCCTCCTCCTCAGTG GAGTGGATGGAGATGGCTCTCCACGTCAGCTTACGTCTCCAA ATTTCTGCTTAGCGAACCTGCTTCAAAGAG
$OlaAA_{EF}65_{bcd}$	3,082	128	CCTAATTATGGGACATCCTCCCCGTTGCCGCAGCAACGCCGC CATAAAAGCTGTCTGAGAGTCTGGAGCATTTGTACAATTGGA GTGCAGTGCAATAAACCGTCTGAGACCCAAGGTTATTAACTG TG
$OlaAA_{EF}75_{acde}$	147	134	GGGAGGGACCTCCTCTTTTCCAGTCTTTTCATTGGTCTGGGGC CTACCCTGGAGAGGGGGTGCCTCTAATCATATCCAGCATGTT TTGTACGAGAAATGTCAGCCAGAAAGGGCTATCTTCTTCCTT CGCCAAA
$OlaAA_{FG}79_{abcde}$	2,645	85	GACTGTCGAGTGGTTTAGGTAGTTTCATGTTGTTGGGGTCCAT TTCTAACTCTGCAACACGAAACTGCCTTAATTGCCACAGTTA
$OlaAA_{FG}89$	904	17	GTAACTGTTTTTATTTT
$OlaAA_{FG}93_{abcdef}$	223	127	CACGTGACTCTGCTGACCAATGGCAGGGGCGTCTGTCTCCCC ATTACTTTCCCACTGTAGTTCTGTGTGGGGCCAAGTTGCTACT TGATTTCTCCACATTGTTATTTTGTGAGGCTGGGTTTACTGC
$OlaAA_{GI}94$	3,952	19	ACTACCTTCATTTTGCACA
$OlaAA_{GI}98_{abcd}$	3,113	73	TGTTAGTTGTATGCAGTCATAAAAGACAATTACCGCCATAAC CTTTTATTGGGTGCAAAGCGCCGCAAGGCGA
$OlaAA_{GI}103_{bcdef}$	2,638	288	GCAGATGGACCCAAACTTCAAAGACTCGGCCAGAGACAGTG CAATAAAACGCCTGGTCTGCTATACTGTCTGGCATTCCAGTTT TAATGACTTTATGGCAGTCCAGACACAATTAGGCCGTTTCCA GAATGACACCCATTTGTTTCTTCTTCTCTTTCTGTGTGACTGAG ATTTCCACAAAAGGCTGGGCGGAAATGATCAGCTTTATTGGA TCCTGCCCGCCGGGGACGCGCAGGACTCACGGTCATTCTGAG C
$OlaAA_{GI}111_{b}$	98	84	GTTTGGTGTAAATCTAGGGTGTATTGCTGTAATATATCACACT ACCTCGTAAAAACGACACTGGGGATTCTGGGCCAACAAATC
$OlaAA_{IK}119_{abc}$	3,725	170	ATGGCGCCTCACCCCTCCCTTCTCCCTCCCCCTTTCTCCCTCT CCCCAGCACCGGGTCATAAATCCGTTGTTGTTTATGAAAATTT ACAACGTAGCGGTCCAACTTTACGACGCGCCTCGGCTGCCTA TTGGCCACAGCGCATCACGTGGGAGGGACCCGTGAACATGA A
$OlaAA_{IK}120$	2,375	21	TTAAATAAAAAACACAACCTT
$OlaAA_{IK}126$	1,416	14	GGCTAATAAAAGAT
$OlaAA_{IK}132_{abcd}$	384	382	TGGCGCAGCGCGACAATTCAAAACAGCAAAGCCCTGCCTTC TCGCTCGCCTTAGCCCAGAGCGCGGAGAGGCGCGCGCCCCTT ACATCCCACCCAGTATTTCCTGCGTGCGCGAGTTTACCTCTGG AGGTCACCGAGCAGGATTTACGACTGGTCAACAAAAGCACG TGATTCCCCGCCGTACCCCCATATTTGGGTGCCTACGTAAGAG AGAATCAAGTCCATGTCCCACTCATTTCCATAATTCATCATA AATT
$OlaAA_{KL}136_{ab}$	6,756	199	AAACTTTATTAAGGGGTTGATTCTGGCTCCCTGACATTTAGGT GCTAAATGAATGGGGGTTTTGTCTATGAATTAGATCGTAAAA ATCATCCGGAGACCGGCCAGATAGGCTCACTGGCCATAAAC GGTCACGTGGTAGCCATTAAAGTAAGTTTTATGGTTTTGGGGA GTTGACAGTATATTGCACATAACATATAAT
$OlaAA_{KL}140_{abcde}$	5,016	84	GGTTCATCGGGAGGTCAGCCGCAGCAGCAAGATAAATCTGC ATCCTCTCGGAGCCACCAGCAGAGCTCGCTTTAGGCCAAGTT C
$OlaAA_{KL}143_{bcd}$	4,450	292	GGTCAAAAGTTTACGCACTGAACAAGTCTCAGTCATCTGTTC ACTCAGAGCCTGATGCCAGCTTTATAATAGGGATCTTGACTC TCCACACAAAAGGCAGAATAGCTTTGCATTACATATGTTGCA CGGTGCACTCCAGGTGAACCCTGGAAGCGCGATGACCTCGC GTTTGGGACCGGGGGGGGGGCAACAAGATTGAGTGGCTTGCG TTTTATTAGGGAGGTTGATTGGTGAATTTCCTTTGAATAAATT GCA
$OlaAA_{KL}146$	2,512	19	TAAGTGTTCTGTTATTTTC
$OlaAA_{KL}149_{bcde}$	1,744	178	TCCAAGGCCGGGGGTGAACCCCCGGTCACTGCGTCTAACAGAT ATGAAAATGTCGCCTCTTTGGGGAAAAAAGGGAGGGCAAAC GCCTTGTTGTTTAACAAAGACTGTCAATGAGTAAGATTAATC AGAAACAAAATGGAAACGGTGTCACTTTAGGGTCAGGCAGA AGTTATCTCTGT
$OlaAA_{KL}157_{abc}$	164	164	CCATTGGTTCCTGTTTACATGATGCCCACAGGACAGCCGGTG ATTGGTGGCTCCTCACACGTGACCAGGCAACTTTGTACATTTG ACAGGGAGTAGGAGGGTTTTGTGGAGATCAGAAAAACGACA GCGCGATAAAAATTAGTATTGTTGCACTTCACAAATTA
$OlaAA_{LM}165$	5,943	15	TGACGGAAAACAAGC
$OlaAA_{LM}172$	5,117	57	TTAATGATTCACGGACTCAAATAAAAGGGATTTAAGTTGACG TTGCGTCACGTGAGC
$OlaAA_{LM}179_{ab}$	4,878	53	AAATAAAAGGGGATTTAAGTTGACGTTGCGTCACGTGAGCAGG GCGCATAATAC
$OlaAA_{LM}181_{abcd}$	4,625	115	CTGTCAGGGACGGATATGTTTTTCCACGCGGACAGAGTCGGA GCCCAAGCTATGGGAGGATGGATTTTATTTTGAGGTATTTCCG CCGATTGTTCGTTACTGTCGTGAGTTATTG
$OlaAA_{LM}182$	5,026	64	GGAAGAAAACCGGAGCCCTGGTCTGGCCATACGATGCAGAT TATATGCTGGTTACCCCCACTTT
$OlaAA_{LM}189$	3,199	14	GGATTTTATTTTTC
$OlaAA_{LM}193_{abcd}$	2,249	201	AAGGCGCGACTTCATTTGCATAATTTTTCTTGTCTATGGTGAC GCTCTGGTTCTGGGCAGTTCACTGGTGTCTGCATGGTACTGGT GAAGGATCTCATAGTCAAGCATCGCCTTGGCTGACGTTAATA TTGCTATGAAGCACTGTCCACTGCTCCATAGCATGACGATAC ATTCAGTCTTCAGGCCAGATAAATCTTATCT
$OlaAA_{LM}194_{acd}$	1,819	169	ATCGATTTCGTCGCTGAATGAGAAAGTATTGCACGGTGCTTTG CATTCATCGTGAGAGGATAAGGTAAGCAGGCCAGAAATAGG CTCCCTCAGTTGTAAATGGTGGAGTTTGTGTGTCGCGCAGTGA TTTATCACCGTATGACTTAGATCTCGGTTCAGGAAGAGTTCA
$OlaAA_{LM}196$	826	72	AAAGTTCACTCCATGGTCATGAGAATGCGAAAACACCCGTCT GGTTTTTCTGGTTTTCTTTGGTAATTTCTT
$OlaAA_{LM}197$	784	17	GGTTTTTCTGGTTTTCT
$OlaAA_{LM}199_{bcd}$	190	184	TGTCAGGTCTGTATTGAAAAGTAAGATGGATCGCCACCATTT CTCCTCCTCACAGTGCTTCTTGTAACCCCAGGTTCACCCAGG GAAGCCATTGGAGGAGAGGCGTCACGTGACCACGGGGTGCC AATGTTATTCTACAAGGGTGTCAAGACCCTGTCAGTTTCTGA AATAAATATTGGGAAAC
$OlaAA_{MN}204_{abcd}$	1,403	91	AGAAGCCTTAAATGTGTTGCGAGGGCACCGAGCTGTCAGACC TTTTGGCGAGTAAGATTGATCGCGCACAGGCTTCCAGGACTC TTTGTTT
$OlaAA_{MN}216$	1,369	25	GTCAGACCTTTTGGCGAGTAAGATT
$OlaAA_{MN}211_{abcdef}$	334	211	TGTAAATAAAAGCAGTCGTGCCCCAGCTGAGAGAGGCGATA TTCACCAGAGTTTTTGGATCAATCAGGCAGACAGTAGCTTCTT TTGATTAAACCCCAAATTGTCATTGGGCAGAAGTAATCATGT GACAGGCAATTCGGTCCAATTTCAACCTTGTCTCCATGAATT CAATAGTTTAATGGTAGCTCGGTCCCCACACGGCCGTAATCA G
$OlaAA_{NO}219$	3,083	30	CTCATAAATCACTCCGTGGCATGAATGAGA
$OlaAA_{NO}217$	1,840	29	TTTATGTATCTCTTGGAAACTTGAAAAAA
$OlaAA_{NO}220$	1,434	24	TAAACATCAAGTCTGCTTGACCTC
$OlaAA_{NO}222$	283	36	GCGTCACCTTGATCGACGAGCGCTTGGAATTTAAAT
$OlaAA_{NO}223$	194	74	ACTGGCTGACCCTGGCCCACGTGACAGCACCCGTTCATTGAT ATCACCCCACGCGCCCTCCTTCCCTTACTGAT

Tetraodon nigroviridis spotted green pufferfish Hox Aa

PFC	Dist to 3' gene	Length	Sequence
TniAA$_{AC}$10	6,067	22	TCCAGGAATGTCTTTAAAGACA
TniAA$_{AC}$16	4,783	27	AATAAAGTTTATCACTGGGTAATAGAA
TniAA$_{AC}$17	3,643	25	TAAATTTAAAGATTTAAAATCCCAT
TniAA$_{AC}$18	3,170	44	CTTGTTGCGCATCTGTAACGACCGCACTAGGTGAT CGTCCAGAA
TniAA$_{AC}$30$_{abc}$	893	42	TGGCTAGACGGCAACTTTCAACTTGACCTTGGCCT CCAGCCG
TniAA$_{AC}$31$_{acdef}$	176	139	CTTTGCATAGCCAGACCACGTGACACCAGCCCGAC CAATGACAGCGCGACCTCGGCGCTCCTGTGCGCGT CAGAGTGGCAGTACTCCCTGGGAACGCGTATCTCC TTTTTTCAGTCCCTCGGCTTTTTAAAAAAGAGCC
TniAA$_{CE}$41	8,070	16	ACATTTAAGGTTAAAA
TniAA$_{CE}$59	1,843	22	GTGGTGTCGCCTGGCCGGAGAG
TniAA$_{CE}$57$_{bc}$	1,789	120	GCTATAAAAACACAGGTCTGTGCGGAGCCGGCTTT GGTTTTTCTGGGCATCCTGTGCTCAATTACTGCTTA AAAGCTTCAGCACAACTCGTAGAATTTGTCTGAAT TGACTGAGAGGCGG
TniAA$_{CE}$60$_{abcd}$	203	114	TGTCAATTTCTGCCGCGCGGTCACGTGATCTCCTCC TCTATGGAGTGGATGGAGATGGCTCTCCACGTCAG CTTACGTCTCCAAATTTCAGCATAGCAAACCTGCTT GAAAGAG
TniAA$_{EF}$65$_{bc}$	3,721	120	CCTAATTACGGGACATCCTCCCTGTTGCCGCAGCA ACGTGACCATAAAAGCCGTCTGAGAGTCTGGGGCA TTTGTACAATTGGAGTGCAGTGCAATAAACCGTCT GAGAACCAAGGTTAT
TniAA$_{EF}$67	1,239	46	GCACAGTGAAGGGTGAACCAAATAAAACAGGTTT TTTAAAAATAAT
TniAA$_{EF}$75$_{cde}$	121	108	ATTGGTTTTACAGTCCGCACAGTGGAGAGGGGGAC GCCTCTAATCATATCCAGCATGTTTTGCACAAGAA ATGTCAGCCAGAAAGGGCTACCTTCTCCCCTCGCC AAA
TniAA$_{FG}$79$_{abde}$	2,514	85	GACTGTCGAGTGGTTTAGGTAGTTTCATGTTGTTGG GGTCCATTTCAAACTCTGCAACATGAAACTGTCTT AATTGCCCCAGTTA
TniAA$_{FG}$82	1,045	24	TAATGTATTATTATGACATAATTT
TniAA$_{FG}$84	914	24	ATATTTTGGAAACTATATAGCACT
TniAA$_{FG}$93$_{abcdef}$	214	127	CACGTGTGTGCGCTGCCCAATGAGAGCGCGCCTTG GCTCCCCATTACAAGCCCACTGTAGTTCTCTGTGGG GCCAAGTTGCTACTTGATTTCTTCACATTGTTATTTT GTGAGACTGTGTTTACTGC
TniAA$_{GK}$98$_{abcde}$	6,866	112	TGTTAGCTGTTTACAGCCATAAAAGACAATTACCG CCATAACCTTTTATGGGGTGCAAAGCGCTGCGAGG CGAGAGGACAACAGAAAAAGACGCGGAGGACGA AGACAGCTG
TniAA$_{GK}$103$_{bcdef}$	6,402	287	GCAGATGGACTCAAACTTCAAAGACGGCCAGAGA CAGCGCAATAAACCCGCCTGGTCTGCTACTCTGTC TGGCATTCCAGTTTAAATTGTTTTATGACCGTCCAG ACACAATTAGTCCGTTTCCAGAATGGCACCCATTT GTTTTTTTCTCCTCTTTCTGTGAGACAACGCTCTGGA CAAAAGGCTCGGCGGAAATGATCAGTTTTATTGGA TTCCACCGACGGGGACGCGCAGGACTCACGGTCAT TTGAAGAG
TniAA$_{GK}$111$_{b}$	4,968	84	GTTTGGTGTAAATCTAGGCAGTATTACTGTCATATA TCAAGCCACCTCGTAAAAACGACACTGAGGATTCT GGACCAACAAATC
TniAA$_{GK}$113	5,215	68	TTAATGAATTCAGTTCCTTATAAAAATCTGCAAATT ATTATTGTATGCCGCTGGAAAAGTGCAATTTT
TniAA$_{GK}$119$_{abc}$	2,714	198	ATGGCGCCTGGTGGTGCCCCCCCTCCTCCCCTTCCT TCTCCCTCCTCCTCCCAGCTCGCAACAGCTCTCCCA CAGGCTCTGGCTCATAAATCTGCTGTTGTTTATGAA AATTTACAACACAGCGACGTAACTTTACGACCGTA CTCGTCTCTCCATTGGCCGCAGCCGTTCACGTGACG ACGGAGCCGTGAACATGAA
TniAA$_{GK}$129	2,004	52	AGTAATTGTCTTTAGATTGACAACATATGTTCAACC CGTCACTGATGCCCTG
TniAA$_{GK}$132$_{abcd}$	372	370	TGGCGCACCACGACAATCCAAAACAAGGGAGTTT GTTCTGGCTTTGGGAGCGCGCAGGCGCACATCCTC CACATCCCACCCAGTAGTTGCTCCGTACTCGAGTTT ACCTCTGGAGGTCACCAGGCAGGATTTACGACTGG TCAACAAAAGCACGTGATTCTCCGGCGTACCCCAT ATTTGGGTGCCTACGTAAGAGAGAATCAAGTCCAT GTCCCACTCATTTCCATAATTCATCATAAATTGTGC AAGGGTGCT
TniAA$_{KL}$136$_{ab}$	6,066	197	AAACTTTATTAAGGCCGAATCTGGCTCTGACATTTG GAAGCTAAATGAATGGGGGGCTTTTGTCTATGAAT TAGATCGTAAAAATCATCCGGAGTGCGGCCAGATA GGCTCACTGGCCATAAACGGTCACGTGGTAGCCAT TAAAGTAAGTTTTATGGTTTTGGGGAGTTGACAGTA TATTGCACATAACATATAAT
TniAA$_{KL}$140$_{abcde}$	4,632	75	GGTTCATGTGGAGGTCACCAAGATAAATCTGCATC CTCTCGGAGCCACCAGCAGAGCTCGCTTTAGGCCA AGTTC
TniAA$_{KL}$143$_{bcd}$	4,083	306	GGTCAAAAGTTTACCACTGAACAAGACTCCGTCAT CTGTTCACTTGGAGCTCGATGCCAGCCTTATAATAG CGATCTTGACTCACACAAAAGACAGAATAGCTTTG AATTACATATGTTGGACTGTGCACTCCAGGTGAAC CCTGGAAGCGCGGTGACCTCGCGTCGAAGACCGG GGGATGGGACCCCCGCCCGCCCCCAGCAACAAGA TTGAGTGCCCGTCGTTTTATTAGGGCGGCTGATTGG TGAATTTCCT
TniAA$_{KL}$149$_{bcd}$	1,667	161	TCCAAGTACGGGGTGAACCCCAGGTCAGCGCGTCT AACAGATATGAAAACGTCGCCCCTTAGAAAAATG GCACGCCTTGTTGTTTAACAAAGACTGTCAATGGG CAAGATTAATCAGAAACAAAATGGAAGCAGTGTC ACTTTTGGGTCAGGCAGAAGTTA
TniAA$_{KL}$157$_{abc}$	164	164	CCATTGGTTCCTGTTTACATGATGCCCACGGGAGA CGCGGTGATTGGTGGCTTTTCACACGTGACCAGGC AACTTTGTACATTTGACAGGGAGTAGGAGGGTTTT GTGGAGATCAGAAAACGACAGCGCGATAAAAAT TAGTATTGTTGCACTTCACAAATTA
TniAA$_{LM}$158	7,161	62	AAATTTTAATTCGCCATGTTCACGTGACTTGTTGAA TAATGAATGCCCTGGGGTCAGAAATA
TniAA$_{LM}$172	4,892	57	TTAATGATTCACGGGCTCAAATAAAAGGGATTTAA GTTGACGTTGCGTCACGTGAGC
TniAA$_{LM}$179$_{a}$	4,874	39	AAATAAAAGGGATTTAAGTTGACGTTGCGTCACGT GAGC
TniAA$_{LM}$181$_{abcde}$	4,589	127	CTGTCACGAACGGATATGTTGTTCCACGCGGACCA AGTGGGCTTCGGGAGGATGGATTTTATTTTGAGGTA TTTCCGCCTATTGTTCATTACCACCGTGAGTTATTG CTGCAGGAGGCAGGGGTCAG
TniAA$_{LM}$193$_{abd}$	2,069	197	AAGGCGCTGCTGATTTGCATAATTTTTTTGTTCGTG GTGACGCTCAGGTTCGGGGCAGTTCACTGGTGTGT GCGTGGTCCCGGTGAAGGATAACAGAGCAACATA GCCTTGGCTGAGTTTGATATTGCCATGAAGCCCTGG TCGTTGCTCCTTAGCATGACGGTACATTCAATCTCA GGCCCAGATAACTCTTATCT
TniAA$_{LM}$194$_{abcd}$	1,646	169	ATCGATTTCGTCGCTGAATGAGAAAATATCGCCCG GTGCTCTGCATTGGTCGTCAGAGGATCAGGTAAGC GGGCCAAAAATAGGCTCCCTCGGTTGTGAATGGCG GAGTTTGTGTGTCGTACGGTGATTTATCACCGTATG ACTTAGATCTCGGTTCAGGAAGAGTTCA
TniAA$_{LM}$199$_{bcd}$	190	184	TGTCAGCCCTGCATTGAAAAGTAAGATGGATCGCC ACCATTTCTTCTCCTCACAGTGCTTCTTGTAACCCT AGGTTCACCCGAGGAGGCCATTGGCGGAGGGGCGT CACGTGACCACGGGGTGCCAATGTTATTCTACAAG GGTGTCAAGACCCTGTCAGTTTCTGGAATAAATATT GGGAAAC
TniAA$_{MN}$204$_{abcd}$	1,468	91	AGAAGCCTTAAATGTGTTGCGAGGGCACCGAGCTG TCAGACCTTTTGGCGAGTAAGATTGATCGCGCACA GGCTTCCAGCACTCTTTGTTT
TniAA$_{MN}$216	1,432	25	GTCAGACCTTTTGGCGAGTAAGATT
TniAA$_{MN}$207	918	19	GTATTTTAGCCATAATCAA
TniAA$_{MN}$211$_{abcdef}$	337	216	TGTAAATAAAAGCAGTCGTTACCAGCTGAGAGAGG CGATCTTCATTTTTTATTTTTTTTTTTGATCAATCATG CAGACAGTGGCTTCTTTTGATTAAACCCCAAATTGT CATTGGACAGAGGTAATCATGTGACAGGCTATTCG GTCCAATTTCAACCTTGTCTCCATGAATTCAATAGT TTAATAGTAGCTTGGTCCCCACACGACCATAATCA G
TniAA$_{NO}$219	2,536	30	CTCATAAATCACTCGGTGGCATGAATGAGA
TniAA$_{NO}$222	281	36	GCGTCACCTCTATCGATGAGGCCCTGGAATTTAAA T

Takifugu rubripes Japanese pufferfish Hox Aa

PFC	Dist to 3' gene	Length	Sequence
$TruAA_{AC}5$	7,175	193	TTATTGGCCTCCCCTTTCGGGTTTGCGGGTTACGTGG AAACCCTGCATCTTTGGCGCATGCAAATTTACATCC AGGCTGATATGCTATGTGAGGGACTCATTATCCAAG TGGAGCAGAAACATTCTTGGTTAGATTTACCACAGC GAGGCTCCAGCCGGCACCGTGTTCAATGGAGAGAC ACAGCTCTGCAGC
$TruAA_{AC}1$	5,423	16	CTGTCCTTGTGCGCCC
$TruAA_{AC}8$	1,704	34	GAAAAAGAGAAGTCATATCAACAGTAAATAACAA
$TruAA_{AC}14$	625	15	ACTGAAAAAAGGAAT
$TruAA_{AC}21$	1,030	235	AAAATAAAGGGTAAAAACATAATGATTGTAAAAGT TAATGTCCTTCAATCGCTCTCAATGGGGCAGGAATT CTCACCCCTCTTCCGCTCCTTTCAAGCATTCAGACA GTTCCTTTAAAGCCTGAATTGCCCGACAGCCTGGAT GGCTAGACGGCAACTTTCAACTTGACCTTGGCCTCC AGCCGCGTCCTATGGAGGAAAAAAGCAGTATTTTCA GTGTTTGCATGCTGTAAACA
$TruAA_{AC}30_{abc}$	888	42	TGGCTAGACGGCAACTTTCAACTTGACCTTGGCCTC CAGCCG
$TruAA_{AC}31_{abdef}$	165	156	CTTTGCATAGTCGGGTCACGTGGTACAACCCTGGCC AATGACGGCGCGTCCTCGGCGCTCCTGTGCACGTCA GAGTGGCAATACTCCCTGGGAACCAGTATCTCTTTT TTCAGTCCCTGGGCTTTTTAAAAAAGAGCCACAAGC CTCAATGTTGGA
$TruAA_{CE}48$	7,169	26	TAATTTAGTTCTATGGTTGTAGGAAA
$TruAA_{CE}53$	6,132	29	TTAAAATAATGCATGTTTTTAACAAAAAT
$TruAA_{CE}55$	2,572	23	ATAAACGACCAACCAGGATTAAA
$TruAA_{CE}57_{bc}$	1,889	120	GCTATAAAAACGCAGGTCTGTGCAGAGCTGCCTTTG GTTTTCCTGGGCATCCTGTGCTCAATTACAGCTTAAA AGCTTCAGCACAACTCGTAGAATTTGTCTGAATTGA CTGAGAGGCGG
$TruAA_{CE}60_{abcd}$	203	114	TGTCAATTTCTACCGCGCGGTCACGTGATCTCCTCCT CCATGGAGTGGATGGAGATGGCTCTCCACGTCAGCT TACGTCTCCAAATTTCAGCATAGCAAACCTGCTTGA AAGAG
$TruAA_{EF}65_{bcd}$	3,300	128	CCTAATTACGGGACATCCTCCCTGTTGCCCCAGCAA CGCGGCCATAAAAGTCGTCTGAGAGTCTGGGGCATT TGTACAATTGGAGTGCAGTGCAATAAACCGTCTGAG AACCAAGGTTATTAACTGTG
$TruAA_{EF}71$	1,325	15	TTTATTGCGGAAAAA
$TruAA_{EF}75_{acde}$	150	137	GGGAGGGACCGCTTATCTTCCAGTATTTCCATTGGTT TTACAGTCCGCACAGTGGAGAGGGGGCGCCTCTAAT CATATCCAGCATGTTTTGCACAAGAAATGTCAGCCA GAAAGGGCTACCTTCTCCCCTCGCCAAA
$TruAA_{FG}79_{abde}$	2,700	85	GACTGTCGAGTGGTTTAGGTAGTTTCATGTTGTTGGG GTCCATTTCAAACTCTGCAACATGAAACTGTCTTAA TTGCCCCAGTTA
$TruAA_{FG}93_{abcdef}$	212	127	CACGTGTGCGCGCTGCCCAATGACATCGCGCCTTGA CTCCCCATTACAAGCCCACTGTAGTTCTCTGTGGGG CCAAGTTGCTACTTGATTTCTCCACATTGTTATTTTGT GAGACTGTGTTTACTGC
$TruAA_{GK}95$	7,934	30	TTTATTTATTAAATAAAACAAATTATATCT
$TruAA_{GK}99_{bcde}$	7,590	98	ATAAAAGACAATTACCGCCATAACCTTTTATGGGGT GCAAAGCGCTGCGAGGCGAGAGGACACAAAACAA AAAAAGACATCGAGAACGACGACAGCTG
$TruAA_{GK}100$	6,809	20	CCTTTGCTTAGCCAGTCCTA
$TruAA_{GK}103_{bcdef}$	7,098	285	GCAGATGGACTCAAACTTCAAAGACGGCCAGAGAC AGCGCAATAAAAACGCCTGGTCGGCTGTACTGTCTG GCATTCCAGTTAAATGGTTTTATGGCCGTCCAGAC ACAATTAGCCCGTTTCCAGAATGGCACCCATTTGTT TTTTCTCCTCTTTCTGTGAGACAACGCTCTGGACAAA AGGCCCAGCGGAAATGATCAGTTTTATTGGATTCCC CGACGGGGACGCGCATGACTCGTGGTCATTTGTATA GAG
$TruAA_{GK}111_b$	5,556	83	GTTTGGTGTAAATCTAGGCTGTATTACTGTCATATAT CAAGCTACCTCGTAAAAACGACACTAAGGATTCTG GCCAACAAATC
$TruAA_{GK}116$	4,073	35	TGTTTTATATCAAAATGTCTAAAAAAAAAAAAAGT
$TruAA_{GK}119_{abc}$	3,321	197	ATGGCGCCCGGTGGTGGGTGACCCCCCCTCTTCCTT CTCTCTCCTCCACCCAGTTCTCACCAGCTCTCCCTCC GGCGCTGGCTCATAAATCGGCTGTTGTTTATGAAAA TTTACAACACAGCGACGTAACTTTACGAGCGGACTC GTCTCCCTATTGGTCGCAGCCGATCACGTGGCGGAG GAGCCGTGAACATGAA
$TruAA_{GK}125$	3,254	15	CCCTCCGGCGCTGGC
$TruAA_{GK}132_{abcd}$	376	374	TGGCGCGCCACGACAATACAAAACAAGCGAGTTTG TCCTGCCTTTGAGCGCACGGAGGCGCGCACCCTCCA CATCCCACCCAGTATTTGCTCTGTGCATGAGTTTACC TCTGGAGGTCACCAGGCAGGATTTACGACTGGTCAA CAAAAGCACGTGATTCACCGGCGTACCCCATATTTG GTTGCCTACGTAAGAGAGAATCAAGTCTATGTCCCA CTCATTTCCATAATTCATCATAAATTGTGCAAGGGT GCT
$TruAA_{KL}136_{abc}$	3,673	204	AAACTTTATTAAGGCCGATTCTGGGTCTGACATTTG GACGCTAAATGAATGGGGGGTTTTGTCTATGAATTA GATCGTAAAAATCATCCGGAGCGCGGCCAGATAGG CTCACTGGCCATAAACGGTCACGTGGTAGCCATTAA AGTAAGTTTTATGGTTTTGGGGAGTTGACAGTATATT GCACATAACATATAATCGCACTGA
$TruAA_{KL}140_{abcd}$	2,046	63	GGTTCATGTGGAGGTCATCTCCGGCACGAAGATAAA TCTGCATCCTCTCAGAGCCACCAGCAG
$TruAA_{KL}149_{bcd}$	1,641	161	TCCAAGTACGGGGTGAACCCCAGGTCAGTGCGTCTA ACAGATATGAAAACGTCGCCCCTTAGAAAAATGGC ACGCCTTGTTGTTTAACAAAGACTGTCAATGGGCAA GATTAATCAGAAACAAAATGGAAGCAGTGTCACTTT TGGGTCAGGCAGAAGTTA
$TruAA_{KL}157_{abc}$	164	164	CCATTGGTTCCTGTTTACATGATGCCCACGGGAGGC GCGGTGATTGGTGGCTTTTCACACGTGACCAGGCAA CTTTGTACATTTGACAGGGAGTAGGAGGGTTTTGTG GAGATCAGAAAAACGACAGCGCGATAAAAATTAGT ATTGTTGCACTTCACAAATTA
$TruAA_{LM}172$	4,830	57	TTAATGATTCACGGGCTCAAATAAAAGGGATTTAAG TTGACGCTGCGTCACGTGAGC
$TruAA_{LM}179_{ab}$	4,812	53	AAATAAAAGGGATTTAAGTTGACGCTGCGTCACGTG AGCGGGGCGCATAATAC
$TruAA_{LM}181_{abcde}$	4,567	128	CTGTCACGGACGGATATGCTTGTTCAACGCGGACCG AGTGGGCTTCGGGAGGATGGATTTTATTTTGAGGTAT TTCCGCCGATTGTTCATTACCACCGTGAGTTATTGCT GCAGGAGGCAGAGGTCAG
$TruAA_{LM}183$	3,996	171	AGCAGTGAACGTGTTGTGATTGTGTCGCTCCGCCGG GGAAAACGAGCATATAATGGAAGACGGAGTGCCAT CTAGGACCGGCCGGCCATGCTGGGGTCACGACGTTC TCGCTTGATTTATGGGCGCGGGGACTTGGGGAAGAT GGTGTGCGAGACCTAATGACTATTCGAT
$TruAA_{LM}186$	3,871	13	CGGGGACTTGGGG
$TruAA_{LM}187$	3,520	30	GTTTTGTTCATGTATGGGCGCCGTTGGAAA
$TruAA_{LM}193_{abd}$	2,012	199	AAGGCGCTGCTGATTTGCATAATTTTTTTGTTCGTGG TGACGCTCAGGTTCGGGGCAGTTCACTGCTGTCTGC GTGGTACTGGTGAAGGATCACATAGCCCAACATAG CCTTGGCTGAGTTTGATATTTCCATGAAGCCCTGGCC ATTGCTCCTTAGCATGACGATACATTCAATCTCGGG CCTAGATAACTCTTATCT
$TruAA_{LM}194_{abcd}$	1,591	169	ATCGATTTCGTCGCTGAATGAGAAAATATCGCCAGG TGCCCTGCATTGGTCGTCAGAGGATCAGGTAAGCAG GCCAGAAATAGGCTCCCTCGGTTGTGAATGGCGGAG TTTGTGTGTCGTACGGTGATTTATCACCGTATGACTT AGATCTCGGTTCAGGAAGAGTTCA
$TruAA_{LM}199_{bcd}$	190	184	TGTCAGCCCTGTATTGAAAAGTAAGATGGATCGCCA CCATTTCTTCTCCTCACAGTGCTTCTTGTAACCCTAG GTTCACCCGAGGAGGCCATTGGAGGAGAGGCGTCA CGTGAACACGGGGTGCCAATGTTATTCTACAAGGGT GTCAAGACCCTGTCAGTTTCTGAAATAAATATTGGG AAAC
$TruAA_{MN}200$	3,053	34	CATTCCATGCTGCTCCAGTTCGAAGAAATAAATA
$TruAA_{MN}204_{abcd}$	1,431	91	AGAAGCCTTAAATGTGTTGCGAGGGCACCGAGCTGT CAGACCTTTTGGCGAGTAAGATTGATCGCGCACAGG CTTCCAGCACTCTTTGTTT
$TruAA_{MN}216$	1,397	25	GTCAGACCTTTTGGCGAGTAAGATT
$TruAA_{MN}211_{abcdef}$	332	212	TGTAAATAAAAGCAGTCGTCCCAGCTGAGCGAGGC GATCTTCATCTGAGTTTTTTTTGGATCAATCATGCAG ACAGTGGCTTCTTTTGATTAAACCCCAAATTGTCATT GGGCAGAGGTAATCATGTGACAGCCTATTCGGTCCA ATTTCAACCTTGTCTCCATGAATTCAATAGTTTAATA GTAGCTTGGTCCCCACACGACCATAATCAG
$TruAA_{NO}214$	2,735	222	AGAAAACTAAAATGAAATTAGTAGAGTAAATAGAG AGCCGTATGCAAATTTCTCAACAGGTTGGCAATGAA TATAATTTGTATAATCGGGTGGCACTGTGATGACAC AGGTTTCACTTTTTACAATTCTTTTAATTGCATTGTA AATCTTAATGTCGTACATCGATATATAGGATGCATA TAAAGTAAATACTAAGATAAAAATCTAAGAACGCT AGCATTT
$TruAA_{NO}222$	284	36	GCGTCACCTTGATCGACGAGCGCCTGGAATTTAAAT

Danio rerio zebrafish Hox Ab

PFC	Dist to 3' gene	Length	Sequence
DreAB$_{AC}$30$_{abc}$	828	39	CTAGACTGCAACTTTCAACTTGACCTTGGCCTCCAGCCG
DreAB$_{AC}$31$_{ad}$	151	106	CTTTGCATGTCACGTGCCTGCTCTCGTCCAATAACAGTTGGATCCTGAC GCACGCGGGGCGCATGGGTATCTGTTACTCCACTGATCCACAGCTTCT CTTTCAGTC
DreAB$_{CE}$44	4,069	19	AATAAGTTGAGGAAATGAA
DreAB$_{CE}$56	2,131	40	TTTTTACGAACTTGGTCTTGTTCACAAATTCTTACATTTT
DreAB$_{CE}$58	813	17	GCTCTTTAACATTTTTA
DreAB$_{CE}$60$_{abc}$	183	87	TGTCAATTTCTGCCCTTGGTCACATGACTGGCGCTCTCTGGAATGGATG GAGATGGATCTCCACGTCAGCTCACGTCTCAAAATTTC
DreAB$_{EF}$64	3,127	32	ATGGAATTATTATTTATTATTAACGAATATTG
DreAB$_{EF}$65$_{abcde}$	2,421	171	GCATCCTTTGAGGCTCGATGGAATTTCGCCCTAATTACGGCACATCCCC CCGTTGCTGCAGCAACTCAGTCATAAAACCTGTCTTAGTCTGGAGCATT TGTAGAATTGCAGTGAGGTGCCATAAACCGTCTGAGAACCAAGGTTAT TAACTGTGACTAAGAGCTAGAAAAC
DreAB$_{EF}$66	2,126	31	CCACTGGATGGCGCACTTTAGATCAATGTCA
DreAB$_{EF}$74	103	19	CACTGAGTTGTATAGTTAT
DreAB$_{EF}$75$_{abc}$	150	105	GGGAGGAGTCATCGGTGCTCGTTGCGATTTAATTGGCTGTTAGCTCACA CTGAGTTGTATAGTTATCAGTAATTATATTCAGAATGTTTTGCACAAGA AATGTCA
DreAB$_{FG}$79$_{bd}$	2,019	62	AGGTAGTTTTATGTTGTTGGGCTCTATTTTATATCCCCGCAACACGAAA CTGTCTTAATTGC
DreAB$_{FG}$86	1,724	14	CAAGGCCTCGGTTA
DreAB$_{FG}$81$_{ab}$	1,650	93	GCCGGATGCGTCTGCTGGGCCGGAGAAAAAAAATACAAAGGCTTTATT GGCCATATTCAAGCCAGACAGCCAAGTTACAACCCGAAGAAAAAC
DreAB$_{FG}$92$_{ab}$	281	37	CATAAAAGCGTATGACTCTACCGTTAAAAATATGACT
DreAB$_{GN}$98$_{bc}$	9,448	44	ATAAAACACAATCACAGTCTTAACTGTTTATAGGGTGCAAAGCG
DreAB$_{GN}$102	4,893	40	TGCTGTATAGGCCACCGATTTTGTTTATTGCAAAAAATAA
DreAB$_{GN}$135	6,703	31	ATTTTCAAGTTATGTTCAGCAAACTTTACTT
DreAB$_{GN}$141	4,750	29	TTGTTTGAGGCCAAATTTTGGTCTAATAT
DreAB$_{GN}$142	4,701	32	TTAAGATCGTAAATTTGAATAATTGTAAAAAT
DreAB$_{GN}$153	1,598	25	TAAAGATACATCCTTGTTTGATTTC
DreAB$_{GN}$159	7,285	44	AACATTATTTATTAAGTTTTATAATACAAAACAACTGGTAAAAC
DreAB$_{GN}$195	978	24	TGATTTTTTTTTATCTCACATTAA
DreAB$_{GN}$204$_{bcd}$	1,159	82	AAATGTGTTCTTAGGGCACTAAGCTGTCAGACCTTTTGGCGTGTAAGAT TGATAGCGTGCAGGCTTCCAGGGCTCTTTGTTT
DreAB$_{GN}$205	1,362	23	AAAATATAACACAAATAATGTTT
DreAB$_{GN}$211$_{bcdef}$	271	179	GAGGCGTTCCTCTCGGACTTTTTTCGATCAATCACACAGACAGTGGCTT CTTTTGATTAAACCCCAAATTGTCATTGGACAGAGGTAATCATGTGAC AAGCAATACGGTCCAATTTCAACCTTGTCTCCATGAAAGCAATAGTTT AATGGCACCGCGGTCCCCATACGGCCGTAATCAG

Oryzias latipes medaka Hox Ab

PFC	Dist to 3' gene	Length	Sequence
OlaAB$_{AC}$30$_{bc}$	663	31	CAACTTTCAACTTGACCTTGGCCTCCAGCCG
OlaAB$_{CE}$32	3,911	116	TATAATTTATTACATTATGGACAGCTTTCATGTCGCCTGAGTGGCT CTGTTTGAACTAAACATTTCCCATCTCACGTTTTAATTGGAACCAG TCGATGCGGGCTGTTGACAGAAGA
OlaAB$_{CE}$33	3,267	44	AAGCCAACGTTTTGACAATAAATGATTGAGATATGCAGTGTGTT
OlaAB$_{CE}$34	1,814	39	CAGTTTTTGTTCTTCTTTTTTTCTCATAATCACTTTAAA
OlaAB$_{CE}$58	790	17	GCTCTTTTCCCTTTTTA
OlaAB$_{EF}$61	3,781	60	TGGAAATCAGACCTGCAGCAGTTAATATTTGATATCCATGACAAA AGGCCCCCTTCATTT
OlaAB$_{EF}$65$_{acde}$	3,336	170	GCATCCTGCCAGGCTCCCAGGAATTCCACTTTGATTGCTGCACAT CCTCCCAGTTGCTTCAGTGACTTGGCCATAAAAGGCGGATTCGTC TGGAGCATTTGCAGTGCGGTGCAATAAAGCGTCTGAGACCAAGG TTATTAACTGTGACTCGGGGGGGGGGGATGAAAAC
OlaAB$_{EF}$66	3,050	31	CCACTGGATGGCGCTCCTTAGATCAATGTCA
OlaAB$_{EF}$68	859	15	AAAGTAAAAACAAAA
OlaAB$_{EF}$75$_{abc}$	144	99	GGGAGGAGTCAGTCTCCTGCTGTTTTAATTGGCTGCAGACTTCAGT GAGAAGCGTGAGTATCAGTAATTATCCGGCAATGTTTTGCTCAAG AAATGTCA
OlaAB$_{FG}$86	867	14	CAAGGCCTCGGTTA
OlaAB$_{FG}$81$_{b}$	763	52	TTATTGGGTATATTTAGCTTAGACAGCTCTAGGTTGTAACCTTAAG AAAAAC
OlaAB$_{FG}$87	340	241	AAATCATGTTTATTGCTTATAAAAGAAGATATTTGCCAATAAAAG CGCAGGGATTTGCTATTAAAATATGGTGAGGAAAATGGTGGCCAT TTAGCTGAGCAGCAAAGGCTGATTTATTATTTATTGTTCAGTCCCA CGGTCACGTGTTTGGGGCGCCCTATCCAGCATTGGACAAGGTTCA GCTTTCTGGGCGCATGAGCATGTAGGGTGCACAAAAAGCGACAC CTTACCATGCAATGAA
OlaAB$_{GN}$104	1,640	31	TTTAATGTTTAAAAAAACTTACAAACACACC
OlaAB$_{GN}$204$_{bc}$	805	53	AAATGTTCTCTAAGGGCAAAGAGCTGTCAGAGCTCTTGGCTGGAA AGATTGAT
OlaAB$_{GN}$211$_{c}$	171	23	TCATTGGGCAGATATCATGTGAC

Tetraodon nigroviridis spotted green pufferfish Hox Ab

PFC	Dist to 3' gene	Length	Sequence
$TniAB_{AC}30_{bc}$	844	31	CAACTTTTAACTTGACCTTGGCCTCCAGCCG
$TniAB_{CE}32$	3,993	117	TATAATTTATTACATGATATGAAACGTTCCTTTGTCGCCATAACG ACGCTGTTTGACCTACACATTTCCTTCCTTTCTTTTAATTGGAACC AGTCCGAGCGGCCTGTTGACAGAAGA
$TniAB_{CE}33$	3,293	52	AAGCCAAGGGTTTTTTTTTTAATAAATAAACGTTTCAGAAACGC AATGTGTT
$TniAB_{CE}34$	1,808	39	CAGTTTGTTTTGGGATTTTTTTTAAAAACATGTTTTAAA
$TniAB_{CE}51$	1,736	23	TGTGTGTTTTGTGTAATCTCCAA
$TniAB_{EF}65_{acde}$	2,287	163	GCATCCTGTGGGGCTTTGAGGAATTCCGCCCTGATTGCTGCACA TCCTCCCAGTTGCTCTAGTAACTTGGTCATAAAACGCGGATTCGT CTGAAGCATTTGGGAAGCAGCGCCATAAAGCGTCTGCGACCAA GGTTATTAACTGTGACTGAGAGGTTGAAAAC
$TniAB_{EF}66$	2,026	31	CCACTGGGCGGCGCCCCTCAGATCAATGTCA
$TniAB_{EF}75_{abcd}$	147	112	GGGAGGAGTCTGTAGTCGCCTAGTGTTTTAATTGGCTGCAGGCT ACAGTGAGAAGCGTGAGTATCAGTAATTATTCAGCAATGTTTTG CACAAGAAATGTCAGCCAGAAAGG
$TniAB_{FG}79_{bcd}$	1,233	54	AGGTAGTCTCATGTTGTTGGGTTCTAATATTTCTCCCACAACAGG AAACTGCCT
$TniAB_{FG}81_{ab}$	779	97	GCCGGATGCGTCTGCCTCCGCAAGAAAAGGGACACAAAAGCTA AACTTTATTGGTGATATTTAGCTGAGACAGCCAGGTTGTAACCT AAAGAAAAAC
$TniAB_{FG}86$	862	14	CAAGGCCTCGGTTA
$TniAB_{GN}101$	3,571	38	AATATTGCAGTAAAGTTTGATCAGCAATCTAGTTTCAT
$TniAB_{GN}106$	1,083	323	GTCCAAAATTTCTACCAGAAGGTGACGGAATATTTGCCTTTTGTC TGTGTCGGAGATAAAGGTGTGAGCCTGAATGGAATCGCTCCTAC AGCCAACACCCACTCACCTCAGCGGACAATCGATCCTGCACAA AAGATGGTCCATACGTTTCTATACCATTACACAAAACCATTCCC TCTTCTGTATGTTTTTTGTGCCATGATTGTACCTTCTTGCTAAAGA AAAACGCGCCAGTTTCGTTTGGAAGTGATTTTC
$TniAB_{GN}121$	569	14	TTTTTAATGAATTT
$TniAB_{GN}141$	4,536	28	TTGTTTTGGTTGTTTTTTTTTTTAATAT
$TniAB_{GN}171$	3,117	313	GTTTGAGCCGGAGCGCGCTGGATTCCCACGATGGCGAATTTGGG AGATCATTTTATCCACCTTCCTGCCCTTCTTCTTTGTTTCCTTCAC CCTCAGCTTTTTTCCCCTGACTTATCTGGACAATAAACTCACATC TTTTACATCTCATTACTTTCGGTATCATTTGTTCCATAAACAGCA CACTCGCCCGGAAAACAATAATCCTCTGATGGCTTTCGTTGATTT TATTAATATCAAGTATTGCAACTGTTCCAG
$TniAB_{GN}204_{cd}$	1,007	60	AGCTGTCAGACCTTTGGGCGAGAAAGATTGATCACACTCAGGCT ACCGAGGTCTTTGTTT
$TniAB_{GN}206$	1,426	30	CCACAATTGAATGTGTAAATTGTCCAAAAT
$TniAB_{GN}211_{abce}$	298	199	TGTAAATAAAGCAATCCCTACCCCCGGAATGAGGCGTTACCCCT CCGACGTTCTCGATCATTCACAGGGACAGTTGCTTCTTTTGATAG AGAAAAAAACCCTCAAATTGTCATTGGGCAGATTTAATCATGTG ACATGCAACGAGCTCTCATTCCAACCGCATCCTTGTGAAAGCAA TAGTTTATGGCACAGAGGTCCC

Takifugu rubripes Japanese pufferfish Hox Ab

PFC	Dist to 3' gene	Length	Sequence
$TruAB_{EF}61$	2,733	59	TGGAAATGCGATCTACTGCAGATGTTATTTGATATCC ATTACAAAGAGTTCCTTCATTT
$TruAB_{EF}65_{acde}$	2,263	163	GCATCCTGTGAGGTTCCGAGGAATTCCGCCCTGATGG CTGCACATCCTCCCAGTTGCTCCAGTAACTTGGCCAT AAAACGCGGACTCGTCTGAAGCATTTGGAGAGCAGC GCAATAAAGCGTCTGCGACCAAGGTTATTAACTGTG ACTGAGAGGTTGAAAAC
$TruAB_{EF}66$	2,029	31	CCACTGGGTGGCGCCCCTCAGATCAATGTCA
$TruAB_{EF}75_{abcd}$	147	112	GGGAGGAGTCTATAGCCGCCTAGTGTTTTAATTGGCT GCAGGCTACAGTCAGAAGCGTGAGTATCAGTAATTA TTCAGCAATGTTTTGCACAAGAAATGTCAGCCAGAA AGG
$TruAB_{FG}80$	1,492	28	AAGCAGGATGTTTTGTTTTTGTATTTTG
$TruAB_{FG}81_{ab}$	806	97	GCCGGATGCGTCTGCCTCGGCAAGAAAAGGGACACA AAAGCTGAACTTTATTGGTGATATTTAGCTTAGACAG CCAGGTTGTAACCTAAAGAAAAAC
$TruAB_{FG}86$	859	14	CAAGGCCTCGGTTA
$TruAB_{FG}87$	306	245	AAATCATAACCTTTATAGCTTATAAAAAGTCATATTT TCCCCAATAAATGCGCGCGAATTTGCTATTAAAATAT GGTGGTGGGGGGGAAGGGGGATGGTGGCCATTGAGT TGAGCAGCAGGCTGATTTATTATTTATTGTTCTGTCCC ACGGTCACGTGTTTGGGGCGCCCTATCCAGTCTTGGG CAAGGTTCAACTTGCTGGACGCACTGGTCCCCTAGAT GGTGCAGAAAGGCACGCAATGAA
$TruAB_{FG}90$	864	28	CAGGACAAGGCCTCGGTTATGATCATTG
$TruAB_{GN}109$	890	13	TGGCAAATCATCT
$TruAB_{GN}112$	4,295	24	TATATTTGAGTATCATTGTTAATG
$TruAB_{GN}131$	2,260	56	CAATATGGTGTGATATATGATACAGCAGCCATAAAT AATTTAGAAACTTCTGTTTA
$TruAB_{GN}204_{bcd}$	1,027	80	AAATGTGTTCTTAGGGCAGGGAGCTGTCAGACCTTTT GGCGTGAAAGATTGATCACACTCAGGGACCGAGGTC TTTGTTT
$TruAB_{GN}208$	1,217	44	ACAGACTTTCGTCTATAAATGAATTTCTCAATCATAA TCGAGCC
$TruAB_{GN}211_{abce}$	297	201	TGTAAATAAAGCAACCCCTACCCCCGGAATGAGGCG TTACCTCTCCGACGTTCTCGATCAATCACAGGGGACA GTGGCTTCTTTTGATAGAGGAAAAAAAACCTCAAATT GTCATTGGGCAGATTTAATCATGTGACAAGCAACAC GGTCTAATTCCAACCACATCCTTGTGAAAGCAATAGT TTATGGCACAGGGGTCCC

PFC	Frog B	Coelacanth B	Human B	Bichir B	Zebrafish Ba	Medaka Ba	Tetraodon Ba	Zebrafish Bb	Medaka Bb	Tetraodon Bb
1					DreBA$_{AC}$1		TniBA$_{AC}$1			
2				PseB$_{AC}$2			TniBA$_{AC}$2			
3				PseB$_{AC}$3		OlaBA$_{AC}$3				
4				PseB$_{AC}$4			TniBA$_{AC}$4			
5						OlaBA$_{CG}$5	TniBA$_{CG}$5			
6					DreBA$_{FG}$6	OlaBA$_{CG}$6				
7						OlaBA$_{CG}$7	TniBA$_{CG}$7			
8					DreBA$_{CF}$8		TniBA$_{CG}$8			
9					DreBA$_{CF}$9		TniBA$_{CG}$9			
10					DreBA$_{CF}$10		TniBA$_{CG}$10			
11					DreBA$_{CF}$11		TniBA$_{CG}$11			
12					DreBA$_{CF}$12	OlaBA$_{CG}$12				
13						OlaBA$_{CG}$13	TniBA$_{CG}$13			
14						OlaBA$_{CG}$14	TniBA$_{CG}$14			
15						OlaBA$_{CG}$15	TniBA$_{CG}$15			
16				PseB$_{CF}$16		OlaBA$_{CG}$16				
17				PseB$_{CF}$17	DreBA$_{CF}$17					
18				PseB$_{CF}$18		OlaBA$_{CG}$18				
19				PseB$_{CF}$19	DreBA$_{CF}$19					
20				PseB$_{CF}$20	DreBA$_{CF}$20					
21				PseB$_{FH}$21		OlaBA$_{CG}$21				
22	XtrB$_{CG}$22		HsaB$_{CG}$22							
23		LmeB$_{CF}$23	HsaB$_{CG}$23							
24		LmeB$_{CF}$24	HsaB$_{CG}$24							
25		LmeB$_{CF}$25	HsaB$_{CG}$25							
26		LmeB$_{CF}$26	HsaB$_{CG}$26							
27					DreBA$_{FG}$27		TniBA$_{CG}$27			
28			HsaB$_{CG}$28				TniBA$_{CG}$28			
29		LmeB$_{CF}$29	HsaB$_{CG}$29							
30			HsaB$_{CG}$30			OlaBA$_{CG}$30				
31		LmeB$_{CF}$31	HsaB$_{CG}$31							
32		LmeB$_{CF}$32	HsaB$_{CG}$32							
33		LmeB$_{FG}$33	HsaB$_{CG}$33							
34		LmeB$_{CF}$34	HsaB$_{CG}$34							
35	XtrB$_{CG}$35		HsaB$_{CG}$35							
36	XtrB$_{CG}$36		HsaB$_{CG}$36							
37	XtrB$_{CG}$37		HsaB$_{CG}$37							
38	XtrB$_{CG}$38		HsaB$_{CG}$38							
39						OlaBA$_{GH}$39	TniBA$_{GH}$39			
40					DreBA$_{GH}$40	OlaBA$_{GH}$40				
42				PseB$_{FH}$42		OlaBA$_{GH}$42				
43			HsaB$_{GH}$43			OlaBA$_{GH}$43				
44			HsaB$_{GH}$44				TniBA$_{GH}$44			
45			HsaB$_{GH}$45		DreBA$_{GH}$45	OlaBA$_{GH}$45	TniBA$_{GH}$45			
46			HsaB$_{GH}$46		DreBA$_{GH}$46					
47					DreBA$_{GH}$47		TniBA$_{GH}$47			
48					DreBA$_{GH}$48	OlaBA$_{GH}$48				
49			HsaB$_{GH}$49	PseB$_{FH}$49						
50	XtrB$_{GH}$50$_{abcdef}$	LmeB$_{GH}$50$_{abc}$	HsaB$_{GH}$50$_{abcdef}$	PseB$_{FH}$50$_{abef}$	DreBA$_{GH}$50$_{bcf}$					
51	XtrB$_{GH}$51$_{ab}$	LmeB$_{GH}$51$_{abcd}$	HsaB$_{GH}$51$_{abc}$	PseB$_{FH}$51$_{abcd}$	DreBA$_{GH}$51$_{bcd}$					
52	XtrB$_{GH}$52$_{bcde}$	LmeB$_{GH}$52$_{abcde}$	HsaB$_{GH}$52$_{bcde}$	PseB$_{FH}$52$_{acd}$	DreBA$_{GH}$52$_{acd}$					
53						OlaBA$_{GH}$53	TniBA$_{GH}$53			
54						OlaBA$_{HJ}$54	TniBA$_{HJ}$54			
55					DreBA$_{HJ}$55		TniBA$_{HJ}$55			
56					DreBA$_{HJ}$56		TniBA$_{HJ}$56			
57			HsaB$_{HJ}$57			OlaBA$_{HJ}$57				
58				PseB$_{HJ}$58			TniBA$_{HJ}$58			
59				PseB$_{HJ}$59				DreBB$_{HJ}$59		
60	XtrB$_{GH}$60$_{a}$	LmeB$_{HJ}$60$_{ab}$	HsaB$_{HJ}$60$_{a}$	PseB$_{HJ}$60$_{ab}$	DreBA$_{HJ}$60$_{ab}$	OlaBA$_{HJ}$60$_{ab}$				
61				PseB$_{HJ}$61				DreBB$_{HJ}$61		
62		LmeB$_{HJ}$62$_{abc}$	HsaB$_{HJ}$62$_{abc}$	PseB$_{HJ}$62$_{abc}$	DreBA$_{HJ}$62$_{b}$					
63		LmeB$_{HJ}$63$_{ab}$	HsaB$_{HJ}$63$_{ab}$	PseB$_{HJ}$63$_{ab}$	DreBA$_{HJ}$63$_{b}$					
64			HsaB$_{IJ}$64					DreBB$_{HJ}$64		
65	XtrB$_{IJ}$65$_{abc}$	LmeB$_{IJ}$65$_{abcd}$	HsaB$_{IJ}$65$_{abcd}$		DreBA$_{IJ}$65$_{abcd}$	OlaBA$_{HJ}$65$_{bc}$		DreBB$_{HJ}$65$_{cd}$		
66		LmeB$_{IJ}$66	HsaB$_{IJ}$66							
67					DreBA$_{IJ}$67		TniBA$_{HJ}$67			
68					DreBA$_{IJ}$68		TniBA$_{HJ}$68			
69					DreBA$_{IJ}$69		TniBA$_{HJ}$69			
70	XtrB$_{IJ}$70$_{ab}$	LmeB$_{IJ}$70$_{ab}$	HsaB$_{IJ}$70$_{ab}$		DreBA$_{IJ}$70$_{ab}$	OlaBA$_{HJ}$70$_{ab}$	TniBA$_{HJ}$70$_{ab}$	DreBB$_{HJ}$70$_{b}$	OlaBB$_{IJ}$70$_{b}$	TniBB$_{IJ}$70$_{b}$
71		LmeB$_{IJ}$71	HsaB$_{IJ}$71							
72									OlaBB$_{JK}$72	TniBB$_{JK}$72
73									OlaBB$_{JK}$73	TniBB$_{JK}$73
74									OlaBB$_{JK}$74	TniBB$_{JK}$74
75						OlaBA$_{JK}$75	TniBA$_{JK}$75			
76						OlaBA$_{JK}$76	TniBA$_{JK}$76			
77					DreBA$_{JK}$77		TniBA$_{JK}$77			
78				PseB$_{JK}$78					OlaBB$_{JK}$78	
79	XtrB$_{JK}$79$_{ab}$	LmeB$_{JK}$79$_{abc}$	HsaB$_{JK}$79$_{abc}$	PseB$_{JK}$79$_{abc}$	DreBA$_{JK}$79$_{bc}$			DreBB$_{JK}$79$_{bc}$	OlaBB$_{JK}$79$_{abc}$	TniBB$_{JK}$79$_{abc}$
80		LmeB$_{JK}$80	HsaB$_{KL}$80							
81									OlaBB$_{KM}$81	TniBB$_{KM}$81
82									OlaBB$_{KM}$82	TniBB$_{KM}$82

83				$PseB_{KL}83$	$DreBA_{KL}83$					
84			$HsaB_{KL}84$	$PseB_{KL}84$						
85				$PseB_{KL}85$	$DreBA_{KL}85$					
86				$PseB_{KL}86$	$DreBA_{KL}86$					
87	$XtrB_{KL}87_{bcdef}$	$LmeB_{KL}87_{acdef}$	$HsaB_{KL}87_{abcdef}$	$PseB_{KL}87_{abcdef}$	$DreBA_{KL}87_{abcdef}$	$OlaBA_{KL}87_{bcdef}$	$TniBA_{KL}87_{bcdef}$	$DreBB_{KO}87_{df}$	$OlaBB_{KM}87_{de}$	$TniBB_{KM}87_{de}$
88			$HsaB_{KL}88$					$DreBB_{KO}88$		
89			$HsaB_{KL}89$	$PseB_{KL}89$						
90		$LmeB_{KL}90_{abcd}$	$HsaB_{KL}90_{abcd}$	$PseB_{KL}90_{abcd}$	$DreBA_{KL}90_{b}$			$DreBB_{KO}90_{abcd}$	$OlaBB_{KM}90_{abc}$	$TniBB_{KM}90_{bc}$
91				$PseB_{KL}91$	$DreBA_{KL}91$	$OlaBA_{KL}91$	$TniBA_{KL}91$			
92								$DreBB_{KO}92$	$OlaBB_{KM}92$	
93								$DreBB_{KO}93$	$OlaBB_{MO}93$	$TniBB_{MO}93$
94				$PseB_{KL}94$	$DreBA_{KL}94$					
95				$PseB_{KL}95$			$TniBA_{KL}95$			
96			$HsaB_{KL}96$			$OlaBA_{KL}96$				
98			$HsaB_{KL}98$		$DreBA_{KL}98$					
99			$HsaB_{KL}99$						$OlaBB_{KM}99$	
100		$LmeB_{KL}100$	$HsaB_{KL}100$							
101		$LmeB_{KL}101$	$HsaB_{KL}101$							
102		$LmeB_{KL}102_{ab}$	$HsaB_{KL}102_{ab}$	$PseB_{KL}102_{ab}$	$DreBA_{KL}102_{ab}$	$OlaBA_{KL}102_{b}$	$TniBA_{KL}102_{b}$			
103					$DreBA_{KL}103$	$OlaBA_{KL}103$				
104			$HsaB_{KL}104_{bcde}$	$PseB_{KL}104_{abcde}$	$DreBA_{KL}104_{abcde}$	$OlaBA_{KL}104_{abcd}$	$TniBA_{KL}104_{abcd}$	$DreBB_{KO}104_{abce}$		$TniBB_{MO}104_{b}$
105				$PseB_{KL}105_{abc}$	$DreBA_{KL}105_{abc}$	$OlaBA_{KL}105_{ac}$	$TniBA_{KL}105_{bc}$			
106				$PseB_{KL}106$	$DreBA_{KL}106$					
107			$HsaB_{KL}107$		$DreBA_{KL}107$					
108		$LmeB_{KL}108$	$HsaB_{KL}108$	$PseB_{KL}108$	$DreBA_{KL}108$					
109						$OlaBA_{LM}109$	$TniBA_{LM}109$			
110	$XtrB_{KL}110_{abc}$	$LmeB_{LM}110_{abc}$	$HsaB_{LM}110_{b}$		$DreBA_{LM}110_{abc}$		$TniBA_{KL}110_{bc}$			
111		$LmeB_{LM}111$	$HsaB_{LM}111$							
112		$LmeB_{LM}112$	$HsaB_{LM}112$							
113	$XtrB_{KL}113_{ab}$	$LmeB_{LM}113_{ab}$	$HsaB_{LM}113_{ab}$		$DreBA_{LM}113_{bc}$	$OlaBA_{LM}113_{bc}$	$TniBA_{LM}113_{bc}$			
114	$XtrB_{KL}114_{b}$	$LmeB_{LM}114_{ab}$	$HsaB_{LM}114_{ab}$		$DreBA_{LM}114_{ab}$					
115			$HsaB_{LM}115$					$DreBB_{KO}115$		
116		$LmeB_{LM}116$	$HsaB_{LM}116$							
117			$HsaB_{LM}117$						$OlaBB_{KM}117$	
118	$XtrB_{LM}118_{abcde}$	$LmeB_{LM}118_{abcde}$	$HsaB_{LM}118_{abcde}$		$DreBA_{LM}118_{abcd}$	$OlaBA_{LM}118_{b}$	$TniBA_{LM}118_{bc}$			
119					$DreBA_{LM}119$	$OlaBA_{LM}119$				
120					$DreBA_{LM}120$	$OlaBA_{LM}120$				
121						$OlaBA_{LM}121$	$TniBA_{LM}121$			
122						$OlaBA_{LM}122$	$TniBA_{LM}122$			
123						$OlaBA_{LM}123$	$TniBA_{LM}123$			
124						$OlaBA_{LM}124$	$TniBA_{LM}124$			
125						$OlaBA_{LM}125$	$TniBA_{LM}125$			
126						$OlaBA_{LM}126$	$TniBA_{LM}126$			
127					$DreBA_{LM}127$	$OlaBA_{LM}127$	$TniBA_{LM}127$			
128			$HsaB_{LM}128$			$OlaBA_{LM}128$				
129			$HsaB_{LM}129$		$DreBA_{LM}129$					
130		$LmeB_{LM}130$	$HsaB_{LM}130$							
131			$HsaB_{LM}131$							$TniBB_{KM}131$
132			$HsaB_{LM}132$		$DreBA_{LM}132$					
133		$LmeB_{LM}133$	$HsaB_{LM}133$							
134		$LmeB_{LM}134_{acd}$	$HsaB_{LM}134_{abcd}$		$DreBA_{LM}134_{abcd}$		$TniBA_{LM}134_{bc}$			
135		$LmeB_{LM}135$	$HsaB_{LM}135$			$OlaBA_{LM}135$				
136						$OlaBA_{MN}136$	$TniBA_{MN}136$			
137						$OlaBA_{MN}137$	$TniBA_{MN}137$			
138						$OlaBA_{MN}138$	$TniBA_{MN}138$			
139	$XtrB_{MN}139_{bc}$	$LmeB_{MN}139_{bcd}$	$HsaB_{MN}139_{abcd}$		$DreBA_{MN}139_{bcde}$	$OlaBA_{MN}139_{acde}$	$TniBA_{MN}139_{acde}$			
140	$XtrB_{MN}140_{b}$	$LmeB_{MN}140_{bc}$	$HsaB_{MN}140_{bc}$		$DreBA_{MN}140_{abc}$	$OlaBA_{MN}140_{abc}$	$TniBA_{MN}140_{abc}$			
141		$LmeB_{MN}141$	$HsaB_{MN}141$		$DreBA_{MN}141$	$OlaBA_{MN}141$	$TniBA_{MN}141$			
142						$OlaBA_{MN}142$	$TniBA_{MN}142$			
143						$OlaBA_{MN}143$	$TniBA_{MN}143$			
144					$DreBA_{MN}144$		$TniBA_{MN}144$			
145					$DreBA_{MN}145$		$TniBA_{MN}145$			
146					$DreBA_{MN}146$		$TniBA_{MN}146$			
147					$DreBA_{NO}147$	$OlaBA_{NO}147$				
148					$DreBA_{NO}148$	$OlaBA_{NO}148$				
149					$DreBA_{NO}149$	$OlaBA_{NO}149$				
150		$LmeB_{MN}150$	$HsaB_{MN}150$							
151		$LmeB_{MN}151$	$HsaB_{MN}151$							
152		$LmeB_{MN}152$	$HsaB_{MN}152$							
153		$LmeB_{MN}153$	$HsaB_{MN}153$							
154		$LmeB_{NO}154_{b}$	$HsaB_{NO}154_{abc}$		$DreBA_{NO}154_{bc}$	$OlaBA_{NO}154_{ab}$	$TniBA_{NO}154_{ab}$			
155					$DreBA_{NO}155$	$OlaBA_{NO}155$				
156						$OlaBA_{NO}156$	$TniBA_{NO}156$			
157					$DreBA_{NO}157$		$TniBA_{NO}157$			
158					$DreBA_{NO}158$		$TniBA_{NO}158$			
159			$HsaB_{NO}159$						$OlaBB_{MO}159$	
160			$HsaB_{NO}160$			$OlaBA_{NO}160$				
161	$XtrB_{NO}161_{ab}$	$LmeB_{NO}161_{a}$	$HsaB_{NO}161_{abc}$		$DreBA_{NO}161_{abc}$	$OlaBA_{NO}161_{abc}$	$TniBA_{NO}161_{ab}$	$DreBB_{KO}161_{bc}$	$OlaBB_{MO}161_{bc}$	$TniBB_{MO}161_{bc}$

Xenopus tropicalis western clawed frog Hox B

PFC	Dist. To 3' gene	Length	Sequence
$XtrB_{CG}22$	95,671	23	CCACTGCAACTCACTAACCTTTA
$XtrB_{CG}35$	11,212	22	GGCCCTAGAGGAGTGTAGGGGC
$XtrB_{CG}36$	3,460	22	GTAAATATGCTTTGCGCTTATT
$XtrB_{CG}37$	1,248	22	AGTACAGAATGGCAACACTCAG
$XtrB_{CG}38$	179	116	CCGATTGATGAATAGCCCTCTGACGCTTTATCAGGCACACGG AGAAAGTTTGACCAATCATTTGCCAGGAGCTCACACGCCAGC AGCCCAAGTGCACCGAGTTGGCGAGCAAGTTG
$XtrB_{GH}50_{abcdef}$	9,020	352	CTCGTAAACTCCTAACAGCTTTTATATGCTGCGGTATAAACAA CATCAACATCTCGGCTTTATGGCGTTTTATAGTTCGCTAAACA CTTTACGGCGTTTTTTTGGCGATTACAAACGCAATTCCTTTCAT CGGAATAGTTAAACCTTTATCAATAATAGGGGGGGAAATTGA TCATTAAAATGTTAAATATGACTGGGCTCGTTTGTTTTAAAAT ATGTTTCAGCCTAAGAAACGGTATGATTTGATAACTTGTA
$XtrB_{GH}51_{ab}$	8,352	30	TGATTTGTTGTAAGGAGAGAGATCTCATTA
$XtrB_{GH}52_{bcde}$	8,209	153	CCAATAGGAGCAGGCTTTGCTAATAGATGCAAACGTTCCTGA AAAGACACAGCAAAATATGAAAGAACTCATTTGCTGGGAAG TAAATCACAGAAAACTGTTTATGAACTGGCAACCCTTCTTGG AAATGTAAAGCGAGAAAGTCTTTAAGTG
$XtrB_{GH}60_{a}$	7,472	17	CAACATGAAACTGCCTA
$XtrB_{U}65_{abc}$	5,298	179	AGATGGCGGCGGAGGAAAGGCTGCACTTAAAGCAGTCATGA AGAAATGCAATAAATTCCTTGTTGTTTTATGAAAATTTACAAC TTTGTGTTAGAAGTTTATGAGTGGCTGGGTCCTGGGATTGGCC AGGGCCGGTCATGTGGACTTGTAGCCGTGAACATGAACTTTTC TCATTTCCC
$XtrB_{U}70_{ab}$	90	61	CCAATCTCCGATAAACTACTAATAGCTAAAGCACTTGGACTA TAAAACACAACAAATCATA
$XtrB_{JK}79_{ab}$	250	192	CCTAACGATTCTCAGATCGTCATTATTTGTAACCATAGAGCAT GGATTACCTCTTGAGGTCATCAGTGAGAATTTACGACTGGTCA ACAAAAGCACGTGATTTCCAAACGCACCCACACCCCCATATT TGGCCGCATACATAGCAAAAAGCAAGTACAGTGTATGGGGAT AATTCATTAATACATCATAAAT
$XtrB_{KL}87_{bcdef}$	11,748	193	TGACATTTCCATGTCAAATGGATTGGGTTTTATGTAGAAGTTC CATGGTACAAATGGCCCAGACCTCAGGCTCAGACCCCCTCAC TGGCTCGCAAAAGTCACGTGTGCTCCATAAAGTTAGTTTTATG GTTTTAGGGAGTTGACAATGTACAATATATTTCACAATCTCCA GAATGTTAAGTGACAGTTTAAC
$XtrB_{KL}110_{abc}$	18,135	71	CAGATTTGGAGAAGATAGATCTATGTTTCATCTTTAATCACGC CAAACCCTGGCCCATTTGTCATGTTTAC
$XtrB_{KL}113_{ab}$	17,291	156	ACAGATGTTCCTCGGTAAATCCAGGTTTTCCCTGTGTTTTTATG TTACAAAGGCACCGGCCACTGCAATAGCTCAAACTCTGACAA GCAAAAAGATAATCAAAATGACAAATGAGTTCTTTTGTGAGA CACAGATCTTGTATTAATTTTCATTTTC
$XtrB_{KL}114_{b}$	16,696	39	AAGAAAAATGAGAGAATTATACAGAAAATCATTAATCAC
$XtrB_{LM}118_{abcde}$	13,264	362	GGTCATAGTGGTCCTGAGTGGATGGATGTCAGCCCCCCCATC CCCCCACCCCTGTCACGTGATTCATTAAATAATTAATGCAGA GGTTGCAGGATAGATATGTATTCGTCAGGAAAATCGCAGACA TGGTCTCCTTGCGTCTGACGTGAAATTCAACTGTCCTTAAAAA ACAGCCTATAGCAGAGACAAGAAAGAGAGAGAGAGGGAGA GAGAGAGAGAGAGAGAGGCTGCCATAGAGAGAGTGGGATTC GAGCC
$XtrB_{MN}139_{bc}$	4,026	20	TTAACGTGAATACAGGGTAT
$XtrB_{MN}140_{b}$	3,795	19	TTGCACTTACAGTTTACAT
$XtrB_{NO}161_{ab}$	266	39	TCAGATTGATGGGCCGGGTTTGATTGAAGTCTCTTTGTC

Latimeria menadoensis coelacanth Hox B

PFC	Dist. To 3' gene	Length	Sequence
$LmeB_{CF}23$	78,400	19	TATTTACACTTTATAAAAA
$LmeB_{CF}24$	69,457	69	TAAAGGAGGCATTAAAAAACAAATTTGAGACGGGGAAAAAGAAGAATTGGTTGTATTGTAAATGAAGTA
$LmeB_{CF}25$	65,545	39	TTGATTATTAATTGGATACAGATACAATGTTATGCCTGA
$LmeB_{CF}26$	64,023	18	GAGCTCAATGTCAATAAG
$LmeB_{CF}29$	36,877	22	GAAAATATAGAAAATGGCTGCT
$LmeB_{CF}31$	34,567	17	CACTGTGTAGTGGCCTG
$LmeB_{CF}32$	10,740	24	GTCAGTTGGATTCTTCCTCTTTCT
$LmeB_{FG}33$	14,696	26	AAAACCTATGCAAGTTCATTTTCAAA
$LmeB_{CF}34$	5,467	67	TTAAATATGTTTTATAATAAATATGCAATTATGCATAAACATGGTGAATTGTCTCAAGTCATTTGAT
$LmeB_{GH}50_{abc}$	916	338	CTCGTAAAATTCTAACAGCCTTTTTTGTAGTATGATATAAACAACAACAGGCCTGGCGTTTATTGCACTTTATAGTTTGTTAAATACTTTACTGCCTTTTTGGTTATTACAACTGCACTTCCCCCTCATGTAATAGTTAAACCTTTATCAATAATAAGGAAATTGATCATTAAAATGTTGGGTATGAACACCTCGTTTGTTTTAAAATATGCTTTAACCGGGAAGTTGCATGATTCCATAAAACGTATTAAAATA
$LmeB_{GH}51_{abcd}$	461	74	TGATTTGTTGTAAAGCACCAGATATCTCATTAATGATGGTGGTCGAATATATTATGTATTATATTCTCCCCTCC
$LmeB_{GH}52_{abcde}$	309	173	GTCCGTGTGGAGTAAGAGAAGCCAATAGGATTTGGTTTTGGTAATAGATGCAAATGATCATGAAAAGACACTGCAAAATATGAAACAACTCATTTGCGGTGAAGTAAATCACAGAAAACTGTTTATGAACTGGCACCTCTTCTTGGAAATGTAATGCGAGAACTCTTTAAGTG
$LmeB_{HI}60_{ab}$	1,643	24	CAACATGAAACTGCCTATTTATGC
$LmeB_{HI}62_{abc}$	1,278	227	AGCATGCGCACATTGCAAAGATCAGTTGTAAAGTGATCCTATAAGTTTACTGTGCAAAAGAATGTATACTGTATACGTGAATTACTTTATGAGGGACCTGATGATATATTTTATTGTTCTTCACCACTTTGTTCTATTGTCTAAACCTGGAAACGGCGGAAGAAAGGGGTACAATAAAAGTTTACAAGCGAGAATCCGTGACATCATTGCTTTGCTCTGCTTCATTT
$LmeB_{HI}63_{ab}$	236	236	ATTATTTATCTTGATTCTATAAGTCCGACCGTATTACAATTTTTCCACATTACAGAGCAACGGGACGCCCAGCTTTCATTGGATCTCTTAGAATCACGTGTCGAAAAATCACGTGGTCCCGGAGGAAAAAGGGGGTCCTTTTTGGTGTAAATCTGGACTCTAATTCCGTAATATATCACGGTACCTCGTAAAACCGACACTAAAACGTCCCGGCCTACAAATCACCCAGCCAAATT
$LmeB_{IJ}65_{abcd}$	6,670	208	AGATGGCGGCTGAAAGAAGACTGCTGTTAAAGCAGTCATGAAGAAATGCAATAAATTCCTTGTTGTTTTATGAAAATTTACAACTTTGTGATAGAACTTTATGAGTGGCTGGGTACTGGGATTGGCCAGGGCCGGTCATGTGGACGGGTAACCGTGAACATGAACTTTTATGATTTCCCATGTGGTTATATTGCAGCATTCTTTTGG
$LmeB_{IJ}66$	5,523	16	ATTTTTAAATATTAAA
$LmeB_{IJ}70_{ab}$	91	61	CCAATCTCCGATAAACTACTAATAGCTAAACCACTTGGACTATAAAACACAACAAATCATA
$LmeB_{IJ}71$	2,270	68	ACTACCTAGAAGGACGCAGAAGGTTATAGTTCAATAATTTTGTATGTGACTGTGTTTGTAGTTTGTCT
$LmeB_{JK}79_{abc}$	252	231	CCTAACGATTCTCAGCTCGTCATTATTTGTAACCATAGAGCATGAATTACCTCTTGAGGTCATCAGTGAGAATTTACGACTGGTCAACAAAAGCACGTGATTCCCAAACGCACCCCCACCCCCATATTTGGCCGCATACATAGCAAAAACGAAGTACAGTGCATTGCTATAATTCATTAATACATCATAAATCGTGAAGCACGGGGGTTATAACGACCAAGATCCACAAATC
$LmeB_{JK}80$	15,285	36	GCATGTTTTATGTGAGATGTTATAACTTATTTATAA
$LmeB_{KL}87_{acdef}$	13,956	203	AAAAGTAGGTATGGCATTTAGATGTCAAATGGATTGGGTTTTATCTAGAAGTTAGATCGTAAAAATCGCACGGACCTCAGACAGATATCCCTCACTGGCTCTCAAAAGTCACGTGGGATCCATAAAGTTAGTTTTATGGTTTTGGGGAGTTGACAATGTACAATATTTCACATTCTCTAGAATGTTAAGTGACACTTTAAC
$LmeB_{KL}90_{abcd}$	9,749	85	ATAATGTGGTTTGCATATTGAAAAGGATAATCTATGCTCATATCTCATCAATAATTCATAGGAAACCGGGATCATTCTGAGGTCA
$LmeB_{KL}100$	8,594	50	CTGTAACACCACAAACCACCAGCAGAGCCCGCTTCAGACCAAGTTCACAG
$LmeB_{KL}101$	5,255	52	ATGGAATTTCTTAACCCGTGACCACGATTGATTGATGAATATTTGGACTTTT
$LmeB_{KL}102_{ab}$	4,488	196	AAGTAATGACCTGGGCAAAATTCAATATGACCGAGCATGCTTTGCGACTGCATTATGGAAGCTGTGAGTGGGAGAGCCCGGGAGAGGGGTGAAACGCAGGTCAGCGCGTCTAACAAATATTAAAATGTTCAGAGGACGCATGACACGCCTAGCTGTTTAACAAAGACTGCCAAAGTATGAGATTAATTCGAAAACT
$LmeB_{KL}108$	60	59	AAACGAGTAAGGGTTAGGAATAAATTTTAGTATATTTTTGTGTGCAATTCAAAGAAATT
$LmeB_{LM}110_{abc}$	22,843	71	CAGATTTGGCGAAGATGGATCCTTGTTTCATCTTTAATCACACCAAACTCTGCCCCATTTGTCATGTTTAC
$LmeB_{LM}111$	22,001	26	TGTACATAGTGGTAATAGAAATAAAT
$LmeB_{LM}112$	21,871	64	TTGTGTTTAGCAAAGCTGATTTATACATAGGCATGGGTCTTTAATAACAAAGGACATGTGTGGT
$LmeB_{LM}113_{ab}$	21,521	153	ACAGATGTTCCTTGGTGTTTGCAGGTTTTGTGTGTATTTTTTATACCACAAAGGAGCCACTGCAATAGCTCAAACTCTGACAAGCAAATAGATAATCAAGAAGACAAATGGCTTCTTTTGTGAGACGCAGCTTTTGTATTAATTTTCATTTTC
$LmeB_{LM}114_{ab}$	20,567	71	AAAGTTCACAGCCATTATTTGTAGACAGAGCCAAGAAAAATGCGAGAATTATACAGAAAATCATTAATCAC
$LmeB_{LM}116$	19,673	78	CACGTTTTCAAACCGATTACCTCAGTTCTATAGTAACAGACTGACCTCTGAAACCCTTGACCTTTTTTAGACTGGACA
$LmeB_{LM}118_{abcde}$	16,226	356	GGTCATAGTGGTCCTTTGAGTATGGAACAGAGAGACCCCCTTCCCCCTCCTGATCACGTGATTCATTAAATAATTAATGCAGAGGTTGCAGAATAGATATGTATTCGTCAGGAAAATCGCAGACATGGTCTCCTTGTGTCTGACGTGAAATTCAACTGTCCTTTTAAAAACAAGCCTATAGCAGAGAAAAAGAGAGAGGGAGAAAGAGAGGCTGCAATAAAACAATTTTGGAAGCAGAGCCACACCGTGGGCTGG
$LmeB_{LM}130$	7,572	111	TTTTATGTATTTGGAGGCAATTCAAAACTACTTCCGTTGATTGATCATTAATCTGGTGAGTTATTACTCTTCTAAGCAAAGGTCATTTGGAGAAAGAGCTCTAATCTGTCA
$LmeB_{LM}133$	3,335	51	CAGAGGCCATTGTTATCAGTGAGCAGGGCGTACCAGTCTATAGGGCAACCA
$LmeB_{LM}134_{acd}$	3,041	132	CTCGTCAGAGGCTAAGGTAAGCTGGACTCAAATAGGCTATCACTTTGTGAATGGCGGAGTATATGTCCCAGTGATTTATGACCATATGACTTCAATCTCGGTTCAAGAAGAGTTCACAAGCTTTAAGCTTCC
$LmeB_{LM}135$	202	41	CCTATTCATGTCATGTGCTTTTCACTATATAAGATGGAT
$LmeB_{MN}139_{bcd}$	3,907	38	TCTTGACTTAACGTGAAAACAGGGTATCTTTGAACAAA
$LmeB_{MN}140_{bc}$	3,794	217	TTGCACTTAGAGTTTACATTTGATGGGTGAAACAACCTTGAAAGGCGTTTAGCCAGTTCAATCGTGACGTCTTTCGATACGTGTGTGTTCTGGTGAACATTCATATATATTTATTGGTTATAGCCAGTTTAAATATTTTCTTTTGTATTATTTATCCCCAAACATTATGTATTTATATTAAGAAAGAGACAAACTGTACTTTTTTAGTATTTACCTG
$LmeB_{MN}141$	3,325	17	AACAATTCTTGAATAAA
$LmeB_{MN}150$	2,960	53	TCACGTGTTGTCCAGAGTGAACCTTTATTGGACCTAAAAGGAAAACAAAGAAA
$LmeB_{MN}151$	2,500	36	AATTGTATACAATGTTTGCGTTTTGGCGTTTAATTA
$LmeB_{MN}152$	2,253	85	TATAATTTAAAAGCCATAAGAGCGTGGAATGTTTGACTGGGATCAATTACCAATACTCAAGGGTTTTTCTGAATACAATTTTCTG
$LmeB_{MN}153$	250	183	CTCTGGAGTTGTAGATCGATTGAAGAAAATAATGCTTTCTTTTTTCCCCCACCAGGCCCTAATTTGCCATTGGCCAAAGATAGTCATGTGACCAGAAATTGGATGCAATTTCGTCCTAGTCTCCAGTTTAATAGAGCAAGGTCCCCAAACGCTTGTAATATCACCAATATACAATTATAAAG
$LmeB_{NO}154_b$	19,653	21	TATTTATTGTATATTATTTTC
$LmeB_{NO}161_a$	264	64	TCAGATTGATGGGCAGGGTTTGATTGAAATCCCTTTGTCATGCAAATGTCAAGCACTTGATGGA

Homo sapiens human Hox B

PFC	Dist to 3' gene	Length	Sequence
$HsaB_{CG}22$	96,213	23	CCACTGGGAGGCTGAAACCTTTA
$HsaB_{CG}23$	95,483	28	TATTTATACAATAAAATTTTTTTAAAAA
$HsaB_{CG}24$	84,917	159	TAAAGGGGAGATCTAAGCAAATGATATGAACGGTCAGGTCAAG AATTGGCCACACAATTGAACGAAGTACATTTTTAGCATGTTGGT AGATATCCCATTGAAAATGCTTGGCATGGAAGCATGGGAATTG AGGAAACTGGTTAGAGATCCCCCTACGGA
$HsaB_{CG}25$	80,145	36	TTGATTTCCCAAGGGCAGGTACCCTTGGCAGCCTGA
$HsaB_{CG}26$	78,392	21	GAGCTCGTTGACTTTAATAAG
$HsaB_{CG}28$	51,295	31	CATAAAGAACTTTCTAGTGCATCCTTGGGCC
$HsaB_{CG}29$	46,437	29	GAAAATTTTTAAAAGCATCTTTTGCTGCT
$HsaB_{CG}30$	44,661	21	CTTATTGATGAGTTGGATAAA
$HsaB_{CG}31$	43,671	16	CACTGTGCCTGGCCTG
$HsaB_{CG}32$	14,496	29	GTCAGTTTACGTAGGTCTTCACCCTTTCT
$HsaB_{CG}33$	30,201	89	TTAAATTTAATTTAAAAAAACAGGATACATTTCAGTAATCCATT AAACATCAGGGGTGAATTGTAACAGAATCCAATTCCATATTTGA T
$HsaB_{CG}34$	14,496	29	GTCAGTTTACGTAGGTCTTCACCCTTTCT
$HsaB_{CG}35$	11,355	23	GGCCCTGGACGTTTCTGAGGGGC
$HsaB_{CG}36$	3,545	22	GTAAATTTCAGCTAAGCTTATT
$HsaB_{CG}37$	1,319	24	AGTACAGGAGGGCTGAAAACTCAG
$HsaB_{CG}38$*	264	118	CCGATTGATTTATGTCGGAGCTGACGCTTTATCAGGCAGTCGGA AAAACTTTGACCAATCATTTTGCAAGGAGAGCTGAGACGGGCT GCTCCACTGTACTTTGTTGGCTGAGAAGTTG
$HsaB_{GH}43$	8,164	15	TCTTCCCCATCTCAC
$HsaB_{GH}44$	5,323	19	TCATTAGGGATGCAAAATG
$HsaB_{GH}45$*	4,945	84	GCAATAAAATAATATGACCGCTATAAAAGTTTATAGCGTATAA ATTTCTGAAGGTTAAGAAACTAAACAGCAGCAAAGCAAACA
$HsaB_{GH}46$	4,502	38	AGACGTCTGGGCGCGGGCAGCCTCTTGATTCTTTTACA
$HsaB_{GH}49$	2,390	35	TTTGTAAAGCCTCCACTTTTGTTGCTCAAAATTTT
$HsaB_{GH}50_{abcdef}$*	1,227	342	CTCGTAAAATGCTAACAGCTTTTATGCGCTGCGGCATAAACAAC AACAGACTCCGGCTTTATTGCGTTTTATAGTTTCTTAAAGAGTTT ACAGCCGTTTTTTGGGGGGCCGGTTATCCAGCCAATTCCCTCCA CACGATAGTTAAACCTTTATCAATAATAAGGAAATTGATCATTA AAGCTTTAAATATGACTACCTCGTTTGTTTGAAAATACGTTTAGG AGAGGAAGCTGTATGATTTGATAACTTGTATTA
$HsaB_{GH}51_{abc}$	748	71	TGATTTGTTGTGAGGCGAGAGATATCTCATTAATGTAGGTAGTTA AACAGATTTAATCCGTATTCATTCTC
$HsaB_{GH}52_{bcde}$*	480	153	CCAATAGGATGCGGGGTCTCTAATGGATGCAAATGATCATGAA AAGACAGTGCAAAATATGAAACAACTCATTTGCAGGGAAGTAA ATCACCGAAAACTGTTTATGAACTGGCATCCCTTCTTCGAAATG TAAAGCGAGGACCTCTTTAAGTG
$HsaB_{HI}57$*	1,889	34	ATTTCTGCTAAGTTCTCCCAACAACATGAAACTG
$HsaB_{HI}60_{a}$	1,868	17	CAACATGAAACTGCCTA
$HsaB_{HI}62_{abc}$	1,623	223	AGCATGCGCGCTGTGGGCAATTGTTACAAGTGTTCTTAGGTTTA CTGTGAAGAGAATGTATTCTGTATCCGTGAATTGCTTTATGGGGG GGAGGGAGGGCTAATTATATATTTTGTTGTTCCTCTATACTTTGT TCTGTTGTCTGCGCCTGAAAAGGGCGGAAGAGTTACAATAAAGT TTACAAGCGAGAACCCGAGACTGGCCCGGCCAGCGCTCCTCAT TT
$HsaB_{HI}63_{ab}$	238	238	ATTATTGATCATATTTTATAAATCCAACGCCACACAATTTTTTCC ACATTACCGGGAGCCGTGGGGAGACGGCCCGGCCATTGGCGGA GGGGACGTCACGTGGGCGGGGTCACGTGGTCCGGAGAGGAAAA AGGGGGTCCTTTTTGGTGTAAATCTGGACTCTAATTCTGTAATAT ATCAAGGAATCTCGTAAAACCGACACTAAAACGTCCCTGCCTA CAAATCATCCGGCCAAATT
$HsaB_{IJ}64$	9,547	23	TTAAATGAAAGGAGTTTAAAAAC
$HsaB_{IJ}65_{abcd}$*	7,032	209	AGATGGCGACTGAGAAAAGGGTTGCTGGTGGAGCAGCCATGAA GAAATGCAATAAATTCCTTGTTGTTTTATGAAAATTTACAACTTT GTGATAGAAGTTTATGAGTGGTTGAATCCAGCGATTGGCCGGCG CCGGTCATGTGGCCGGGCGATCGTGAACATGAACTTTTTATCAT TTCCCTGGTGGTTATAATGCAGCATTCTTTTGG
$HsaB_{IJ}66$	5,717	15	ATTTTTGAAATTAAA
$HsaB_{IJ}70_{ab}$*	103	61	CCAATCTCGGATATACTACTAATAGGCGCGGCGCTCGGACTATA AAACACAACAAATCATA
$HsaB_{IJ}71$	2,259	87	ACTACCTAGAAGTAAGAAGAGGAGCCTCAGAAGAAAACAAAG TTCTATTTTATTAATTTTCTATGTGTTGTGTTTGTAGTCTTGTCT
$HsaB_{JK}79_{abc}$*	267	246	CCTAACGATTCTCGGATCGTCATTATTTGTAACCATAGAGCATG AATTACCTCTTGAGGTCATCAGCGAGAATTTACGACTGGTCAAC AAAAGCACGTGATTCCCTAACGCCCCCCACCCCCCTTCCAACC CCCCCCCCATATTTGGCCGCATACATAGCAAAACGAAGTACAG TGCATCGCTATAATTCATTAATACATCATAAATCGTGAAGCACA GGGTTATAACGACCACGATCCACAAATC
$HsaB_{KL}80$	13,558	53	GCATGTTTAATGTGAACTCTCCCCTCCCCATCTGTGTTCTAACTT ATTTATAA
$HsaB_{KL}84$	13,483	19	GTTTCAGCTGGAAACTTCT
$HsaB_{KL}87_{abcdef}$*	12,115	204	AAAAGTCGCTATGACATTTAGATGTCAAATGGATAGGGGTTTTA TCTCGAAGTTAGATCGTAAAAATCGCCGAGAAGTCAGACAGAT ACCCCTCACTGGCTCGAGAAAGTCACGTGAGGTCCATAAAGTT AGTTTTATGGTTTTGGGGAGTTGACACCGCGCAGTATATTTCACA TTCTCCAGAATGTTAAGTGACACTTTAAC
$HsaB_{KL}88$	12,418	40	CCAACTGATCTTCCCTCTTGATTAGGAGTTAGGGTCCTTT
$HsaB_{KL}89$	11,577	35	GGGAAAAGGGTTTGTGTGGGGGATCCATGCTCCCT
$HsaB_{KL}96$	9,043	18	CATTTTTGGAGTTTTTCA
$HsaB_{KL}90_{abcd}$	8,297	84	ATAATGGAGTTTGCATCCTGAAAGGGGAAATCAACGCTCATATC TCATCAATAATTCATAGAGTCCGGGATCATGCAGAGGTCA
$HsaB_{KL}98$*	8,288	84	TTTGCATCCTGAAAGGGGAAATCAACGCTCATATCTCATCAATA ATTCATAGAGTCCGGGATCATGCAGAGGTCAGCAGACGGG
$HsaB_{KL}99$	7,273	23	CGCTGCTCTGGACCAGAGGCAGA
$HsaB_{KL}100$*	7,248	51	CTGTAACATCCCGGAGCTGCCAGTAGAGTCCGCCTTAGACCAA GTTCACAG
$HsaB_{KL}101$	4,511	55	ATGGAAAATCAAAACAGGGGCGTGGCCGACTGACTCCTATCCG AATGGGACTTTT
$HsaB_{KL}102_{ab}$*	4,357	200	AAGTAATGACCTACGCAAAATTCAATATGACCCAGCGAACTGC GCGAGCATATTATAGTAACTGCCTGCTCGTGGGGGAGGCTGGGA GAGAGGTGAACCGCAGGTCACGGCGTCTAAAAATTATTAAAAT GTTTGAGAGCCTCGTGACGCGCCTAGCTGTTTAACAAAGACTGC CAAAGTATGAGATTAACACGAAAACT
$HsaB_{KL}104_{bcde}$	1,606	69	TACCCTGTAGATCCGAATTTGTGTAAGGAATTTTGTGGTCACAA ATTCGTATCTAGGGGAATATGTAGT
$HsaB_{KL}107$	1,472	34	TTTCTGAATGAGGACAGTCTGGTGACTGGCCACA
$HsaB_{KL}108$*	59	58	AAACGAGTCAGGGGTCGGAATAAATTTTAGTATATTTTGTGGGC AATTCCCAGAAATT
$HsaB_{LM}110_{b}$*	23,888	13	CATCTTTAATCAC
$HsaB_{LM}111$	23,080	26	TGTACAGAGTGACAATAGAAATAAAT
$HsaB_{LM}112$	22,942	90	TTGTGTGTGGTTGAAGATTTTTTTCAAAGAGTCTGATGGGAATTT TTATCTCAGAGGACTCCAAATTGTGGTGGCTGGTTTGTTTTCCTT
$HsaB_{LM}113_{ab}$*	22,517	156	ACAAGATGTTTCTGGGTGTTTGCAGGTTTTCAGAGTATTTTTATATT ACAAAGAAGCTAGCCAGTGCCATAGCTCAAACTCTGACAAGCA AATAGATAATCAAGAAGACAAATGGCCTCTTTTGTGAAGCCTTG CTCCAGTATTAATTTTCATTTTC
$HsaB_{LM}114_{ab}$*	21,660	72	AAAGTTCACAGCCATTCTGTGTAGACAAGAGCTAAGAAAAATG TGAGAATTATACAGAAAACCATTAATCAC
$HsaB_{LM}115$	21,308	20	CACAACTGAAAGAAGAAACA
$HsaB_{LM}116$	20,477	79	CACGTTTTCTAGGCGATTAGCTCAGTATTAGAACCACAAAATGA CCTCGGAAGCCCCTTGACCTCGCTGAGACTGGACA
$HsaB_{LM}117$	19,434	26	TTGGTTTGCATTTTGTAAATTGTTTC
$HsaB_{LM}118_{abcde}$*	16,792	393	GGTCATAGTGGTCCTTTGAGCCTGAAATCGAAAGACCCCCTTCC CCCTCCCGGTCACGTGATTCATTAAATAATTAATGCCGAGGTTG CAGAATAGATATGTATTCGTCAGGAAAATCGCAGGCTCGGTCTC CCTGTTTCTGACGTGCAATTCAACTGTCCTTTGAAAAACAAGCC TGTACCCCGAGAAAAAAAAAGAGAGGGGGGAGAGAGAGAGGG AGAGAGAGAGAGAGAAAGAAAGGGAGACAGCAATAAA
$HsaB_{LM}128$	12,557	22	TCCTCCCTCAAACTCCCAGCCT
$HsaB_{LM}129$	10,193	27	TAATTTAATTGATTTTATCATAAATCA
$HsaB_{LM}130$	7,873	108	TTTTATATTTGGAGGCAATTCGGAAGGCCTTTCGCTGATTGGTCA TTAATCCACTGAGTTATTGTCTGCTAAACAGAGGTCACTGGGAG AGAAACATCCAGTCTGTCA
$HsaB_{LM}131$	7,665	20	TGCAGCCTCCCTTCCTCTGC
$HsaB_{LM}132$*	4,597	36	TGCAGACACCTACATTTTTGGCTCCTGTCTTCCTCC
$HsaB_{LM}133$	3,473	51	CAGAGGCCATTGTTACGGAGAGTAGGGCGTACCAGTCTTATAGG GCAACCA
$HsaB_{LM}134_{abcd}$*	3,072	133	CTCGTCGGTGGCTAAGGTAAGCTGGACTGAAATAGGCTATCAGT TTGTGAATGGCGGAGAGTATGTCTCAATGATTTATGGCCCATAT GACTCCAATCTCGGTTCAAGAAGAGTTCACAAGCTTTAAGCTTC C
$HsaB_{LM}135$*	248	40	CCTATTGATGTCAGTTCCCTTTTCAGTTCCTAAGATGGAT
$HsaB_{MN}139_{abcd}$*	4,581	38	TCTTGACTTAACGTGAAAACAGGGTATATTTGAACAAA
$HsaB_{MN}140_{bc}$*	4,478	232	TTGCACTTAGAATTTACATTTTAATGGATGTAAAAACAACTGTG AGAGATGTCTGGGCCTGCAGAAGTCCAGCATTGCTCAAAAAAG CGTGTGTTCTAGTGAACATTTTCATATATATTTTATTGGTTATAGC CTGTTAAAATATTTTCTTTTTTGTATTATTTATCCCCCTACATTAT GTATTTATATGAGGGAAAAAAAGGAAAAAATTGTACTTTTTTAG TATTTACCTG
$HsaB_{MN}141$*	3,994	17	AACAATTCTTGAATAAA
$HsaB_{MN}150$	3,495	52	TCACGTGACCCGAAGCCCAACCACCATTGGGTCTAAAATGAAA ACAAAGAAA
$HsaB_{MN}151$	3,013	33	AATTGTAAGCGATGTGCCCGCATTGCTTAATTA
$HsaB_{MN}152$*	2,732	107	TATAATTTAAAGGCATAAGAGCGTGCAAAGTTTGATTGGGATCA AATAACGCTCAAGGGTTTTTCTTTCTTTCTCTTTCTCTTCTTCC TTGAATAACATTTTCTG
$HsaB_{MN}153$*	328	180	CTCTGGTGTTTTTGAATCAATTAAACCAAATAATGCTCTCTGTTT TCCACCAGGCCCAGACGAGCGATTGGCGGAGGCCGGTCCCGTG ACCACGAATTCCCTGTAATTTCGCTGGAGTCCTGGGTTTAATAG AGAGAGTCCCCATACGCTTGTATTTATCAGCAATATACAATTAT AAAG
$HsaB_{NO}154_{abc}$*	11,844	73	TATTTATTAAAATTCTTTAATAATAGGAAAAGGGGAAAGTATTT ATTGTACATTATTTTCATAGATTAAATAA
$HsaB_{NO}159$	5,635	20	CCTTCAGGTTAGAAGAAAAT
$HsaB_{NO}160$	2,483	25	CCTCCTCCCATCTCAATTTTCCTTT
$HsaB_{NO}161_{abc}$*	289	64	TCAGATGGATGGGCTGCGGGTGATTGAAGTGTCTTTGTCATGCT AATGCTTGGGGGGTGATGGA

Polypterus senegalus bichir Hox B

PFC	Dist to 3' gene	Length	Sequence
$PseB_{AC}2$	16,089	26	AAACAACGTCACCTTTCTGGAGAAAT
$PseB_{AC}3$	5,963	28	AATTAAATTAGTCAATACGATTAGTGCT
$PseB_{AC}4$	3,355	22	TTTAATGTCACGTTGAATACAT
$PseB_{CF}16$	-107	14	TTTTTGCTTTTGTT
$PseB_{CF}17$	-1,106	27	AAATCAGTGTTTAACCTTGACCTTGAA
$PseB_{CF}18$	7,372	27	TTTGTTACGTTACTGCGGTTCATCAGT
$PseB_{CF}19$	336	24	TTGAAATGGTCAATGTGTACATTT
$PseB_{CF}20$	53	53	TTGAAAAGTCAATTGAGAAGAACCCTTTGCCTTCGTTAAAATATGTCCCAGCA
$PseB_{FH}21$	in 10	16	TGATCTGATTCATATT
$PseB_{FH}42$	2,782	26	ATGAAACCTGTATTTGAGAATGCAGA
$PseB_{FH}49$	2,104	34	TTTGTAATGTTTCATTAAGTTTCCGCAGAATTTT
$PseB_{FH}50_{abef}$	986	342	CTCGTAAAAATTTAACAGCTTGTGTTCCTTTGCATTCCTGTATAAACAACGCGCAACCTTGCCGTTTATGGGATTTGTTAAACCATTTACCGCTTTATTTTCTATTACGGCAGATACAATATCCACGGGTATACAACTTAAATATACATGAAACACATATTTATAACAAATCATTAAAACGTTAAATATGACCACCTCGTTTGCTTTATATAAACGTTTACTGGTGGAATACTGCGTTTTCATAAGATGTTCTA
$PseB_{FH}51_{abcd}$	528	74	TGATTTGATGCAAAGCGGCAGATATCTCATTAACTAACCCGGTTGAACATATTAATGCGCATATTCTCCCCTCC
$PseB_{FH}52_{acd}$	352	163	GTCCGTGTGGAAAACCTGAAGCCAATGGGATGGAGATCCGGTAATAGATGCAAATTATCATGAAAAGACTCTGCAAAATATGAAACAACTCATTTGCGGCGGAGTAAATCACAGAAAACTGTTTATGAACTGGCACCCCTTCTTGGAAATGCAATGCGAGAAC
$PseB_{HI}58$	1,944	27	TTGGTGTGTGAAGAAACAAAGAACTGA
$PseB_{HI}59$	1,815	36	AAATGTGAAACTTTTTATTAGAATGGTTTCTTAATT
$PseB_{HI}60_{ab}$	1,680	24	CAACATGAAACTGCCTATTTATGC
$PseB_{HI}61$	1,624	16	AAATCAAGATATTTTT
$PseB_{HI}62_{abc}$	1,586	219	AGCATGCGCACAGTTAAAAAAAATGTAACGTGACCCTCTGTTTACTGTGAAAAGAATGTATACTGTACTCGTGAATTACTTTATGGGGACTCCGTGATATACTCTGTTGTTTTGCGCTACTTTGTTCTATTGTCTAAACCTGGAAGCAGCGGAAGAAAGGGGCACAATAAAGTTTACAAGCGAGAATCCGTGACATCATTGCTGCGCTCTGGTTCATTT
$PseB_{HI}63_{ab}$	233	233	ATTATTAGACATATTTTATACATTTAAACTGTATTACAAATTTTCCACATTACTGGAGCTGCAGGGATCCCGAATTCTATTGGATCAGCGAGAAGCACGTGTCTTGTCACGTGTCTCCACAAGAAAAAGGGGTGCTTTTTGGTGTAAATCTGGACTCTAATTCCGTAATATATCACGGTACATCGTAAAACCGACACTAAAACGTCCAGACCTACAAATCACTCGGTCAAATT
$PseB_{IK}78$	779	25	CATTTGGCTGACTTTATATATGTTT
$PseB_{IK}79_{abc}$	252	231	CCTAACAATTCTCATCTCGTCATTATTTGTAACCATAGAGCATGAATTACCTCTTGAAGTCATCAGTGAGAATTTACGACTGGTCAACAAAAGCACGTGATTCCCTAACGCACCCCCACCCCCATATTTGGCCGCATACATAGCAAAAACGAAGTACAGTGCATTGCTATAATTCATTAATACATCATAAATCGTGAAGCACAGCGTTATAACGACCAAGATCTACAAATC
$PseB_{KL}83$	12,413	65	AACATTATTCCAAGTCTGGAGTACTGTAAAAGCACCTTTTGGCATGCTCGTGATGTTATAGGCAT
$PseB_{KL}84$	12,176	17	GTTTCATGTGAACTTCT
$PseB_{KL}85$	12,239	133	GTTGTTATCGTTATAAATCTTATGTGACAAGTGCGATGTTGATGCTTTCTTAAAAGTCGGCATGTTTCATGTGAACTTCTAAGTGTTATAACTTATTTCCAATCTATACCAACTGTATAATTTTAAGTTTACA
$PseB_{KL}86$	11,914	222	GTTCATGTGTGTATACTTTTACTCGAATGACAAATGTACAATATGTTATTATTCATGTGCGAAAAAGGTTGTAAATAATTATACATTTTTAATGTGATCAAGAATATTTGTAGTAAGCAGTGTGAATGACAAGCATTCTGTGGAAAATGGCATCTGTCATGTGCGACTGCCACTTAGTATTTTGCGAATAAGCTTTTTTGTATTTGTTTGTAGCTTAACTTG
$PseB_{KL}87_{abcdef}$	10,865	201	AAAAGTAGGTATGACATTTCGATGTCAAATGGATGAGGGTTTTATCTAGAAGTTAGATCGTAAAAATCGCCCAGACCATAGACAGATACCCCTCACTGGCTCTCAAAAGTCACGTGGGGTCCATAAAGTTAGTTTTATGGTTTTGGGGAGTTGACAATGTACTATATATTTCCCGTTCTAGAATGTAAGTGACGGTTTAAC
$PseB_{KL}89$	10,359	31	GGGAAAAGATTTTTTGTCATTTATGCTCCCT
$PseB_{KL}90_{abcd}$	8,221	84	ATAATGTGGTTTGCATATTGATAGGAGTAATCTGCGCTCATATCTCATCAATAATTCATTGAAGCAGGGATCATTTCGAGGTCA
$PseB_{KL}91$	6,830	19	CAGAGCCCGCTTCAGACCA
$PseB_{KL}94$	5,703	39	AAGTTTATGTGGTGTGTATTCCTGTATACTATTATTATT
$PseB_{KL}95$	1,593	20	ATTTTTTTACTGTTTGTTCA
$PseB_{KL}102_{ab}$	3,824	195	AAGTAATGACCTGGGCAAAATTCAATATGACCGAGCAAGCGGTATGCATTACTATAGAAGTCGCAAGTGGGAGAGCCCTCTAGAAGGGGTGAAACGCAGGTCAGCGCGTCTAACAAATATTAAAATGTACTGGGATGCATGACACGCCTAGCTGTTTAACAAAGACTGCCAAAGTATTAGATTAAACT
$PseB_{KL}104_{abcde}$	1,219	77	TCTATATATACCCTGTAGATCCGAATTTGTGTGAACAGAGAAGCGGTCACAAATTCGTATCTAGGGGAGTATGTAGT
$PseB_{KL}105_{abc}$	621	75	GTGATTTAGGAGCTTGGTATCCCGAATTAGTTGATGAATTTTTTATCGATCCAAACAAGCCCAGATTTATCTCTG
$PseB_{KL}106$	441	13	TTTTCATGAACAC
$PseB_{KL}108$	61	60	AAACGAGTAAAGGGATAGAAATAAATTTTAGTATATTTTTGTGTGCAATTCAAAGAAATT

Danio rerio zebrafish Hox Ba

PFC	Dist to 3' gene	Length	Sequence
DreBA$_{AC}$1	2,734	42	AAACAAGTACACATTTAATCCGTTAAGAAACTTATATTCCAG
DreBA$_{FG}$6	1,621	12	AGAATATTAACA
DreBA$_{CF}$8	12,954	14	ATTAATCTTATCAC
DreBA$_{CF}$9	10,950	19	TCACAGTTGAACTTGAGAC
DreBA$_{CF}$10	4,962	25	CTTACAGTCCTAATTAAATATTTAA
DreBA$_{CF}$11	4,527	13	CATGAAATTCACA
DreBA$_{CF}$12	2,335	23	GTTAGTCAGAAAATGGCATTCAG
DreBA$_{CF}$17	12,628	27	AAATCAGCTTTTAACCTTGACCTTGAA
DreBA$_{CF}$19	578	27	ATTGAAATAGTGTGTATAGTCACATTT
DreBA$_{CF}$20	268	55	TTGAAAAGCCAATTGAAAGGGACACTATTGCCTTGGCCCAAA ATATGCTCCAGCA
DreBA$_{FG}$27	1,863	30	TAGATATTTTTTCCAAGATACTAGCATTAT
DreBA$_{GH}$40	5,275	26	TGTAATAAATTAATGCATACAGAAAA
DreBA$_{GH}$45	4,705	83	GCAATAAAATAATATGACTGTAATAAAACTTTATAGGGTATAA ATTTCTGAAGGTTAAGAACTAAATGGCTGTAAAGCAAACA
DreBA$_{GH}$46	4,300	35	AGACGTCTGGGATGACCAACGTTAATTTGTTTACA
DreBA$_{GH}$47	2,445	14	CAGTCGCTCTTCAT
DreBA$_{GH}$48	1,591	311	AACATGCAGCACGTGTTTCTATATTTAAAAAGTAAAAAAGTAA ACAGGCTTTAGTTCGATGTGAGCGTTTTTTGTTTTGCATGTTTACT TAGATATATATTGGCATGATTTTTAAATACGAATCTAGTTTCAT TACAACAATTGGACATCTGCAATTTTACCCTCACCAGAACGCA CCACGCACAATGTTTAAGCTGTTATAGAAATGTAGCATCGTTT ATTTATTTATTTATTTATTTATTTATTTATTTATTTA
DreBA$_{GH}$50$_{bcf}$	1,068	331	TAACAGCTTTTTGATGCACTACGTCAACAACAAACAGTAGGGG CGTTTATTGCATTTTATAGTTTGTTAAACCATTTACAACCACATT TACCATTACACCTGACGCTCAGGTCCATTTCAACATCCCCCTCT TTCTGTAAATACATTAATAGAATATTGTTGTCAAAGAATTGAAT ACGACTCGGCCTACCTCGCTTATTTTTGACGCGCGGGCCACAG TCAGTTCGCTTATTCCTGTAAAATTGAACGTGCTTA
DreBA$_{GH}$51$_{bcd}$	550	50	TCTCATTAACTTACCCGCTTGAACATATTAATGTGCATATTCTC CCCTCC
DreBA$_{GH}$52$_{acd}$	368	139	GTCCGTGTGGGGATGCGAGAAACCATAGTGGAACGAGGATCA GACAGAAGATGCAAATGATCATCAAAACACACCGCAAAATAT GAAACTACTCATTTGCAGCCGAGTAAATCATAGAAAACTGATC GGGAACTGGCAC
DreBA$_{HI}$55	1,971	27	AGTGTTTGTGTGTGTCAGTCAATCAAA
DreBA$_{HI}$56	1,598	31	TAGTTTTTAAGTTAAATGGTGCACTTTTAAC
DreBA$_{HI}$60$_{ab}$	1,520	24	CAACATGAAACTGCCTATTTATGC
DreBA$_{HI}$62$_{b}$	1,219	32	ACAATAAAGTTTACAAGCGAGAATCCGTGACA
DreBA$_{HI}$63$_{b}$	189	189	ACATTAGACAAGGCTCTTGGAAGCTGTATACTCATTGGAGGAT ATATCATCACGTGCTGCCCATCACGTGGCCCCTGAGGAAAAAG GGGTGATTTTTGGTGTAAATCTAGACTGTAATTCTGTAATATAT CATTGGACCTCGTAAAACCGACACTAAAACTTCTTAGCATATA AATCACTTCTCAAATT
DreBA$_{IJ}$65$_{abcd}$	5,885	208	AGATGGCGACGGCAGAGCTACTGCTGTTAAAACAGTCATGAA GAAATGCAATAAATTCCTTGTTGTTTTATGAAAATTTACAACTT TGTGATAGAACTTTATGAGTGCCTGGGTCCTGGGATTGGCCGA GGATGGTCATGTGGACGAGTAACCGTGAACATGAACTTTTTAT GATTTCCCAAGTGGTTATATTGCAGCATTCTTTTGG
DreBA$_{IJ}$67	2,278	51	TTGAATTAACATTGTGTTAACACATTAAGGCAGGAGTTTGTTGA TTACCCC
DreBA$_{IJ}$68	1,387	22	TATTTTGAGTTACGCTCAAAAA
DreBA$_{IJ}$69	141	26	CCCGCACTGGCACATAGCCACCCAGA
DreBA$_{IJ}$70$_{ab}$	89	61	CCAATCTCCGATAAACTACTAATAGCTAAACCACTTGGACTAT AAAACACAACAAATCATA
DreBA$_{JK}$77	909	50	TCAACAAACCATATATCTTTTTGGAACATAAATCATGATCTCA CATTCCT
DreBA$_{JK}$79$_{bc}$	236	215	TCGTCATCATTTGTAACCATAGAGCATGAATTACCTCTTGAAGT CATCAGTGAGAATTTACGACTGGTCAACAAAAGCACGTGATTC TCAAACGCACCCCCACCCCCATATTTGGCCGCATACATAGCAA AAACGAAGTACAGTGCATTGCTATAATTCATTAATACATCATA AATCGTGAAGCACAGCGTTATAACGACCAAGATCTACAAATC
DreBA$_{KL}$83	12,508	64	AACATTACCTACAGAGTACTGTACATCGCGGCACCTTTCGGCA TGTTATGGATGTTTTAGGCAT
DreBA$_{KL}$85	12,316	131	GTTGTTCTTGTTACAAATATTGTTTATGTGTCAAGTGCTATGTTA ATGCTTTCTTGAGAGTTAGCATGTGAGCTCTTAAATGTTATAAC TTATTTACCTATACACGTGTATAACTACCTTTTGAGTTTACA
DreBA$_{KL}$86	11,959	217	GTTCATACGTGTGTATAGTTGTATTCCAAATATATAATTGTATA TTACTGGATCATGTTTATGTTCCAAAAAAGGTTGTATATATTGA ACATTTTTATTGTGATCAGTTGGCTATTTGTAGTGGGCAGAATA AACGGCAATGTGGAAAAAAGTCATGTCAATCTGCCTGTTAGCA TTTTGCCAAAATGCTTTTTGTATTTGTTTGTAGTTTAACTTG
DreBA$_{KL}$87$_{abcdef}$	11,215	201	AAAAGTAGCTATGACATTTACATGTCAAACGGATGAGCGTTTT ATCTTGAAGTTAGATCGTAAAAATCGCCCAGGCCACAGACAG ATACCCCTTACTGGCTCTCAAAAGTCACGTGGGGTCCATAAAG TTAGTTTTATGGTTTTGGGGAGTTGACAATGTACTATATATTTCA CATTCTAGAATGCAAGTGACGGTTTAAC
DreBA$_{KL}$90$_{b}$	8,533	39	TTGCATATTGATAGGAGCAATCTGCGAGCATATCTCATC
DreBA$_{KL}$91	7,654	19	CAGAGCCCGCTTTAGACCA
DreBA$_{KL}$94	5,716	42	AAGTTTAATGCTGCCAGTCTCTTTTTTCAAAATTTAATTATT
DreBA$_{KL}$98	8,534	83	TTTGCATATTGATAGGAGCAATCTGCGAGCATATCTCATCCAT AATGCATAGGGACAGGATCATCCAAGGGTCAGCAGACGGG
DreBA$_{KL}$102$_{ab}$	4,301	197	AAGTAATGACCTGGGTACAATTCAATATGAGCCAGCAAACTAT GCATGCATTACTATAGAACTTTAGAGTGGGAGAGACCCAGAGT AGGGGTGAACCGGAGGTCAGTGCGTCTAACAAATATTAAAAT GTACTGGGACTCGTGACACGCCTCGCTGTTTAACAAAGACTGC CAAAGTGCGAGATTAATATGAAAACT
DreBA$_{KL}$103	1,024	19	AACATTTGGTTTGTTATTG
DreBA$_{KL}$104$_{abcde}$	1,150	77	TCTATATATACCCTGTAGATCCGGATTTGTGTAAACAGACGCA CAGTCACAAATTCGTATCTAGGGGAGTATGTAGT
DreBA$_{KL}$105$_{abc}$	635	77	GTGATTTAGGAGGCTTTGTAACCTGAATTAGTTGATGAATTTTC TATCGATCTTAAACAAGCCTAGATTCATCTCTG
DreBA$_{KL}$106	457	13	TTTTCATGAACAC
DreBA$_{KL}$107	1,440	36	TTTCTGATTGTAGGAAAATTATGTTTCAGAGCCACA
DreBA$_{KL}$108	60	59	AAACGAGTAAAGGGATACATATAAATTTTAGTATATTTTGTGT GCAATTCAAAGAAATT
DreBA$_{LM}$110$_{abc}$	22,379	71	CAGATTTGGTGAAGATGGATCCGCGTTTCATCTTTAATCACGCC AAACTCGGCCCCATTTGTCATGTTTAC
DreBA$_{LM}$113$_{bc}$	21,228	557	GTCTTGAAGATCACTGCCTTTCTCTCTGGTGATTTATATCGCTG CATTGCGAGTAATCCTGGTACATTACGCCTTGTAATCTTATTCT AAATCACAAAACCCTTTCTTTCTGTCTCGCCTCTTTTTAATTTAA CGGTGCACAAACCCTCTGTTCAACGACGCCGTGAATGCGCCTA AAATATGGATCCATAGTAGAAAAAAGCTGAAGTTAAATAGTTT CTCGGCTCGTGGGCTTTTAATGCTTTAATGAGTGTA
DreBA$_{LM}$114$_{ab}$	20,131	75	AAAGTTCACCGCCATTATTTGCAGACAGATGCCTTAAGAAAAA TGTGAGAATTATACAGAAAAATCATTAATCAC
DreBA$_{LM}$118$_{abcd}$	15,298	174	GGTCATAGTAATCCCGCGAGCGTGCAAGCCGTGGTGCCCCCCA TCCCCCTCCTGGTCACGTGATTCATTAAATAATTAATGCAGTGG TTGCAGAATAGATATGTATTCGTCAGGAATATCGCAGACATGG TCTCCTAGTGTCTGACGTGAAATTCAACTGTCCTTTTAAAAACA
DreBA$_{LM}$119	11,555	26	TTATTTTTCTAAACTAAGGAAAATAT
DreBA$_{LM}$120	6,750	22	TGTAATAACTGTTTTGTAAATG
DreBA$_{LM}$127	192	147	CTATTCATGTCATGGCGTCTTTCAGTGCAGTAAGATGGATTTAC CTCGTGCCCCTCTTATTATGTGCCTCTTATAACCTTTCAGGTCA GCTTCCAAGAGGCCATTGGAAAGAGACCGTCACGTGACAACT GGTGCCAATGTTCTTCC
DreBA$_{LM}$129	9,435	27	TAATTTAATTGATTTTATCATAAATCA
DreBA$_{LM}$132	4,465	36	TGCAGACACCTACATTTTTGCCTTGTGCCTTCCTCC
DreBA$_{LM}$134$_{abcd}$	2,969	133	CTCGTCAGAGGCTAAGGTAAGCTGGTCTGAAATAGGCTATCAG TTTGTGAATGGCGGTATGTGTGTCCTAGTGATTTATGACCGTAT GACTCCAAACGCGGTTCAAGAAGAGTTCACAATGCTTTAAGCT TCC
DreBA$_{MN}$139$_{bcde}$	3,459	41	TTAACGTGGATACAGGGTATATTTGAACAAAATGCATGTCC
DreBA$_{MN}$140$_{abc}$	3,366	229	TCTGGGCATCATTGTTGCACTTAGAGTTTACATTAAATGGGTGA GGAAGAAAAGTAAATCTTATTTTGAATCGGAAGACCTTCAATC AGCGTCTTTCGCTAAGTGTGTTCAAGTGAACATTCATAAATATA TATTTATTTGTTATAGCCAGTTTAAATACTTTCTTTTTTGTATTAT TTATCCCCATGTATTTATATCGATAAAAATGTACTTTTTTAGTAT CTACCTG
DreBA$_{MN}$141	2,895	17	AACAATTCTTTAATAAA
DreBA$_{MN}$144	1,213	41	GGCTGCTGTCAAGCGCTTTGCCGGCGAAAAGATTGATCACC
DreBA$_{MN}$145	791	17	ATAAAACATATAAACAA
DreBA$_{MN}$146	242	194	CCAGGCCCCGAATTTCAATTGGCTGAGATGAGTCACGTGACCA GGAATTGGCTGCAATTTCGCCCATAGTCTTCAGTTTAGTAGACC CTGGTCCCCATACGCTAGTAATATCACCAATATACACAATTAT TAAAGCCCGCAAATGCCGCCATAGCAGGAGCGCAAAATAAAT CATAGCTTTTTTGGATGTTATTT
DreBA$_{NO}$147	14,210	94	TTTAGACCTAGTTTATTTATTGAGATCCTTTAATAGTAAATTTCA AATTTATTTATTGTACGTTATTTTCATAGTTGAAATAATAAAAA TAAAA
DreBA$_{NO}$148	13,453	72	CTGCTCTCGCCCTGATGCAAAATGGCGCTCGAAACACAATGAA GGATATGAACAATAAAGCGGCAGCCATTA
DreBA$_{NO}$149	13,066	20	ATTTATATTATAAACATTAA
DreBA$_{NO}$154$_{bc}$	14,161	34	TATTTATTGTACGTTATTTTCATAGTTGAAATAA
DreBA$_{NO}$155	6,699	33	AGGTTTGTTAATTGTTTGGCAATAAAAATTATG
DreBA$_{NO}$157	5,114	20	TAAATAAATAAATAAATAAA
DreBA$_{NO}$158	2,044	106	AGCCATATTCGACAGTTTTAAAAAAATAAACTACTGATAATAA ACATAAGACTGAGTTTCAGTATTTTTATTGTATTTTTGTCAAGA GACGTTTGATTTTTTTATT
DreBA$_{NO}$161$_{abc}$	281	63	TCAGATTGATGGGCTGGTTTGATTGAAGTGGCTTTGTCATGCAA ATGTCAAGCGCGTGATGGA

Oryzias latipes medaka Hox Ba

PFC	Dist to 3' gene	Length	Sequence
$OlaBA_{AC}3$	4,855	21	AATTAAATATTTTTCAGTGCT
$OlaBA_{CG}5$	25,610	37	AACCAGTAACTGAACTATATATAAGTCATAATAAATT
$OlaBA_{CG}6$	19,300	16	AGAATAGTATTTAACA
$OlaBA_{CG}7$	9,969	199	AAAACATTTTCAGTTTGATGCAAAGAATTTTTGCAAATAAAATT TTTCTATACTTTTTTTGCCTCTAACCCAAAGTATGTAATAATTAA AAAAAAAATAAAAAATTTTGCTTCACTGGATTTAGCCATTTTTG TGAGCATCAGCTGTGCAGCACAGCCTCGTTGTCTTTGTTGGAGC TTATTGTTGTGCTCCTTGCCTG
$OlaBA_{CG}12$	8,237	23	GTTAGTTAGACCTCTGCATTCAG
$OlaBA_{CG}13$	1,565	82	AATTAACGCGCTCTCTGGTTAGCAGCGAACCCAGCAGGGCGGA GGCCCCGAAAAATTAAAATTCCCCGAGCTTTGCGGTCTG
$OlaBA_{CG}14$	1,260	26	AAAGCGGGTTTTTTTTTCTTTTTTTC
$OlaBA_{CG}15$	101	101	AGCAAAGCAGAAGAAAAGAAGAAAAGAGTGACGTAAATCTCA TAATGTCCCTCTTCTTCTGCTTTAACAAACTCGCGGCTGCACGC GGATAATGTCCGAGA
$OlaBA_{CG}16$	4,219	23	GTTTCAACATTTTTGTTTTTGTT
$OlaBA_{CG}18$	10,345	27	TTTGTTGAGTGAGCACAGGGGATCAGT
$OlaBA_{CG}21$	3,582	15	TGATCTCTTCATATT
$OlaBA_{CG}30$	19,382	47	CTGTTTTCCCCTTTGTTGATGTTCTCGCTTATTATGCATTGGATAA A
$OlaBA_{GH}39$	9,226	63	AGAAAGAAACAGCATTTCCTTTTCTTTTCTTATTGAAGCACTTC CTTTATTGGTTATTTATTG
$OlaBA_{GH}40$	5,500	29	TGTAATAGCCTATAAATATAGAGAGAAAA
$OlaBA_{GH}42$	7,066	26	ATGAAAGAGGATTCCAAACATGCAGA
$OlaBA_{GH}43$	8,371	15	TCTTCCGTTTCTCAC
$OlaBA_{GH}45$	3,377	83	GCAATAAAAGAATATGACGGCAATAAAAGTTTATAGCGTATAA ATTTCTGAAGGTTAAGAACTAAACGGCTGTAAAGCAAACA
$OlaBA_{GH}48$	1,753	318	AACATGCAGTCGGATTATTTTGATTTTTCAAATTCAAGATACTTC CAATTCTCCAAAATACCAGATTTTCTTTGTTTTGCGTGCGTAAA AGTGTTTGTGGTCACGGGGGTCCAAGTTTTCAGTGCCGCAGTTAG ACTTAAACAATTGCCCTGCAGTAGACCAAATTCTCCAGGCCAT AATCTGGAACGGAACGCAAGAAACACTTATGTTTGGGTAGGAT ATGGTCCCATCCATGTTTATTGGATGAGAAACAAAA
$OlaBA_{GH}53$	235	212	CGCTGTGGTTTGCTCTGGCTGCTCACCGTCGCTCTTAGCAGCCG GCAGCCGCGCGCCTGTTGTGGCTGTGAGGGGAAACACCACAAC GACGGTCACCGACACCGGCGGTAGACCTGCATTGCCTGCCATC CCCCCACCCGCCTTGGTCTCTGTGTGTTATTAGGTCTGCCACAA AAAAACCAACAACACACACCTGAAAGCCACTCTGACTC
$OlaBA_{HJ}54$	9,435	31	AAAAGACAAAAAACATGACAAACGTCATCAG
$OlaBA_{HJ}57$	24,265	37	ATTTCTAAGACGTGCGTCAGACGACTTTGTAAAACTG
$OlaBA_{HJ}60_{ab}$	28,111	24	CAACATGAAACTGCCTATTTATGC
$OlaBA_{HJ}65_{bc}$	6,481	146	ATGAAGAAATGTAATAAATTCCTTGTTGTTTTATGAAAATTTAC AACTTTGTGATACAAGTTTATGAGTGGCCGCGCGTGGGGATTGG ACGAGGGGTTGGTCATGTGGACGCGCTTAACGTGAACATGAAC TTTTTATGATTTCCC
$OlaBA_{HJ}70_{ab}$	256	66	CCAATCTGCGATAAACTAAGTGATAGCAGCAGTCGGACTGTCG ACTATAAAGCACAACAAATCATA
$OlaBA_{JK}75$	1,118	45	CTGAAATGCATCCCGAGCCGGGAAATCAACAAATCATAAATCA GC
$OlaBA_{JK}76$	676	71	TTTATGGCTTCTGGTCTCGGGCTTCTAATTTCCAGCCTCAGCAAA GGTCAATTAACAGGCCTGAGTGATGA
$OlaBA_{KL}87_{bcdef}$	10,515	193	TGACATTTACATGTCAAACAGATGAGGGGTTTTATCTCCGCGTC GGATCGTAAAGACCAGGCCGAGCCTCAGACTGACACATCTCAC TGGCTCTCCTCTGGTCACGTGTGGTCCATAAAGTTAGTTTTATGG TTTTGGGGAGTTGACATTGTACTATATATTTCACATTCTAGAAA GCAAGTGACGGTTTAAC
$OlaBA_{KL}96$	8,524	14	CATTTTGTTTTTCA
$OlaBA_{KL}91$	6,950	19	CAGAGCCCGCTTCAGACCA
$OlaBA_{KL}102_{b}$	4,567	118	GGTGAACTGGAGGTCAGCGCGTCTCACCGACATTAAAATGTGC ACAGGGGAACGCGGGGTGACACGCCTCGCCTGTTAAACAAAG AGTGTGCCAAAGTCGCAGATTAATTTGAAAACT
$OlaBA_{KL}103$	1,050	19	AACATTTGGGCATTTATTG
$OlaBA_{KL}104_{abcd}$	1,310	71	TCTATATCTACCCTGTAGATCCGGATTTGTGTGACGGTCGATGA AACAATCACAAATTCGCTTCTAGGGGA
$OlaBA_{KL}105_{ac}$	682	76	GTGATTTAGGAGGCCGTGCGCCCAGAATCACTTGATGAATTTTC TATCGATCCTAAACAAGCCAGATTCATCTCTG
$OlaBA_{LM}109$	26,025	190	ACACAAAGAAGTGGGTTTTTTGCGGCTTTTGCACACAGAATGCG CTCTGGCGCACAGGCAGGAGGGCGCGATGATGCGTCCGTCACG GCCTTGAGTGATCGCGGTACATTACGCGCAGTAATCTTGTTTTA AATCACAAGCCTTCCCCCCGTTTCACTCCCGTTCTCCCGTTAAT TTAACGGCAGCTCAG
$OlaBA_{LM}113_{bc}$	256,339	104	ATAGCTCAAACGTTGACAACCAAATAGATAATCAACAAGACG AATGGCTTCTTTTGTGCAGCGCGGCTCCCGTATTAATTCTCATTT TCTTTTCTGAGAGCAGG
$OlaBA_{LM}118_{b}$	18,643	26	CTGACGTGGAAAATTCAACTGTCCTT
$OlaBA_{LM}119$	11,949	29	TTATTTAAATATAAACAAAACGTAAATAT
$OlaBA_{LM}120$	7,000	24	TGTAATGTTCATTAAATTTAAATG
$OlaBA_{LM}121$	5,846	227	GATGTGTAAGATGAGAACCAGAATCAATAAGAGCTCATATCGC TCCCACTCAGCCTCCATCCTCTCTGCTGCTTTTTGCTTTTATTCC GGGAATAATTCCAGGTCAGCAGTACCTGTCAGCGGCTGCTCCT GTCCCACAACCTTCGATTAGACTACAAGAGGTTGGTGCTCCCTA TCTGATTTTTCAGCCCCTGCATCTTTCTTCCCTTTCTTTTCCGCAT TTTATC
$OlaBA_{LM}122$	5,232	14	ACTCGTGTAATGTA
$OlaBA_{LM}123$	3,199	207	GGAGAACCGCGGGGACGATCAGGAGGAGGACGGCTGATCACA GAAAATGCGATGGGGTAACGGTTCCGTCTGGATCTTTGTCTGAG ATTCTCGGCCCGGATGGCTGCCGGCAGAGCCGCTCAAGCCAGA GAAATTCGAGGCCATTGTTATCAGCCAGCAGCGGCTGCAGCCA GAGCTGGTGTCGGGGCGGAGGGTGGAGGTGGAGGG
$OlaBA_{LM}124$	2,850	107	CCGTTCATGCCGCAGGGAGTCCGGCGCGCGCCGCCCGGTCTGC TGCTGCTGCGGCTCCAGGAGGGGTGGTGCTGCTGCTGCTGCTGG TGGTCGCCAGCAGCCGTGCT
$OlaBA_{LM}125$	2,564	24	CCACCGGTGCTGAGGAAGGTAAGG
$OlaBA_{LM}126$	2,068	442	TTTAATGGCTTTTCTCTGTCTCCTCCACCTGCAAAGCTCCTCATG GCAGGCGCCATTTGTAATTTTATATCCCGGGATAAATCGCAGAG TGGACACCAAGGCCGAGCTCGTCCCCCCCTCCCTGCCTCCCTTC CTCTCTGTGTTGTTGTTTGGGCGGCCGTGTCTCTGTCACCGGCCT GTGTGAGTCGATTATTCAGGGAGGCGTCTAGTAAAGTTGGTGAT TGTATTAAGCTGTAGTCCTACATGCAAGGGGCA
$OlaBA_{LM}127$	203	146	CTATTCATGTCACGGAGCCTTTCAGGGGAGGAAGATGGATTTAA CTCTGTCATTTCTTTATTCTGTCTCTCTAACCTTTAAGGTCAGATT GCAGGGGACCGTTGGAAAGACACCGTCACGTGACCTCCGGTAC CAAATGGTCTTCC
$OlaBA_{LM}128$	12,880	22	TCCTCCTGCCTACCTGCAGCCT
$OlaBA_{LM}135$	204	41	CCTATTCATGTCACGGAGCCTTTCAGGGGAGGAAGATGGAT
$OlaBA_{MN}136$	9,069	20	AATCAAGAGCTATCTTTTCT
$OlaBA_{MN}137$	8,843	270	AGCTCTCCTGCCTGGTAGATGAAATAGTTGAGATTTTAATGGGC AGTTAAGGAGGTGGTTGTATTTATTATGCCGGGTTTTCTCCCCCT GCAGGCAACACATGTTGTTGTCAGCAGAGCTGAGCTGTGTGCTG GGCTCAGGCCCTCCATCAGCTGGAGAGCGACAGAGAAATGCA ACGCTGCTCCACAAAAGCACAGAGAAAAACCACTGGACGGCC TCCCACTCGCTGTCTACTCTTTTTTGCTTCCTGCAGTA
$OlaBA_{MN}138$	8,486	199	AACAGGGTTTTATAGGCTTAACTCTCAGTCATGCCTACATGCAA CATTAAGCTCACTTCCTTTCACAAGCAGAGTTGCCAGCTGTTGA CGTGTTCAAACAAACCTGTGAGAGGACATGCAAGGCGCATTTT AGTGTCCTTACCTTTGGCACATCATCGACACAGGGTAAATACC AAGAGTTGAGAGAACTATTTATTTA
$OlaBA_{MN}139_{acde}$	8,186	54	TCTTGACTTTTAATGTGGAAACAGGGTATATATATTTGAACAAA ATGCATGTCC
$OlaBA_{MN}140_{abc}$	7,951	260	TCTGGGCATCATTGTTGCACTTAGAGTTTACATTTTGATGGTTAA AGGTTAAAAAAATAATAATGAAATCTTATTTTGAAGACGGAAG GCCTTCCAATCAAAGCGTCTTTCACCAATGCGTGTTCAAGGGAA CATTCATATATAAATATTTATTTGTTATAGCCAGTCTAAAAGGA CTTTCTGTTTTATATTAATATTATTTATCCTTCATGTATTTATATT GAGGAAAAAAAATACTGTACTTTTTTAGTATTT
$OlaBA_{MN}141$	7,391	17	AACAATTCTTTAATAAA
$OlaBA_{MN}142$	5,129	104	GCTGCTGTCAGCTCTCTCCAAGAGAAAGATTGATCACCGCATTT CTTTTCTCCCACTCGGGGCTCCTGTGGCCCTCCTCAAAACTTCC CTGTTAAACTCCCAAA
$OlaBA_{MN}143$	4,504	619	CATATTGTAAACGGTACTGGCCAGCTCCTCTCGCCTGATGTTAC CCGCCATTACATCAGCCCAGAGGAGTGGAGCAAATGCCCATTG TTTGCGCCTGTCATCGGCTGGGATGCGCCGGACTTGCCGCCCACA CGGCGCACATGGCGACCGGTCGCCTTCTCCCTTTCTTGCCCCTC AGAAATGGCCCCCCGTTAGGGCCAAACAATGGCACAGGGATTC TCTTTGCCTCCATATTTGGTGGAAAAAAATAATAATC
$OlaBA_{NO}147$	8,175	95	TTTAGACCTTAGTTTATTTATTGAGATTCTTTAATAGTGAAAATC CAAATTATTTATTGTACATATATTTTCATAGTGGAATAATAATA ATAAAA
$OlaBA_{NO}154_{ab}$	8,161	62	TATTTATTGAGATTCTTTAATAGTGAAAATCCAAATTATTTATTG TACATATATTTTCATAG
$OlaBA_{NO}148$	7,478	72	CTGCTCCCCTCCAGAGTCAAAATGGCGCTGGAGACAAAGGAGG ATGTGAACAATGAGCAGTGGCAGCCATTA
$OlaBA_{NO}149$	7,096	17	ATTTATCTCTGCATTAA
$OlaBA_{NO}155$	961	30	AGGTTTCTCTGTTAAACTGGCAGAATTATG
$OlaBA_{NO}156$	740	288	GCTCCTCTTTTATTCCACACCGCTCTGCGCTCTGACACATCCGC GCTGATATCAGATTGATGGGCCGGTTTGATTGAAGTCTCTTTGTC GCGCTGATGTCACTGCGAGTGATGGATGAGGCGCGTGCGCCAC TATTGACGCGCGCGCGCACCAACTTCACTCCTGAGCCGAAAAC GACCCCATCCCCTCCTCCTGTACGCACACACACACACAAACAC TGCCGTGACCAGCTGACAACTTCTGGCCCGGTCACCA
$OlaBA_{NO}161_{abc}$	687	64	TCAGATTGATGGGCCGGTTTGATTGAAGTCTCTTTGTCGCGCTG ATGTCACTGCGAGTGATGGA
$OlaBA_{NO}160$	812	21	CCTCCTCACGCACGTTCCTTT

Tetraodon nigroviridis spotted green pufferfish Hox Ba

PFC	Dist to 3' gene	Length	Sequence
TniBA$_{AC}$1	3,439	36	AAACAATATGGTGTTAATGCTCAAATGGTATTCCAG
TniBA$_{AC}$2	13,381	26	AAACAAAGGCAGCTTTTTTGAGAAAT
TniBA$_{AC}$4	908	23	TTTAATAATACATAATAATACAT
TniBA$_{CG}$5	30,138	38	AACCAGTAATCAACATGTAGCGCAACAACTAATAAATT
TniBA$_{CG}$7	11,747	19	TTGTTGTCCTGCCTGCCTG
TniBA$_{CG}$8	15,031	19	ATTAATACCTGTGTATCAC
TniBA$_{CG}$9	12,637	18	TCACAGTTCCCGTGAGAC
TniBA$_{CG}$10	6,064	25	CTTACAGATTGGGAGACATATTTAA
TniBA$_{CG}$11	5,598	12	CATGAATTCACA
TniBA$_{CG}$13	2,578	81	AATTAAGGCGCTGTTTGGTTGGCTGCTCGGCTGCAGGGTGGAG GCCCTCCATAGATTAAAAATTCCCACAGCCTGCGGTCTG
TniBA$_{CG}$14	2,238	26	AAAGCGATCTTGTTTTTGTCTTTTTC
TniBA$_{CG}$15	686	118	AGCAAAGGGCAGAAGAGAGAAGGAGAAATAGACAGAGGGA AAGAGTAAAATCTCGTAATGTCTCTCTGCTTGTTCTCTGTCTAG TAAAATCGCGGTGCGTCGCAGATAATGTCCGAGA
TniBA$_{CG}$27	4,461	37	TAGATAAAAACTAATATTTTACAATAATTAACATTAT
TniBA$_{CG}$28	22,698	22	CATAAAGATTTCAACTTGGGCC
TniBA$_{GH}$39	9,294	66	AGAAAGAAAACCGCAGCTGTGCGCTGGACAGAAATGCAGAA ACAGATGCTTTTTGGTTATTTATTG
TniBA$_{GH}$44	10,338	20	TCATTATTGGAAGAAAAATG
TniBA$_{GH}$45	4,408	83	GCAATAAAAGAATATGACGGCAATAAAAGTTTATAGCGTATA AATTTCTGAAGGTTAAGAACTAAACGGCTGTAAAGCAAACA
TniBA$_{GH}$47	9,023	13	CAGTCGTCCTTCA
TniBA$_{GH}$53	260	214	CGCTGTCGGTGTGCTGGTGTTTGTGGTCGTTACTCTTGCCGGCA GCCGTCACCTTTACGCGCCCCGCTCCTCTGTTGTGGCCGCGGG GGGAAACACCACAACACCAGCGGTAGACCTGCTCCACTTTCT CTTTCTCCCTCCCCTCCGTCTCCCTCTTTTCGTGCCCGAGTCTCT CCGCTTCCAACAGCCTCTTCCTTCACTTCTACTCTGACTC
TniBA$_{HJ}$54	9,170	28	AAAAGAAATAGTGCGTTACCGCCATCAG
TniBA$_{HJ}$55	17,063	28	AGTGTTTACGATCTCTTGCTGTATCAAA
TniBA$_{HJ}$56	16,665	31	TAGTTTCTTCTTTTTGTTGGAAGCATTTTTAA
TniBA$_{HJ}$58	6,599	27	TTGGTGTCTCAAATATCAAACAACTGA
TniBA$_{HJ}$67	2,551	47	TTGAATTCAAATATGTCACTGTATCCCGAGGAAGACGTAATTA CCCC
TniBA$_{HJ}$68	1,652	27	TATTTTGAAGGCCTATTTTCTCAAAAA
TniBA$_{HJ}$69	373	26	CCCGCACTGGCACTCAGCCACCCAGA
TniBA$_{HJ}$70$_{ab}$	257	76	CCAATCTGCGATCAACTAAGTGATAGCGGCAGCAGCAGCCCG GCGGACTGTCGGCTATAAAACACAACAAATCATA
TniBA$_{JK}$75	1,161	44	CTGAAATGCATCTCATCCAGGGAATCAACAAATCATAAATCA GC
TniBA$_{JK}$76	694	69	TTTATGGTTTATGGTAGCATTTTCTAATTTCTGCCAAAGTAAAG GTTAATTAACAGCTCTGAGTGATGA
TniBA$_{JK}$77	1,137	54	TCAACAAATCATAAATCAGCAGATTTCCGTGCACGGATCAGG TTTAACATTCCT
TniBA$_{KL}$87$_{bcdef}$	14,162	193	TGACATTTACATGTCAAACGGAAGAGGGTTTTTATCTCCAAGT TGGCTCGTAAAGATCAGGCAAGGCCTCAAAATGATACCTCTC ACTGGCTCTCCGCTTGTCACGTGGGGTCCATAAAGTTAGTTTTA TGGTTTTGGGGAGTTGACAATGTACTATATATTTCACATTCTAG AAAGCAAGTGACGGTTTAAC
TniBA$_{KL}$91	8,415	19	CAGAGCCCGCTTCAGACCA
TniBA$_{KL}$95	4,712	20	ATTTTTTAGCAGTCTGTTCA
TniBA$_{KL}$102$_{b}$	6,232	118	GGTGAAAGAGAGGTCAGCGCGTCTCACCAATATTAAAATGTG TACCGGGGACGCGGGGGGTGACACGCCTCGCCCGTTAAACAAA GAGTATGCCAAACTGGCAGATTAATTTGAAAACT
TniBA$_{KL}$104$_{abcd}$	2,012	71	TCTATATCTACCCTGTAGATCCGGATTTGTGTAACGATCATTAA AGCAATCACAAATTCGCTTCTAGGGGA
TniBA$_{KL}$105$_{bc}$	840	52	GAATTAGTTGATGAATTTTCTATCGATCCTAAACAAGGCAGAT TTATCTCTG
TniBA$_{KL}$110$_{bc}$	29,183	44	CATCTTTAATCACGCCAAACTCGGCTCCCATTCGTCATGTTTAC
TniBA$_{LM}$109	27,710	191	ACACAAATATTTTCATTAAATTTCAATGTTGGTTTTGTCTATAG CCTGTCTTGAGAAATCAGCAGCCTGCGATGATTAATCTGTGCC GGCCTTGAGTGATCTTGGTACATTACGTCCAGTAATCTCGTTTT AAATCACAGACCGTCCCTTCGTTTCACTCTCGTTCTCCCGTTAA TTTAACGGCTGCTCAG
TniBA$_{LM}$113$_{bc}$	27,288	104	ATAGCTCAAACCCTGACAACCAAATAGATAATCAACAAGACG AATGGCTTCTTTTGTAAGGAGCGGCTCCTGTATTAATTTTCATT TTCTTTTCTTAGAGCAGG
TniBA$_{LM}$118$_{bc}$	18,479	120	TCACGTGATTCATTAAATAATTAATGTGGAGGTGGCAGAATGG ATGTGTATCGTCAGAAATATAACGGTAAGACATGCGGCTGCTT CAGCGGGTCTGACGTCGAAAATCCAACTGTCCTT
TniBA$_{LM}$121	6,383	235	GAATCAATAACATCTGAGATCACTTTAACCCCGGCCTTCATCC TTCTTCCTTTCCTTATTAGCCTACTTTTACTCCGAGATTTGGCCT GGAATATAACCATCAGCATTACCTGTCAGTGGTCGCCGCTGTC TCGTGCGTTCGATCGGACTACCAGAGGTCTGCGCTCTTTATCG GATTTTTACCGAGAGCTTTCTGATTGTATATCTTCTTTTCCCTCTT TCTGCCCCTCTTTATC
TniBA$_{LM}$122	5,673	15	ACTCGTGCTAATGTA
TniBA$_{LM}$123	2,833	204	GGAGAATGGGACCGTGAGCGCGGACGATTCACCGCCAAATAA ATAAGATGGGGTAATATTTTGTTGGAGTTTTTTTTTTTTTTTTCC TCCACAAACGCACGCCGGCTCTGTCGTGTGGCCGCGGTGGAC ATTTTAAGGCCATTGTTGTCGGTAAGCAGCGGTGCAGGCAGCG CTGGTGCTGGGGCGGAGGGTGGAGGTGGAGGG
TniBA$_{LM}$124	3,117	96	CCGTTCATGCCCCACGGTGTCCGGCCCGCCGCTGCGCCGCTGC CGCCGCCGCTGGCTCCCGGAGGTGTCGTTGTGGTGGCGGTGCG TCGCCGTGCT
TniBA$_{LM}$125	2,837	24	CCACCGGTGCTAGAGAAGGTAAGG
TniBA$_{LM}$126	2,236	435	TTTAATGCATTCTGGGTGCTTCCCTGCCCGCATTGCTCCTTATG GCAGGCGCCATTTGTAATTTTATATCCCAGGATAAAACAGACA CTGGACACCAAGGCCGACCCCTCTCTCTCTTTCTGTGTCTTT CTCTCTCTCCCTCCTTCCCTGTTGTTTTTGTTTGGGTGGCTGTTT CTCGGTTGGTGTAAGTAGATTATTCCGGGAGGCGTCACAGTGA AGTTGGTGATTGTATTAAGATGTGGCCCTACAAGCA
TniBA$_{LM}$127	204	147	CTATTCATGTCATGGTGCCTTTCAGTGCTGGAAGATGGATTTAT CTCTCTCATTTCTTTATTCTGCGTCTCCTAACCTTTCAGGTCAG ATTGCAGGGGACCGTTGGAGAGATACAGTCACGTGACACCCG GTACCAAATGGTCTTCC
TniBA$_{LM}$134$_{bc}$	2,766	78	CAGTTTGTTAGCGGCGGAGTGGCTGGCTCAGTGATTTATGGCC CGTATGACTCCAATCCCGGTTCAAGAAGAGTTCAC
TniBA$_{MN}$136	5,567	20	AATCAATGATGGGCTTTTCT
TniBA$_{MN}$137	5,354	299	AGCTCTCCTGCCTGGTAAATGAAATAGTTGAGCTTTTAATGGG CAGTTAACGAGGTGCTTAGATTTATGACACCGAGCTCGCTCCC CCTCCAGGCAACACATGTTGTTGTCAGCCTAGCTAAGCTATGT GGCGGCTCTTGCTCTCCATCAGCCAAAGAGAGCAAGGGAGAG CAGAAATGGATGTGTGGAGCTGCGGCTAACACTTGTCTTTCTC TCTTCAAAAAACCACTGGGCGCTCTTTCGCTGCGCCGTGTC
TniBA$_{MN}$138	4,889	268	AACAGGGTTTAAGAAGCTACGAGCTAAGCTAGCTCCAGTAGA ACATGCATAGGCAGTGTGAGCTAGCCACTGCTTTATACACACA TTCAATGCTAAGCTAGCCTTCCTTTTGGTTGACACACAACTAT AGCCAGGTTGCTCTGGCTGTTGACTTGTTGACTTGTTCAAATAA ACACATAAAAGGACATGGAAAAGACATTTTTATGTTGTTATTT TAGCACATCATCGACACAGGGTAATAAGAAGAGCAGAGAA
TniBA$_{MN}$139$_{acde}$	4,482	53	TCTTGACTTTTAATATGGATACAGGGTATATATTTGAACAAAA ATGCATGTCC
TniBA$_{MN}$140$_{abc}$	3,765	255	TCTGGGCATCATTGTTGCACTTAGAGTTTACATTTTGATGGTTA AAGTTAAGAAAATAATAATAATGAAATCTTATTTTGAAGACG GGAGACCCTCCAATCAAAGCGTCTTTCGCCAATGTGTGTTCAC GTGAACATTCATATATAAATATTTATTTGTTATAGCCAGTTTAA AAAGACTTTCTGTTTTGTATTATTATTTATCCTCCATGTATTTAT ATATAGAAAAAAAATGTACTTTTTTAGCATTTACCTG
TniBA$_{MN}$141	3,289	17	AACAATTCTTTAATAAA
TniBA$_{MN}$142	1,393	128	GCTGCTGTCAGGCTCTTTGGAGGGAGAGAAAGATTGATCACCG CACTTCTTTTTTTTAACCCGCTCTCCTCGGGCCCTCCCCAAAAA TAGGCTATCCTCTGCTAAACTACTTTAGATCTTTTCCCAAA
TniBA$_{MN}$143	661	639	CATATTGTAAACGGTGCTGGCTAGCTTCGTTCACCTAATGTTA ATTGTCATTACATCACCCAGTTCCCCCAAGGAGGAAAAAAAA AAGATCCAGAGATGCTCTTTTGTTTGTGCGTAACACTGGACCA GGATGCGCCAGCCCGCCGCCCACACAGCGCACATGGCGACCG ATCGCCTTCCCGTTTTTTTTTCTTCCCAGAATTTACCCCCCGTTT AGACAAAACAAAAGTAAGTACCCCTGTGATGGAGCCTCTT
TniBA$_{MN}$144	1,394	43	GGCTGCTGTCAGGCTCTTTGGAGGGAGAGAAAGATTGATCACC
TniBA$_{MN}$145	929	16	ATAAAAAGCAAAACAA
TniBA$_{MN}$146	341	196	CCAGGCCCGGACCCCGCGATTGGTGAGTTTGAATCACGTGAC CAGGAATTGGCTGCAATTTCGCCCCATAGTTCTTCATTTTAGCC TACCCAGGTCCCCATACGCTATATCACCAATATACACAATTAT TATATAGTCACCGCATTTACGCCATTGCGGTCGGGAGACAGGC GCTCGTGCGTGTATTTATTTATTT
TniBA$_{NO}$154$_{ab}$	10,101	63	TATTTATTCAGATTCTTTAATAGTGCAAAATCCAAATTATTTAT TGTACATATATTTTCATAG
TniBA$_{NO}$156	838	293	GCTCCTCTTTTCTTCTCTGCTTCTCTTTGCTCTGACAGATCCGCG CTGATATCAGATTGATGGCCCCGTTTGATTGAAGTCTCTTTGTC GTGCTAATGTCACGGCGATTGATGGATGAGGCGCGTGCGTCCG GGCAGCCGCGCGCGCACACACCCAACTTTAAAACCGAATAGG TCCCCTCCCTTAATGCACACACACACACTACTCCCCCCGTCCC TGTGACCGGCTGACAAACTTGTGGCCTCCGGTCACCAG
TniBA$_{NO}$157	4,128	20	TAAATAAACAAACAAATAAA
TniBA$_{NO}$158	1,185	33	TTGTATTTTTTAAATGGATATTAACTTTTTATT
TniBA$_{NO}$161$_{ab}$	787	38	TCAGATTGATGGCCCCGTTTGATTGAAGTCTCTTTGTC

Danio rerio zebrafish Hox Bb

PFC	Dist to 3' gene	Length	Sequence
$DreBB_{HJ}59$	5,541	38	AAATGTTTAATTTGAGTTTAATAATTTTTTTCTTAATT
$DreBB_{HJ}61$	5,332	20	AAATCATATAGGATATTTTT
$DreBB_{HJ}64$	7,249	27	TTAAATAAAAGTCTGAAGTTCAAAAAC
$DreBB_{HJ}65_{cd}$	3,061	142	TTTATGAAAATTTACAACTTTGTGATAGAACTTTATGTGTGGTTCTGCC AGTGGATTGGCTGACGGAGGTCACGTGAGCAGCTCTCTGTGAACATG AACTTTTTATGATTTCCCAAGTGGCTATAATGCAGCATTCTTTTGG
$DreBB_{HJ}70_{b}$	50	22	CTATAAAACACAACAAATCATA
$DreBB_{JK}79_{bc}$	230	209	TCGTCATCATTTGTAACCATAGAGCATGAATTACCTCTTGAAGTCATC AGAGAGAATTTACGACTGGTCAACAAAGGCACGTGATGCCCTAACG CTCTCCCATATTTGGCCGCATACATGGCAAAAACGAAGTACAGTGCA TTGCTATAATTCATTATTACATCATAAATTGTGCGGCGCTGGATTTTA ACGACCAAGATCTACAAATC
$DreBB_{KO}87_{df}$	9,352	94	GTCACGTGAGGTCCATAAAGTTAGTTTTATGGTTTTAGGGGGGTTGAA AATGCGCTACATAATTCACATTCTTGAATGTAACTGACTGTTTAAC
$DreBB_{KO}88$	9,782	42	CCAACTTTGATCCACGCAAAACATATGCAACTAGTTTCCTTT
$DreBB_{KO}90_{abcd}$	8,468	82	ATAATGCGTTTTGCATATTGATAGGGTAATCTGGGCGCTGTTCTCATC AATTATTCATTGGGAAGGGATCACTCGGAGGTCA
$DreBB_{KO}92$	724	18	AGAATTTAGCGTTTACTT
$DreBB_{KO}93$	229	143	TTTGTCATGTAAATACGAATCGTTTGATGGACACCTGCAACGCCCAG AGTGCTACCGCAGGTCGCACACTATTGGCGCAAAATTAGTCACATGA TCTGTCACTTAACTGATATGAGCTCAGATACAACTTTGCGTGGTTTGA C
$DreBB_{KO}104_{abce}$	6,652	73	TCTATATATACCCTGTAGATCCGAATTTGTGTGAATATACAGTCGCAA ATTCGTGTCTTGGGGAATATGTAGT
$DreBB_{KO}115$	8,249	17	CACAACATTGAGAAACA
$DreBB_{KO}161_{bc}$	229	31	TTTGTCATGTAAATACGAATCGTTTGATGGA

Tetraodon nigroviridis spotted green pufferfish Hox Bb

PFC	Dist to 3' gene	Length	Sequence
$TniBB_{IJ}70_{b}$	57	22	CTATAAAAGACAACAAATCATA
$TniBB_{JK}72$	1,093	123	CATGTTAGAGACGCGTAGAGCGCTAGAGCTTCTTTGGGCCCCCGC ACGGCGGGGTGTCCTCTGCACCCAGGTCCTCCTCTGTTCGTCCTC GTGCTCGTGACTTTATTTAGGGGACTATGCAAA
$TniBB_{JK}73$	848	216	TTGTCCTTCCCTGTCCATCTTGTGCTGTAGTTGTCCTCTATTCGTGT GTAGGCTCGGTGTTTACGGCTGTGCCATCATCATCCACGCGTGTA CAGATGAGAAATGTTACCTGTTATTTATTGTTGTTTCTCCGTGTTC AATGCACATCTGGATATTTGTGCAATATTTTGCTTAGGAAATTGT ACATAAATAATTAAATAAAGGCTTTTTGATGTA
$TniBB_{JK}74$	532	29	CAATAACCCAGCAAAGAAATGCGCGTAAA
$TniBB_{JK}79_{abc}$	252	230	CCTAACGACCATTATTTCGTCATCATTTGTAACCATGGAGCATGA ATTACCTCTTGAAGTCATCAGTGAGGATTTACGACTGGTCAACAA AGGCACGTGATTCCCGAACGCTCTCCCATATTTGGCCGCATACCT GGCAAAGTACAGTAGGGCTTCATTGCTTATAATTCATGATTGGAA CCATAAATCGTGCAAGCACACAGGATTATAGCGACAAAGATCTA CAAATC
$TniBB_{KM}81$	7,329	114	GAAAAAGACAGGAGTCCATGAGTACTGTAAAGCTATTGAAAGCG CAGCACCTTCCAGCATGCCATTGTGGTAGAAGTAAAGTCATCTGT GGTCTCAAAACGCTGTTTTAGTTGA
$TniBB_{KM}82$	7,088	469	TCTCTCTTTTAAGTGTTTCAACTTGAGCTCTTTATTGTGACTTCTCG ACGGTCTAACAGGAGTAAAATTCACTGTACAGTCAGATAAAAGA TAAAAGAAAGAAGAAAGACACGTTTCCAAACCGTATGCTGCTCT TCACTGAATGTAATGTCCTCCGGAGGCCATTCAAAGCGCTTTGGT GTAAGTTTTCCTGCTATTTTTTGACCTTGTGGCCAGACCCTTTTCAT GAACGAGCCCATGTGTAGCTGTAGGAGAT
$TniBB_{KM}87_{de}$	6,247	85	GTCACGTGAGGTCCATAAAGTTGCTTTTATGGTTTTGGGGAGTAG ACAATGTACAATATAATTCACAACCTTGAATGAAAGTGAC
$TniBB_{KM}90_{bc}$	5,319	64	TTGCATATTGATAGGGGGTAATCTGTAGCCGGATCTCATGCATAA TTCATACCGGCGGGGATCA
$TniBB_{MO}93$	236	145	TTTGTCATGTAAATACGAGGCGTTTGATGGACGAGCGCGCAGCCT TGACAGAGCGCTCCGGGTCCTCCGCCATTGGCGTGCAGCGGGTC ACGTGGTGTGTCAGGAAGGCGATATGAGGGGCGGAAGGCAGTTT TACAGCTTTGAC
$TniBB_{MO}104_{b}$	3,420	22	ACCCTGTAGATCCGAATTTGTGT
$TniBB_{KM}131$	3,542	20	TGCAGCGCTGGTCCCTCTGC
$TniBB_{MO}161_{bc}$	236	31	TTTGTCATGTAAATACGAGGCGTTTGATGGA

Oryzias latipes medaka Hox Bb

PFC	Dist to 3' gene	Length	Sequence
$OlaBB_{IJ}70_{b}$	57	22	CTATAAAAGACAACAAATCATA
$OlaBB_{JK}72$	1,048	113	CATGTTAGAAGCGTGGTCTTTTATTTCTGTTTCTGCACGGCAACACTGT CCGCATAATCAGAGCCTCCTCTGTCTGTCTTTGTGCTTGTGTATTTATT TAGGGGACTATGCAAA
$OlaBB_{JK}73$	822	214	TTGTCCTTCCTCGTGAACCTGCTCCGTGACTCTGTGCTTTGTAGTCTGT AGGCTCGTCGTTTACACCTGTGCCATAAACGGTTCCATTTGTACAGAT GAGATATGTTTCATGTTATTTATTGTTGTTAATTGTGTGTTCAATGCAC AACTGGATACTGTGTGCAATATTTTGTTTAAAGGGAATTGTAAATATG AATAAAGGATTTTTGATGTA
$OlaBB_{JK}74$	523	28	CAATAAATTGACATAAACTCTGCGTAAA
$OlaBB_{JK}78$	738	24	CATTTGTACAGATGAGATATGTTT
$OlaBB_{JK}79_{abc}$	252	230	CCTAACGACCATTATTTCGTCATCATTTGTAACCATGGAGCATGAATT ACCTCTTGAAGTCATCAGTGAGGATTTACGACTGGTCAACAAAGGCA CGTGATTCCCGAACGCTCTCCCATATTTGGCCGCATACCTGGCAAAG TACAGTAGGGCTTCATTGCTTATAATTCATGATTGCATCCATAAATCG TGCAAGCACACAGGATTATAGCGACAAAGATCTACAAATC
$OlaBB_{KM}81$	10,164	205	GACAATAAACTGAAAAGCATGAGTCTTGTAACGGGAGGCAGCGCCT TCCACAACTGAACAACTTTTTGGGGAGAGTAAATTAGAGAACGTGAA AAAAACAATTACGAGTACTGTAAAGCTATTTAAAAAGCGCAGCACC TTTCAGCATGCCATTGTGATAGAAGAAGAAGTATTCTGTGCTCTTAAA CATGCCATTTTTAGTTGA
$OlaBB_{KM}82$	9,967	471	TCTCTCTTTAAAGTGTTTTCTGTCTATTGTATAACTTCTTGACGGTCTA ACAGGAGTAAATTCATTGTATAGTTACAAAAAAAGAAAAAAAAGAA TAAAAATATCAAAAGCAAGGCAACTTTTAAAACCATATGCTGCTTTT CATTGAATGTAATGTACTCCGGGGTCATTGAAAGCGCTTTAGTGTGAC TTTTACTGCTATTTTTGATGGTTTACTTGTGGCCAAGATCATAAACTTT CATAAACAAGCCAATG
$OlaBB_{KM}87_{de}$	9,057	85	GTCACGTGAGGTCCATAAAGTTGGTTTTATGGTTTTGGGGAGTAGACA ATGTACTATATAATTCACAACCTTGAAAGAAAGTGAC
$OlaBB_{KM}90_{abc}$	8,280	73	ATAATGAGTATTGCATATTGATAGGGGTAATCTGCAGCCGGATCTCA TGCATAATTCATGGTGGCGGGGATCA
$OlaBB_{KM}92$	637	17	AGAATTCAATTTTACTT
$OlaBB_{MO}93$	233	144	TTTGTCATGTAAATGCAAGCAGTTTGATGGACGAGCCTGCAGACTTG ACAGGCCGCTTTAGGTCATCAGCTATTGGAGTTTATCGGGTCACGTGG TGTGTCAGGAAGGTGATATGAGGGTGGAAGGCAGTTTTACAGCTTTG AC
$OlaBB_{KM}99$	6,922	22	CGCTGCACTGAATAGAGGCAGA
$OlaBB_{KM}117$	8,349	26	TTGGTTTTCAAAGCTGAAAATGTTTC
$OlaBB_{MO}159$	743	16	CCTTCAGTTAGAAAAT
$OlaBB_{MO}161_{bc}$	233	31	TTTGTCATGTAAATGCAAGCAGTTTGATGGA

PFC	Coelacanth C	Frog C	Human C	Zebrafish Ca	Medaka C	Tetraodon C	Zebrafish Cb
1					$OlaC_{CD}1$	$TniC_{CD}1$	
2				$DreCA_{CD}2$	$OlaC_{CD}2$	$TniC_{CD}2$	
3					$OlaC_{CD}3$	$TniC_{CD}3$	
4				$DreCA_{CD}4$	$OlaC_{CD}4$		
5				$DreCA_{CD}5$	$OlaC_{CD}5$	$TniC_{CD}5$	
6					$OlaC_{CD}6$	$TniC_{CD}6$	
7				$DreCA_{CD}7$		$TniC_{CD}7$	
8	$LmeC_{CD}8_{cde}$	$XtrC_{CD}8_{d}$	$HsaC_{CD}8_{bde}$	$DreCA_{CD}8_{abcde}$	$OlaC_{CD}8_{abcde}$	$TniC_{CD}8_{abcde}$	
9	$LmeC_{CD}9$		$HsaC_{CD}9$				
10	$LmeC_{CD}10$		$HsaC_{CD}10$				
11	$LmeC_{DE}11$		$HsaC_{DE}11$				
12	$LmeC_{DE}12$		$HsaC_{DE}12$				
13					$OlaC_{DE}13$	$TniC_{DE}13$	
14					$OlaC_{DE}14$	$TniC_{DE}14$	
15					$OlaC_{DE}15$	$TniC_{DE}15$	
16					$OlaC_{DE}16$	$TniC_{DE}16$	
17					$OlaC_{DE}17$	$TniC_{DE}17$	
18					$OlaC_{DE}18$	$TniC_{DE}18$	
19					$OlaC_{DE}19$	$TniC_{DE}19$	
20					$OlaC_{DE}20$	$TniC_{DE}20$	
21					$OlaC_{DE}21$	$TniC_{DE}21$	
22				$DreCA_{DE}22$	$OlaC_{DE}22$		
23			$HsaC_{DE}23$				$DreCB_{DE}23$
24	$LmeC_{DE}24$		$HsaC_{DE}24$				
25		$XtrC_{DE}25$	$HsaC_{DE}25$				
26	$LmeC_{E}26_{abc}$	$XtrC_{DE}26_{abcde}$	$HsaC_{DE}26_{abcde}$	$DreCA_{DE}26_{bcde}$	$OlaC_{DE}26_{bcd}$	$TniC_{DE}26_{bcd}$	$DreCB_{DE}26_{b}$
27					$OlaC_{EF}27$	$TniC_{EF}27$	
28				$DreCA_{EF}28$	$OlaC_{EF}28$	$TniC_{EF}28$	
29				$DreCA_{EF}29$	$OlaC_{EF}29$	$TniC_{EF}29$	
30				$DreCA_{EF}30$	$OlaC_{EF}30$	$TniC_{EF}30$	
31				$DreCA_{EF}31$		$TniC_{EF}31$	
32			$HsaC_{EF}32$				$DreCB_{EJ}32$
33		$XtrC_{EF}33$	$HsaC_{EF}33$				
34		$XtrC_{EF}34$	$HsaC_{EF}34$				
35	$LmeC_{EF}35_{abc}$	$XtrC_{EF}35_{ab}$	$HsaC_{EF}35_{ab}$	$DreCA_{EF}35_{abc}$	$OlaC_{EF}35_{ac}$	$TniC_{EF}35_{ac}$	
36	$LmeC_{FG}36$	$XtrC_{FG}36$	$HsaC_{FG}36$				
37					$OlaC_{FG}37$	$TniC_{FG}37$	
38	$LmeC_{FG}38_{ab}$	$XtrC_{FG}38_{a}$	$HsaC_{FG}38_{ab}$	$DreCA_{FG}38_{ab}$	$OlaC_{FG}38_{ab}$	$TniC_{FG}38_{ab}$	
39					$OlaC_{FG}39$	$TniC_{FG}39$	
40	$LmeC_{FG}40_{abc}$	$XtrC_{FG}40_{abc}$	$HsaC_{FG}40_{ac}$	$DreCA_{FG}40_{bcd}$	$OlaC_{FG}40_{bcd}$	$TniC_{FG}40_{bcd}$	$DreCB_{FG}40_{abc}$
41					$OlaC_{FG}41$	$TniC_{FG}41$	
42					$OlaC_{FG}42$	$TniC_{FG}42$	
43			$HsaC_{FG}43$	$DreCA_{FG}43$			
44				$DreCA_{FG}44$		$TniC_{FG}44$	
45				$DreCA_{FG}45$		$TniC_{FG}45$	
47			$HsaC_{FG}47$				$DreCB_{EJ}47$
48		$XtrC_{FG}48$	$HsaC_{FG}48$				
49		$XtrC_{FG}49$	$HsaC_{FG}49$				
50	$LmeC_{FG}50_{ab}$	$XtrC_{FG}50_{a}$	$HsaC_{FG}50_{ab}$	$DreCA_{FG}50_{ab}$		$TniC_{FG}50_{ab}$	
51		$XtrC_{FG}51$	$HsaC_{FG}51$				
52	$LmeC_{FG}52_{abc}$	$XtrC_{FG}52_{bc}$	$HsaC_{FG}52_{abc}$	$DreCA_{FG}52_{c}$	$OlaC_{FG}52_{abc}$	$TniC_{FG}52_{abc}$	
53					$OlaC_{GH}53$	$TniC_{GH}53$	
54			$HsaC_{GH}54$				$DreCB_{EJ}54$
55	$LmeC_{GH}55$	$XtrC_{GH}55$	$HsaC_{GH}55$				
56				$DreCA_{GH}56$	$OlaC_{GH}56$	$TniC_{GH}56$	
57	$LmeC_{GH}57$		$HsaC_{GH}57$				
58	$LmeC_{GH}58$		$HsaC_{HJ}58$				
59					$OlaC_{GH}59$	$TniC_{GH}59$	

60					$OlaC_{GH}60$	$TniC_{GH}60$	
61					$OlaC_{GH}61$	$TniC_{GH}61$	
62					$OlaC_{GH}62$	$TniC_{GH}62$	
63					$OlaC_{GH}63$	$TniC_{GH}63$	
64					$OlaC_{GH}64$	$TniC_{GH}64$	
65				$DreCA_{HJ}65$	$OlaC_{HJ}65$		
66					$OlaC_{HJ}66$	$TniC_{HJ}66$	
67	$LmeC_{HJ}67_{b}$	$XtrC_{HJ}67_{b}$	$HsaC_{HJ}67_{b}$	$DreCA_{HJ}67_{abc}$	$OlaC_{HJ}67_{abc}$	$TniC_{HJ}67_{abc}$	$DreCB_{EJ}67_{b}$
68					$OlaC_{HJ}68$	$TniC_{HJ}68$	
69					$OlaC_{HJ}69$	$TniC_{HJ}69$	
70					$OlaC_{HJ}70$	$TniC_{HJ}70$	
71					$OlaC_{HJ}71$	$TniC_{HJ}71$	
72					$OlaC_{HJ}72$	$TniC_{HJ}72$	
73					$OlaC_{HJ}73$	$TniC_{HJ}73$	
74			$HsaC_{HJ}74$	$DreCA_{HJ}74$			
75	$LmeC_{HJ}75$		$HsaC_{HJ}75$				
76			$HsaC_{HJ}76$				$DreCB_{EJ}76$
77	$LmeC_{HJ}77$		$HsaC_{HJ}77$				
78	$LmeC_{HJ}78$	$XtrC_{HJ}78$	$HsaC_{HJ}78$				
79	$LmeC_{HJ}79$		$HsaC_{HJ}79$				
80	$LmeC_{HJ}80$		$HsaC_{HJ}80$				
81	$LmeC_{HJ}81$		$HsaC_{HJ}81$				
82					$OlaC_{JK}82$	$TniC_{JK}82$	
83				$DreCA_{JK}83$	$OlaC_{JK}83$	$TniC_{JK}83$	
84					$OlaC_{KL}84$	$TniC_{KL}84$	
85					$OlaC_{KL}85$	$TniC_{KL}85$	
86					$OlaC_{KL}86$	$TniC_{KL}86$	
87	$LmeC_{KL}87_{abcd}$	$XtrC_{KL}87_{abcd}$	$HsaC_{KL}87_{abcd}$	$DreCA_{KL}87_{bc}$	$OlaC_{KL}87_{ac}$	$TniC_{KL}87_{ac}$	
88	$LmeC_{KL}88$		$HsaC_{KL}88$				
89	$LmeC_{KL}89$		$HsaC_{KL}89$				
90	$LmeC_{KL}90$		$HsaC_{KL}90$				
91	$LmeC_{KL}91$		$HsaC_{KL}91$				
92					$OlaC_{KL}92$	$TniC_{KL}92$	
93				$DreCA_{KL}93$		$TniC_{KL}93$	
94					$OlaC_{KL}94$	$TniC_{KL}94$	
95					$OlaC_{KL}95$	$TniC_{KL}95$	
96				$DreCA_{KL}96$	$OlaC_{KL}96$		
97					$OlaC_{KL}97$	$TniC_{KL}97$	
98				$DreCA_{KL}98$		$TniC_{KL}98$	
99		$XtrC_{KL}99_{ab}$	$HsaC_{KL}99_{ab}$	$DreCA_{KL}99_{b}$	$OlaC_{KL}99_{ab}$	$TniC_{KL}99_{ab}$	
100		$XtrC_{KL}100$	$HsaC_{KL}100$				
101		$XtrC_{KL}101$	$HsaC_{KL}101$				
102	$LmeC_{KL}102$	$XtrC_{KL}102$	$HsaC_{KL}102$	$DreCA_{KL}102$	$OlaC_{KL}102$	$TniC_{KL}102$	
103				$DreCA_{LM}103$	$OlaC_{LM}103$		
104				$DreCA_{LM}104$	$OlaC_{LM}104$		
105				$DreCA_{LM}105$	$OlaC_{LM}105$		

Latimeria menadoensis coelacanth Hox C

PFC	Dist. To 3' gene	Length	Sequence
$LmeC_{CD}9$	5,426	23	TTCTTTCCCTTTCAAGCTCCATG
$LmeC_{CD}10$	3,444	22	ACTAGACAGTGACCTTGAACTG
$LmeC_{CD}8_{cde}$	99	99	CACGTGATTCACTATGGACCAATGGTGCAGACCCTTAGGTTTA ACTATGTTTAATGTCAGATAGCAATAAAGTAGAAGCTCTTGGT CAGGCCCGCGGAA
$LmeC_{DE}11$	6,857	14	GTAATTACTTAAAT
$LmeC_{DE}12$	4,280	43	AAACGTTACAAGGAACATATTTTCTTGTACTTTCACAGTCTGA
$LmeC_{DE}24$	2,544	12	AAATCTGAATCT
$LmeC_{E}26_{abc}$	in 11	194	AAATCTGCAATTGATTTTCATAATGTTCCTGCGGTGTTTTGAAA CCAATCATTTGAACCTCAGAGATCTTACCTAAGTGAACCACTC CTACGTATCTAAGTGCTCCAAATTGATATATGGCAATATCTACT TTGGATCACGTGTTCTTAGGGAGACTTAGACGGGTAGCGCGTC ATCTCGCCTTCCCAAATTTT
$LmeC_{EF}35_{abc}$	269	230	CTTTGTTCAACTTGCAGCTCCTTGTTTTTTTTTTTACCCCGAAAA ATGTGGTAGATGTCGCTGTTGCACCACGTTCATCTTCCACATCT GAACAACTTGACCCCTGTGACGTCATTGGTGTCTGAATCATCA AGTCCATTTTCAGTTGTGATTGGTTTGAAAGTCACGTGGGAAGG CCAATGCGTTGATAATTATGTTGCTGATTTTTTTCCTAAAAAAA AGATGTCAGC
$LmeC_{FG}36$	8,999	29	TTTGCATGTATCCTATTTGTAAAGAAAAA
$LmeC_{FG}38_{ab}$	8,839	163	ATAACGTTGATTTAAATATTATCCAGGTGACCACAATAAGTCA AGGTCATAAAACAGTAATGTCAGGACAGTCCTGGTAAGCGCTG GAGGTGGATTTTATGATCTGCAAATATAATGTGCTGCAGCAGT AAAAGATGCATTTAAAGGTGACTGTGGAGGAGGG
$LmeC_{FG}40_{abc}$	7,965	71	CAGCTGATCTGTGGTTTAGGTAGTTTCATGTTGTTGGGATTGGC TTTTAACTCGGCAACAAGAAACTGCCT
$LmeC_{FG}50_{ab}$	750	18	TTTATGAACAATCAAATG
$LmeC_{FG}52_{abc}$	172	160	GCGCATTGATCCGCACTGTATTTTTGAGTAAATATAGTCACGTG AGTAGATGGAACCAATGGCAGGAAGAGAAGCCTTAAAAATAA TTACCTGCCCTGATTGTTCTATGAGCAGATAAAAAGTACACAT ACAGTCCATACAATAATCTTATGAATGTAAA
$LmeC_{GH}55$	3,457	138	GTTTGTCTCAATGCTGTTTGTCCTGAATGAAACTGAACAAAAC AAACCTCAGTTCTGGCTAGACGTCTGGCTTTAATTGTTTTATGG TTTAAATAAGGTGGATGCTTTCCTCTTTGAAACGGGATTACTGG AATGTTT
$LmeC_{GH}57$	286	129	GCCTGTCTTTGGCTGCTGAATTGATTAATAGAAATCGCGTGTGA AAGCCTTACCACCGGGATTTTGGAGAATTTGCCTGTGGTGTTAA AGGAAGAAGCAGTGAGGAAAACTGAGATAGAGCTCTACCTA
$LmeC_{HJ}58$	16,211	86	TGAAGTTCCGTTTTATGGATGATGTTAAAATGAAATGTTTACGA CCTTCACTTACTTTTCTAGTGAACACAATGGAGAGACTCACA
$LmeC_{HJ}67_{b}$	10,485	27	AATTTACAGCTGAGTAATAAAAGTTTA
$LmeC_{HJ}75$	5,922	17	AGATATTGTATGTCATT
$LmeC_{HJ}77$	1,578	87	AATAAACTGGTCACTTTCTTCTTTTATGCCTTTTATAGCTTTATT GTCACCATTAACTCGCTTACATTCTTTCACAGGAAAAAGATC
$LmeC_{HJ}78$	193	170	GGGAATCAACACAAGCAAAACCATTTTTTCCTTTCTTGTCATCT AGCTTGCTATTGGTCCAAATCAAGTCATCTGACTTTGTCATTTT GTCTGTCCTGGAGTGGAGCCGTCCCTATAACAATCTAGTTCAG AGTACAAACCGGAGCTCAGAAATAAATATTAAAGAAATC
$LmeC_{HJ}79$	2,483	52	TGGCACAAATTTTATTTTCTCTTAAAGTACATTTAATTAGCCAG AAGTGGAC
$LmeC_{HJ}80$	2,026	168	TGTGTGAATCATTGTAGTTGTAATTGTGACATCTGTATATTATTA TTTATGTTTGGTATTGTACTGATAAACAGCTAAAAGTGACTGAT TTCATGAAGTAGAAAATTTGTGTTTTGCTTTTAAAATGCACAGC AACCATATTGTTAAGTATTTAATAAAACTGATATT
$LmeC_{HJ}81$	238	229	ACATATCAGGCACGTTTTGCAGACCTCCATCAATAACCTCCTG GGGTCATCAAGCCAAATTTATGACTGGCCAACAAAATCACGTG ATTCTATTTAAACATCCCATATTTGGGCAGCGCACGTAGAATA AAGAAAAGAAGAATCTCCACCTATAAATTGTGCACTTTTGAGA CAAAACCCTTAACTTCAAAGGGTCACAAATCAAGCTTAATCA AAAGGTGTAAATTTT
$LmeC_{KL}87_{abcd}$	21,028	205	AGGTTAATTATGAGGATGGCCCTTAGCAAAGGCCCTGAATGGC TCAAAAGGAACACGTGATGTCATTAAAGTTAGTTTTATGGTTTG GGGGAGCTGACAAACCCACAATATATTTACATCATATATAACC TTAACTGTCCACCATCGCAGCTGCTAGGGCTTTTTAATGCTGGG ACTGTGGGACAAGACATTTAGTGCGGATTAC
$LmeC_{KL}88$	17,002	36	CAAAAGCATTGGAGCTGCCTTCTAAGTAGTGATTGG
$LmeC_{KL}89$	16,659	31	GGATTTGCATGTCTATTGGGCTTGTTGGATT
$LmeC_{KL}90$	13,346	20	TTCTGGGAAGATATTCAAAC
$LmeC_{KL}91$	10,496	18	TAGGGAGGAGTAATATTT
$LmeC_{KL}102$	157	151	GGTTCCTTATCCGGGGACTAGCTCCTTTTGTCTCATTGGATAAC TAAGTCACATGGCTAAACTAACTTTACAGGGTCGCCAGCTAGT AGGAGGGCTTTATGGAGCAGAAAAACGACAAAGCGAGAAAA ATTATTTTCCACTCCAGAAATTA

Xenopus tropicalis western clawed frog Hox C

PFC	Dist to 3' gene	Length	Sequence
$XtrC_{CD}8_{d}$	47	28	AATGTCAGATGGCAATAAACTAGAAGCT
$XtrC_{DE}25$	2,849	63	TTTTATGGTGGTCAGGAAATTGACATTGGTCAAATTCAAGTGCTCT GACTCTCCAGCTGTCTC
$XtrC_{DE}26_{abcde}$	257	255	AAATCTGCAATTGATTTTCATAATGTTTCTGCGGTGTTTGAAAACC AATCGTTTTTGAACCTCTCTCAGATCTTACCTAAGTGAACCCTCCT ACGCATCTAAGTGCTCCGGACTGATATATGGCAATATCTACTGGG GCATCACGTGTTCCTGGGGAGAGACTAAGACGGACAGCGCGTCA TCCAGCCATCCCAAATTTTTCCCCCTCTGCAGATCGCCTCCAAAA CATCTATCTATAGAGCCAGAAAGGGAGAA
$XtrC_{EF}33$	9,061	59	CATTCTGCTCTTACCTAATTTTGTTTGGCCCAAAGAATGCTACTGT ACTTTGTATCCTC
$XtrC_{EF}34$	8,882	21	GTGCACAATTATATAAGTTTA
$XtrC_{EF}35_{ab}$	267	200	CTTTGTTCAACTTTAAGGTCCTTGGGTCTATACCTCACAAAGTCAG ATAGATGGCGCTCTTACTCCACGTTCATTTTCCTCCTTTCGACAAC TTGACCCCTGTGACGTCACCTGCGTCTGAATCACCAAGGCCATTT TCAATCCTCATTGGCTTGGGAGTCACGTGGTGGGGCCAATGCGTG GATAATTATGGTGCTGAT
$XtrC_{FG}36$	9,236	30	TTTGCATGTATTTTCCTTTGTGGAGAAAAA
$XtrC_{FG}38_{a}$	8,986	149	ATAACGTTGATTTAAATATTATCCAGGTGACCACAATAAGTCAAG GTCATAAAACAGTAATGTCAGGACAGTCTTGGTAAGCGCTGGAG GTGGATTTTATGATCTGCAAATATAATGTGCCGCAGCAGTAAAAG ATGCATTTAAAGGTG
$XtrC_{FG}40_{abc}$	7,734	74	CAGCTGATCTGTGGTTTAGGTAGTTTCATGTTGTTGGGATTGCTTTT TCTTAACGCGGCAACAAGAAACTGCCT
$XtrC_{FG}48$	4,580	37	ATTGGGTCATTAAAAACAAAAAGGAGGTAAAAGGCGT
$XtrC_{FG}49$	1,874	29	GGGGAAAGAAGCATGGATCAGCTGCTGAA
$XtrC_{FG}50_{a}$	826	12	TTTATGAACAAT
$XtrC_{FG}51$	294	26	ATGGGTGCACAGGCAGAGCCCAGATG
$XtrC_{FG}52_{bc}$	154	141	TTTTTGGGTAAATACAATCACGTGGGACGCGAGAGCCAATGACA AGCGTGAAAGGCTGAAAAAATAATTACCTGCCCTGATTGTTCTAT GAGCAGATAAAAAGTACACATACAGTTCATACAATAATCTTATG AATGTAAA
$XtrC_{GH}55$	3,420	142	GTTTGTCTCACTCATGTGTGTGTCTTTGAATGGATGTGAACAAAAC AAGACCATAGGACTGGCTAGACGTCTGGCTTTAATTGTTTTATGG TTTAAATAAGGTGCATACTCTGCTCTTTGAAACGGAATTATTGGA ATGTTT
$XtrC_{HJ}67_{b}$	9,009	27	AATTTACAGCTGGGCAATAAAAGTTTA
$XtrC_{HJ}78$	194	170	GGGAATCAACACAAGCAAAACCAATTTTTCCCTTTCTTGACATCT CGCCTCCTATTGGCTCCAAGTGGGTCAGCTGACTTTGTCATTTTGT CTGTCCTGGATTGGAGCCGTCCCTATAACAATCTAGTTCAGACTA CAAACTGGAGACAGAAATAAATATTAAAGAAATC
$XtrC_{KL}87_{abcd}$	36,678	205	AGGTTATTTATGTGAATGGCCCTAAGGAAAGGCCCTGAATGGCTC TCACGGAGCACGTGATGTCATTAAAGTAAGTTTTATGGTTTGGGG GAGCTGACAAAGCTACAATGTATTTACATCATATATAATCTTAAC TGTCCACCATCGCAGCTGCCGGGTCTCTTTAATGCTGGGACAGTG GGACAAGGCACTTAGTGTGGATTAC
$XtrC_{KL}99_{ab}$	4,718	163	ACATTATAACTAGTTATTGAACTAGCTCTGAGATCTAAAGGCCAT TTGTGCTGAGACAGATTTCAATGGAGTTTGCCCATCAATAATCTTT GGCAGTGACCTATTGGTGGAAGTCAAGCAACCGAGGTGAAATTC AGGTCACGCTGTCTAACCGGTATTAAAA
$XtrC_{KL}100$	2,806	56	TCACAGGACCAACTGACATAGTGGGGTAACCCATAGTTATAGGA TTGTTTGGGAGA
$XtrC_{KL}101$	1,805	257	GCAGCCACCGCAACGAGCAGGTCCCAACCAGTGCAAAGACTCT AAACCATTCCTCCTGCTTGATTTATGGCTTTTTACTGCCCTATAAA AGCTGTTACAAGGAACCTAAGCCTGCAACACCGCTGGCACATTT AGGCTAATAAATAAACATAAACAGCTAACTTTATGTGTCACTTT TATTGTTATCAAATAAAATATTGCATTGCCCCTGCAAGCTATGAT TCCGTGCTATAATTGTTATCAAGCAGAACAAG
$XtrC_{KL}102$	156	151	GGTTCCTTATCCGGGGACCTGAGCGCACGTTGTGATTGGCTGGAG GAGTCACATGGTGAAAGTAACTTTACGGGGTCGCCAGCTAGTAG GAGGGCTTTATGGAGCAGAAAAACGACAAAGCGAGAAAAATTA TTTTCCACTCCAGAAATTA

Homo sapiens human Hox C

PFC	Dist to 3' gene	Length	Sequence
$HsaC_{CD}9$	5,643	23	TTCTTTCCGCAGATCCCTCCATG
$HsaC_{CD}10$	3,635	22	ACTAGACAGTGACCTTGAACTG
$HsaC_{CD}8_{bde}$*	172	172	GAAGTTAGAGTGCGGGATGGGATGGTGGGGGGGGGGGATCGGT TGTCCCCACCCCTCCCCCTGGCGGCCGTGCCCACGTGAGTGGGG CGGCCAATGGGTGACTGGTGCAGATTTAACTATGTTTAATGTCA GATAGCAATAAAGTAGAAGCTGCCGGTCGGGCCCCGCGGAA
$HsaC_{DE}11$	9,984	13	GTAATTCTTAAAT
$HsaC_{DE}12$	6,843	42	AAACGTCATAAAGAAAAGTTCATTTACAGCTTAAATGTCTGA
$HsaC_{DE}23$	6,264	22	AGAAAGAAGGAGAGGAAAATAA
$HsaC_{DE}24$	4,786	17	AAATCTCTGGCGAATCT
$HsaC_{DE}25$*	2,505	65	TTTTATGGTGGCCCAGGAATAAACTTTGGTGGTCAAATTCAAGT GCTCAGTCCAGCAGCTGTCTC
$HsaC_{DE}26_{abcde}$*	259	257	AAATCTGCAATTGATTTTCATAATGTTTCTGCGGTGTTTGCAAAC CAATCGCCTGAACGTCCCCGATCTTACCTAAGAGAGAACCCCT CCTACGTCTGCGAAGTGCTCCGAACTGATATATGACAATATCTA CTTTGGATCACGTGCTCAGAGAGAGAGAGACTAAGACGGATAA CGCGTCATCTCGCCTTCCCAAATTTTCCCCCCTCGCTAGACCGG GTCCAAAACCTCCATCCGGAGCCGGCAGGAGAGGAG
$HsaC_{EF}32$	9,811	20	GGAAAATTATTTTATTTTAT
$HsaC_{EF}33$	8,852	64	CATTCTGTTCAATATTATGATTTTATTTGACCCTTTTATTAACCCC TCTCCTCTTAAAATCCTC
$HsaC_{EF}34$	6,692	17	GTGCACGTGGGAGTTTA
$HsaC_{EF}35_{ab}$*	330	204	CTTTGTTCGCGGGGAAGGGCTCCGGTGCCCCTACCCCGAGGCAG CTGCTAGATGGCGCTGTTACTCCACTCTGCGCGCTCCGCCTGCC GACAACTTGACCCCGCTGACGTCACGGCCGTCTGAATCATCAA GGCCATTTTCAAATCCCATTGGTCTAGCCGTCACATGGTGAGAA CCGAATGCGCGGATAATTACGGAGCTGAT
$HsaC_{FG}36$	9,975	30	TTTGCATGTAGCTTCCTTAATGGAGAAAAA
$HsaC_{FG}38_{ab}$*	9,764	161	ATAACGTTGATTTAAATATTATCCAGGTGACCACAATAAGTCAA GGTCATAAAACAGTAATGTCAGGACGGTCTTGGTGCGCGCGAG AGGTGGATTTTATGATCTGCAAATATAATGTGGTGCTGCAGTAA AAGATGCATTTAAAGGTGACTGGAGGAGGG
$HsaC_{FG}40_{ac}$	8,443	72	CAGCTGATCTGTGGCTTAGGTAGTTTCATGTTGTTGGGATTGAGT TTTGAACTCGGCAACAAGAAACTGCCT
$HsaC_{FG}43$	7,032	30	TGTTTTGACTCCCAAGGAAGCTGGAGGAGA
$HsaC_{FG}47$	6,650	18	TCTCTGTGCCCCCATTTC
$HsaC_{FG}48$	5,053	129	CTTTCTGTCTGAGACGCCCCAAACCCTGAGCTCTCCTTCCCAAG GCTGGAAAAAAAATCGGGAAAAACGGCGGAATTTGCACTTTCA TTCATTGGGTTATCTGCAACTGAGAGGGGCTGGTTAAGGCGT
$HsaC_{FG}49$	2,076	29	GGGGAAGACAGTGGCGCCAAGGAGCTGAA
$HsaC_{FG}50_{ab}$*	1,183	18	TTTATGAACAATCAAATG
$HsaC_{FG}51$	347	34	ATGGGTGGACATGGGCACAGGGGCTTCTCAGATG
$HsaC_{FG}52_{abc}$*	175	164	GCGCATTGATCCGCGCCGTATTTTTGGGTAAATACGATCACGTG GGGGCCGGGGAGCCAATGAGCTGTGGGGAAAAGGCTGGAAAA ATAATTACCTGACTTGATTGTTCTGTGAGCAGATAAAAAGTACA TATACAGTTCATACAATAATCTTATGTATGTAAA
$HsaC_{GH}54$	5,999	42	TGTTATGGTAGAACAGCCCTGTGTTAATATATTGAAGTCACT
$HsaC_{GH}55$*	3,848	137	GTTTGTCTCCCTGTTCTGGGTTCTCAATGGGGCCGAACAAAACA GCAGCGCGGAGCTGGCTAGACGTCTGGGCTTAATTGTTTTATGG TTTAAATAAGGTGGACACTTTCCTTTGAAATCGGATTATAGGAA TGTTT
$HsaC_{GH}57$	400	126	GCCTGTCTTCATGTCGTGGATTGATGAACGCGAATCGCGTGTAA GCGCCGCCACCGCCGGGAGTCTGAGGAATTCGCCTGGGCTGTTA GAGGAAAGAGCTAAGTGAGAGAGCGCGAGCTCTACCTA
$HsaC_{HJ}58$	17,102	83	TGAAGTTTCGTTTTATGGTAGCAGATAAATTGAGAAGTTTACGA CTGTCATTTGCTTTTATAGAGAATAGAATGACACTCACA
$HsaC_{HJ}67_b$*	11,757	27	AATTTACAGCTGAGTAATAAAAGTTTA
$HsaC_{HJ}74$	7,266	54	TTCTCTAAGGTCCCCCAAATACCCAGATCTGGTTACTGGATGAA CTGCGTTTTA
$HsaC_{HJ}75$	6,257	17	AGATATAGCATGTCATT
$HsaC_{HJ}76$	6,080	27	TATTAAATTTAAAAGAGTATCATTGAA
$HsaC_{HJ}77$	1,717	90	AATAAACGAGTTTTTTCTCAGCATAGTCCCGTTTATGGCTTTATT GCTACCCATTGATTACTCCGCTTTGTTACACAGAAGGAAAAGAT C
$HsaC_{HJ}78$*	200	173	GGGAATCAACAGAAGCAGAAGCGATTTTTTTCCCCCTTCCTGAC ATCTGGCTTGCGATTGGCTGGGAGGGGGTCAGCTGACTTTGTCA TTTTGTCTGTCCTGGATTGGAGCCGTCCCTATAACCATCTAGTTC CGAGTACAAACTGGAGACAGAAATAAATATTAAAGAAATC
$HsaC_{HJ}79$	3,075	57	TGGCACAATTGATGTGTTTTGATTCCCTAAAACAAAATTAGGGA GTCAAACGTGGAC
$HsaC_{HJ}80$	2,561	247	TGTGTGAAGATTTTTAGCTGTATTTGTGGTCTCTGTATTTATATTT ATGTTTAGCACCGTCAGTGTTCCTATCCAATTTCAAAAAGGAA AAAAAAGAGGGAAAATTACAAAAAGAGAGAAAAAAAGTGAAT GACGTTTGTTTAGCCAGTAGGAGAAAATAAATAAATAAATAAA TCCCTTCGTGTTACCCTCCTGTATAAATCCAACCTCTGGGTCCGT TCTCGAATATTTAATAAAACTGATATT
$HsaC_{HJ}81$	255	238	ACATATCGAGATGCTTTTCGCCGGCTTCCATCACTAACCTCCCG GAGGTCATCAAGCCAAATTTATGAGTGGCCGCTCGAGTCACGTG ACTCTATTTAAGGCTCCCTTATTTGGGAAGAGCGCATAGGATAA AGAAAGAGATATCTCCACCTATAAATTGTCCACTTTGGAGAACA AAAAACCCCTCAACTTCAAAGAGTCACAAATCACCCTTAATCA AAAAGGGTGCAGAAATTTT
$HsaC_{KL}87_{abcd}$*	16,174	206	AGGTTATTTATGTGAATAGCCTGAAGGCAAGCCCGTGAGTGGCT CTTGGGGGACCACGTGATATCATTAAACCAAGTTTTATGGTGTA GGGAGAGCTGACAAACCCACAATATATTTACATCATATATAAT CTTAACTGTCCAGCCACGCAGCTGCTGGTGAGGTTTAATGCTAG GACAGAGGGCCTGGCAGTTAGAGGGGATTAC
$HsaC_{KL}88$	13,050	33	CAAAAGTGGAGATTTTTTGAAAGTTCTGATTGG
$HsaC_{KL}89$	12,787	29	GGATTTCAGTGTTTGAGTTTACATGGATT
$HsaC_{KL}90$	10,282	17	TTCTGGAAAACTCAAAC
$HsaC_{KL}91$	8,173	18	TAGGGAACAAATATATTT
$HsaC_{KL}99_{ab}$*	4,717	164	ACATTATAACTAGTTATTGAACTCGCTGGAAGATCTAAAGGCCA TTTGCGAAGAGACAAATTTCAATGGAGTTTCCCCATCAATAACC CCATGGCAGTGAACTATTGGAACAAGTCAAACACCCGAGGTGA AATGCAGGTCACTCTGTCTAACCAATATTAAAA
$HsaC_{KL}100$	2,751	58	TCACAGCTCACCTTTCCAGGTGAGGCAGCCGCATCCAGGTACA GGCCTAGAAGGGAGA
$HsaC_{KL}101$	1,720	262	GCAGCCAGCGGCCGCCAGCCCTGCGCCCGCCGCCGACACAAAG AGTGCGGGCGATTCCGCGAGACTGATTTATGACGTTTTACAGCC CTATAAAAGCTGTAGCGACGAGGCAAGCGCTCTACCACCGCT GGCAGATTTAGTCTAATAAATAAAACATAAACAGTTAACTTTAT GTGTCACTTTTATTGTTATCAAGTAAAATATAGCTGAGCCCTGGC AAGCTATGATTTTAAACTATAATTGTTATCAAGTACA
$HsaC_{KL}102$*	151	151	GGTTCCTTATCCGGGGACTGGGTTGCTCCGTGTGATTGGCCGGA GGAGTCACATGGTGAAAGTAACTTTACAGGGTCGCTAGCTAGTA GGAGGGCTTTATGGAGCAGAAAAACGACAAAGCGAGAAAAAT TATTTTCCACTCCAGAAATTA

Danio rerio zebrafish Hox Ca

PFC	Dist to 3' gene	Length	Sequence
$DreCA_{CD}2$	4,973	167	TACCCTTGACCATGACTATGCAGTCGCTTTGACCTTGACATCAC AGAGTCTTGTGAATAATAAACAGAGGGGCTGGACACCACTGCC GTGGTTTGTTTTCAGGCAAACAAGGTGTTATTTCCCTGTCTAGA CGGCGAGAGTTAATGACATACCATAAATGCATAAAA
$DreCA_{CD}4$	3,046	23	AAAATATTCTAAGCTCGAAAAAT
$DreCA_{CD}5$	2,422	93	CTTTGTTTACGCTATGGCCCTGTCTGGACGGAAAAGGATTCCAA GACAGCAATCTTACAGCGGCGCAGACAGTGCACTAGACACTG ACCTTGA
$DreCA_{CD}7$	1,302	43	AATCCAAATTTAAAATTATAAAATATATTTACATAATATATTT
$DreCA_{CD}8_{abcde}$	775	775	CCTCTGCCAGCTTGTAAAATTCAAATAAAAGAATTGCGAGATG TGTAACAATCTACCATGAGTTTGACTGCGAGTATTTATAGGATG TATCTGTTAAAGTTGTTCATGTGAGGTGTACTAGAGCATATTAT CTAAAATATAGGCTACGCTCTTAAACATTTCTTATAGAATATCA TCATCTTTTACGGAGCCAATGGTTTCAACCTGCTGACCCCACTA CAGTAAATATGAGTGGAGTTCAGCACTGAGCATCAG
$DreCA_{DE}22$	1,975	206	GTTAGATATTTGTCTTCCTTAGGCCTTTTTAATTATTTAACTTTTA TTTGTTTTCTGTAGAACAAAGAAGTTCTTGCTTTTCCGCCGTTCG GACTTTTATGGCGGCTGCAGTGAGAATTTTAGTGGTCAAATTCA AGTGCCTGGTCTGGGCAGCTGTCTGCTTTCTTTTGTACTTGGTTGA AGGCATATTGCACACCAACCTTTTGT
$DreCA_{DE}26_{bcde}$	167	165	CTCCTACGCATCCATGTGTTCCAAATTGATATATGACAATATCT ACTTTCGATCACGTGTTGCGGGGCGACTTAGACGGATTGCGCGT CATCTCGCCTTCCCAAATTTTCTCCTTCTGCAGCTCGAATCCAA AACATCATATCTGTAGTACGGACTACAGGAGAA
$DreCA_{EF}28$	5,968	286	TTTTCACCATTTAGAACTGTGACTTGTACAATCTGTAATAATGT ATTCCTTGTACAGAGATTATGCATTGTGCATTCGTGAATGTAAA TATAATGTTATTTTGAGTATCCTCAACTTATCCCTGCTGCTATAT TTCTTTACATTGCCTTTGCAGCCTTGTGAGCTTTGTTGTTCAATT CTTTAAATTATTTGCTGTACGCCCAAAGTTCTCTATCTTTTTTTTT TTCTTAAGGGGAAGTTCGGCGGTCCGATTTG
$DreCA_{EF}29$	5,234	145	CTTTTAGTACCTTCTAAATGGAATCGCCACTCTTCTAGTCTAGA CACTGCAATAAAGCTAGGGCCCTAAATTGCCTTTTTCTGACTTA ATTGCACCGTCTGTTCTAGGGGAATGGTGTTCACTTCCGCAGTA AATAGTAAACAAG
$DreCA_{EF}30$	4,757	253	TATAATTAAAAGTTGTCTTTGAACTTCAATAATGTCAAGGCCTT CACCTTTAACCTATTGCAATAAAGCGAAATCAAGGAGCCCCTT TTGTCAGGAAACGTGGTGAACAGAACAGATCCAATCAGCGCA AGTGGGAAAGAACCAAGGAATGTCTAAGAAGTAACACTTTGA CCTCTCGCATTAAACTACAGCAGGTAACTTTTCCAGTTTTAGCA CTCCCATTAAATGCACTTTATTGTCTCTGTAAAATGTA
$DreCA_{EF}31$	4,091	71	CAAGGCTGGAAGGCTTGTCTTTTTGTCACACCCACCAACCTTGG CCCTGGCTCAGTGACGTAGTTAATTTG
$DreCA_{EF}35_{abc}$	297	223	CTTTGTTCTGCTGTTGGCTTCCTTGTTCTCAATACACGGCTAGAT GGCGCCATTGCTCCATGCTGCACCACATAAATTCAACAACTTG ACCCCAGTGACGTCAAGACGGTCTGGCGCATCAAGGCCACTTT CAAAACCTTATTGGTCTGAAAATCACATGACAAAACGCCTTGA ATCCATAATTATGTTGCTGATATTTTTCGCCCCCCCAAAAGATG TCAGC
$DreCA_{FG}38_{ab}$	8,336	160	ATAACGTTGATTTAAATATTATCCAGGTGACCACAGTAAGTCA AGGTCATAAAATTGTAATGTCATGACGGTCCTGGAAAGCGCTG GGGGTGGATTTTATGATCTGCAAATATAATGTGCTGCTGCAGTA AAGATGCATTAAAAGGTGAGTGGAGGAGGG
$DreCA_{FG}40_{bcd}$	7,264	157	GTGGTTTAGGTAGTTTCATGTTGTTGGGATTGGCTTCCTGGCTCG ACAACAAGAAACTGCCTTGATTACGTCAGTTCGTCTTCATCAA GGGCGACAATTGCCGTTAAATTACACCCTGTGCAATGAAGTTT CCGCAAATGGCATTGTAAATGAACTG
$DreCA_{FG}43$	5,670	20	TGTTTTGCGACGGGAGGAGA
$DreCA_{FG}44$	824	64	TTATTAACGACACGTTTATGAACAATCAAATGGTCCTCGTAAA AATTTATTGAAGGACATAAAA
$DreCA_{FG}45$	409	29	GACAGAGATTGATTACACCAACAAATAGT
$DreCA_{FG}50_{ab}$	810	18	TTTATGAACAATCAAATG
$DreCA_{FG}52_{c}$	100	82	AAATAATTACCTGTCTTGATTGTTCTACGGTCAGATAAAAAAGT ACACATACACTCCATATAATAATCGGATGCATGTAAA
$DreCA_{GH}56$	2,723	206	AAATTGCCTTTTTGTGGGCCTTAACAATACGAACAGGATGCGGT GATTTCAAAGCCTGATCGACCACGTTTCCGCCCTAGAACTAAA ATGCCAGTTTTACAGCCCTGTTGGAGCTTGGTGTTGTTTGTCTCT GATGCAGACGAACAAAGCCAACCTGGCTAACTGGCTAGACGT CTGGGCTAAATTACTTTATGGTTTAATGGACG
$DreCA_{HJ}65$	11,180	391	TAAACACTTATGATTATCGAGCATTTTGTCGTCTTCGGAAGACA CGAATAGCTCGAAGCCAATCTTAACTATTTTAAAAGATTGTATC ACGGCACTACCTAAAACAAAGTTTTGCTAGAAAGAATGAATTT ATTTATTCTACGTGCTGCTTTGTTAAAAATAATTATACATTTTCC CCCAACAACTTAAACTGCCTATAGAAAATGCACATTTATGTGT AGTTGAACAAACACGCTGATTACTACCCTACCTTAC
$DreCA_{HJ}67_{abc}$	6,174	130	ACATTCCTAGTTGTTTAGAGGGGGAATTTACAGCTAAGTAATAA AAGTTTACGACTGAATACACGCGGTCATTGGTTGCGTCAGACC ACGTGGTGTTCACTCTATGAACATGAACTTTGTGCTGTTGTCT
$DreCA_{HJ}74$	2,864	53	TTCTCTCCAGTCACATCAACATACCAACTGCAGACTGAGTGTG GTGTGTTTTA
$DreCA_{JK}83$	482	226	TGACGCTCTTGTCTAACGGTCTCGCCCTTTTTAGCCAAGACCCT TGCATGATTTCCATACCGAGAAGTTATCGGACACGTTTCCCTGT CTATCAATAACCTCCTGGGGATCAAGCCAATTTATGACTGGCC AGGAGCTGCACGTGATTCTATTTAAACATTCCATATTTGGGCAT TACACGTCGTACCAAGAAAAAAAGAAAATGATTTCCTCCACCT ATAAATCC
$DreCA_{KL}87_{bc}$	8,555	91	TCATTAAAGTGGGTTTTATGGCCTGCAAGACCTGACAAACCCTC AATATATTTACATCATATATAATGTTAACTGTCCGTAATCGCAG CTG
$DreCA_{KL}93$	4,817	19	AAAGAATTTATTTATTTTT
$DreCA_{KL}96$	3,001	50	TATATTTCAGAATAGCTTAGGTCACTTAAACTACAATTCACTCG ATTTTT
$DreCA_{KL}98$	2,893	32	ATTAAATATCATATAGGAGTAAAATGTCATTC
$DreCA_{KL}99_{b}$	2,503	87	CATCAATAACTCCTTGGCAGTGAACTATTGGAACGAGTCGAAC GCGAGGGGTGAAATGCGGGTCAGGCTGTCTAACTAATATTAAA A
$DreCA_{KL}102$	151	151	GGTTCCTTATCCGGGAACTACATCGCCCCATGCCATTGGGCCAT CGAATCACGTGGTAAAAGTAACTTTACAGGGTTGCTCGCTAGT AGGAGGGCTTTATGGAGCAGAAAAACGACAAAGCTAGAAAAA TTATTTTCCACTCCAGAAATTA
$DreCA_{LM}103$	12,836	310	TTATTTATAAAATATAATATATTTTGTTTTCGTAGGTATCTTGTG CGTAAAGGGGGTTCCTTTAAGTCCACAGTTATATAAAATGCAT GTTATATAGCATGGATTATTGCGAATACCTATGGATTATTTTTTA CGCGACACAGGGTTTAATTTATCTAAAACTGAATCAATATGAT GTTTGTTTTTGTTATATGTTTTTTCAAGAGAAATGTTTTGGGCGA AGCAGGTTATGGACATTTTATTCAAATGTTATTC
$DreCA_{LM}104$	12,328	132	GTTGTGTTTAGGCTATTAACAGCAGATTGTTGTGAGATGCAATA AAACCAGTTTAGTGTATTTGTGCTAATTATATTGGTCAATAAAA CGTAAGTGTCTATTAGTTTAACCTAGACTTGGGTGACTTTTATT
$DreCA_{LM}105$	10,733	402	TTATGTGACTGAACAGCAGAGAGAAATGAATGTGCACGCCTAA ATCATAGCATCCCGGCAGTCTCTTTCGAGACGCTGTCGCCTCCA TGTCGCGAGGCTGAAAAGATACTTTATTGAAGTTTGTGGCAACT AAGAAAAGTCGAGTTCAATTTGAGGACACATTTTCTGTCCCATT AGTTTCATTTCAGCCGGGTCGGCCGACCTCTCACACCCCTTGAC CCCACTGGCTCTCTCAACGTGTACCTTCTGTGTGCA

Oryzias latipes medaka Hox Ca

PFC	Dist to 3' gene	Length	Sequence
$OlaC_{CD}1$	6,532	172	TCCATCAACTTTCTCGACCACGTCAACAAGACCCCTGCTCCCAGT TTAGAGAAAATGCTCTAAAAACTGTGAAAATGTTCAAATCGCTTC ACAATTTTGTTTTATAAGAATGTACATATATAAATATATAATATTT AAAACGATATCTGTATTTCAGACAAGTTATTGCGTG
$OlaC_{CD}2$	4,553	167	TACCCTTGACCATGACTATGCCGTCACTTTGACCTTGACATCACAG TAGATTGTGAATAATAAATAGGGGGCTGGACATCAAGGCTTTCGG GTTGTTTGTAGGCAAACAAGGTGTTTTCCTCCTGTCTAGACAGCCG GGCTTAATGGCATGCCATAAATTAATAAAA
$OlaC_{CD}3$	3,363	209	GGCTGATAACCTCTGGTGGTGACCAGAGGAAAATATCAGGAAGT GCTGCAACTGTTGCTAATTTATGGATCTTTTGGCTCCCCCAATTTG ATTTATTTTAATGTAAGCAAATTTGGCTTTTATTGCTGGAAAACAT TTTAAAATATTTTTTTTACATTTTATTTATGTTTATGCAATTCCTTAC AAATTCTGCACTCACTCACTCAGAC
$OlaC_{CD}4$	2,726	22	AAAATATGTATACATAAAAAAT
$OlaC_{CD}5$	2,169	95	CTTTGTTTAGACTATTGCCTTGTCTGGACGGGGCACAGGATTCCAG ACGGCTGATTTCAATACTGCGCAGACGGTGGACTAGACAGTGACC TTGA
$OlaC_{CD}6$	1,659	190	GCTTCTCCGGTTAAAACATTTCATATTTTCCTTCTATTTAAGGGCA GAGAGAGCTGCGGTGTCTTTGCTCAGGATACTTGGTCACAAGTGA CAGATACGGCTCTCAAATATCGTCTGTGTTGCAAACCGTTCTCAA GTCTGTGTTTAAATGAATCCTAATGTTCTCACATAACACTTCATTG GCACATTT
$OlaC_{CD}8_{abcde}$	741	741	CCTCTGATATCTTGTAAAATTCATATAAAAGAAGTGTGAGAGGTA TAACTAGCCCACCAGAGTTTGGCTGCGAGCGTTTATTGGAGTTATC TGTTAAAGTTGCTCGTGTGAGGTGTACAAGAGAATATTGTGTGAA ATACACCCTCTTAAAGTAACCTTGTAGAATATCGTCTTCTTTTACA GAGCCAGTGGTTTCAACCTCCTGACCTCCTTACTGTAAATACGAGT GGTGTGCGGCAATGAGCATCAGTCTTT
$OlaC_{DE}13$	9,455	14	AATGCATGCACTCA
$OlaC_{DE}14$	8,909	19	GTTCACGGTCACGGGCAGG
$OlaC_{DE}15$	8,685	16	AAAACGGGTTTGAGCT
$OlaC_{DE}16$	8,405	110	AGCATGTTCTTGATGTGAAGTAGATATGACTCATTTTCTGTTTTGCG GGCCACTTTTTATGCAGGTGACCCTTTTTCTACTTTTTTTTCACTTA AGTAAAATATAAAAGC
$OlaC_{DE}17$	8,162	119	GTATAACCTTGTAAAATAAAATCACACAAATTAGTATTTGCATCC TGATTAAGGATACCTTGCCGCTATATAATAGTGCTTTATGTTATTG ATCGTCTCGTACAATATGGCATTCCATA
$OlaC_{DE}18$	7,900	78	TTTGCGCTTGTAAGACTGCTACGTTGTCTTGTAAAAGTAATAACGA TATGAATGTATATGCGTTATGTGTTCAATAAA
$OlaC_{DE}19$	7,493	195	TTATTTTATGGGTACATCATCAACCTTTTAACTACATTTCTTTTCAA CAACAACCCCCCTTTGCTGTTGTTCTTTAACCTTTTTTTCCTTTTAA ACCCTTGAAGTTCGACCTAATCAATGTCACCACTGTGCTACAAGC CTTTAACGCTGCAGCCAAACAAGCAAAACAGCTCTGTAAAGCAG TAATAGGGTTAC
$OlaC_{DE}20$	7,092	32	CCTGCTGGCAATGAAGTGTAAAGCCACGCAGC
$OlaC_{DE}21$	3,566	279	AGCTCATACGGCCATGCTTAACTGCCCGACTTACATCCAATTAAA CTTATCTAGGCCAATAGGGGACAGTAATATGATTAGACCGGCTTC CCGCATTCAGCTCAGACTCCAGTTTACTGGGCAGCAGTGTCTCCGT GCCCATTGTCAGTGTTTTGGTGTGTGGGGAGAGTTTGTATCGTTTGG AGCAGGATTTATCACCACAGAAGCAAAACACCATGATTTTCAAA GTTTTCATCTCTGTTTAAAAGACGATGA
$OlaC_{DE}22$	2,029	199	GTTAGAATCGAAATTCCTCCCATTCTCAATGTTTATGGTTTTCTCTG GAAGAAAGGGATTTAGCGTTACTACAGCCATTCGGGCTTTTATGA TGGACACAGAGAGGATTTTAGTGGTCAAATTCAAGTGCTGTGTCT GGACAGCTGTCTGCTTTCTTTGTACTTGGGTGCAGGATGCATCTTG GGAGACAACCTTTTGT
$OlaC_{DE}26_{bcd}$	174	139	CTCCTACACATCCAACTCTTCCAAATTGATATATGACAATATCTAC TTTCGATCACGTGTCTGCGGGGCGACCTAGACGGATTGCGCGTCA TCTCGCCTTCCCAAATTTTTCCCTTCTGCAGCAGCTCGAATCCAAA AC
$OlaC_{EF}27$	6,383	56	CTCACTGAGTTGTAATCGTGGGCATAATCATTTTCTAAAAGAGCCC ATAACACCAA
$OlaC_{EF}28$	5,987	286	TTTTCACCTTTGAAAACTGTGATGTACAACTTGTGAAAGTATATAC CCCAAAATGTACAGAGTATAGCCTATATGCAATGTGCATTTGTAA ATGTAAATAGGCGTATGTTGACCCCCCCTCCCCTCCTACTCCTATC CCCGCTGCTATATTCATCACACCGCCCATGTAACTCCAATAGCTTT AACGCCACCTCTCGTTAAACATTTGCTTCAAAAATTGTTGGTGCTT TGGATTTCAGAGAGGATGGGCAGATC
$OlaC_{EF}29$	5,316	145	CTTTTACTGGCCACTAAATGGAATCGCCACTCTTAGTCTAGACACT GCCATAAAGACACAGGCACTAAATGGCTATTTTTTCCTCTTAATTG CACCGTCTGTTCTGGGGAAACAGCATTCACTTCCGCAGTAAACAG TAAACAAG
$OlaC_{EF}30$	4,922	237	TATAATTAGTCTTTGAACTTCAATAATGTCAAGGCCCCCACCTTTA ACCTGCTGCAATAAAGTGAAATCAGGGAGCCCGGAAATGTGCTG AGAGGCAGAAGGAGCCAATGAGAGTGGGCGGGAGGATCGTGGAA TGTGTACGACACCAACTTTGACCCTTTGCCTTAAACTGCGGCAGGT AACTTTACCAGTTTTAGCACACCCAGTAAATACACTTTATTGTCCT TGTAAAATGTA
$OlaC_{EF}35_{ac}$	322	235	CTTTGTTCTTCCTGATGTGGGCCTGCTGATTTGGACAATGAACGGT CATTAGGTGGCGCTCTTGCTCCACGGAAAAGCCTTCGGTTTTTCCA ACTTGACCGCGCTGACGTCACTTCAGTCTGGAGCTTCAAGGCCAC TTCCACGCCGCTATTGGTCTGAAACTCACATGACATGTCTCCGAA CATCCATAATTATGTTACTGATATTTTTGCACCCCCCTCTAAAAGA TGTCAGC
$OlaC_{FG}37$	6,955	66	ACACACACACGCACGCGCTCACACCTACTGTGCCTTTCTAGATGA ACAAAACATAGCAAACTGCAC
$OlaC_{FG}38_{ab}$	6,521	162	ATAACGTTGATTTAAATATTATCCAGGTGACCACAGTAAGTCAAG GTCATAAAATTCTAATGTCAGGTTCGCCCTGGAAAGCGTTGGGGG TGGATTTTATGATCTGCAAATATAATGTGCTGCAGCAGTAAAGAT GCGTTTAAAGGTGTGTGTGGAGGAGGG
$OlaC_{FG}39$	6,143	21	AGGTAGGACTAATGCTTTTAT
$OlaC_{FG}40_{bcd}$	5,693	162	GTGGTTTAGGTAGTTTCATGTTGTTGGGGTTGGCTTCCTGGCTCGGC AACAAGAAACTGCCTTGATTACGTCAGTTCGTCTTCATCAAGGGC ACAATTTCCGCTGAATTACCGCTTACAGTCAGTCACCCAAGTCCC CGTAAATGGCATTGTAAATGAACTG
$OlaC_{FG}41$	4,650	398	CCTTGTTTTCTCAGCTCCCTCATTATGGCTTTAGTACCGCTGGAAG ACAGTCTAGAATCCAAAGTCACTTTCATTCAGGCCCCTGAATGAG TGTCTCCCTTGGAATGCATCACTCACAAAGACTTTACAGCGCTGG AAACAGGCTGTAAGAGCAGCACAGTTTCACCCTAAAACCATATG CACATAAAGTCATTTATTTCACTTTGTTTGTGGCGCACAGGAACTGA AGCCGCAGGACAGACACTTTCTCGACGCA
$OlaC_{FG}42$	824	112	TATTAACGGCACGTTTTTGAATAATCAAATGATCCTCGTTAAAATT TATTGTTGGGTATAAAAAACATGAATGGAGAGCTGTCGAATATAG TGCGCTGTAGTGGAGCTGCTT
$OlaC_{FG}52_{abc}$	188	168	GCGCATTGATCCACTGTACAATTTTTGTGGGGGTAAATATAATCAC GTGTCAGCGAGGCAGCCAATAGGAGCCCGGGAAGCTCTGAGAAA TAATTACCTGCCTTGATTGTTCTATGGCCAGATAAAAAAGTACAC ATACACTCCATATAATAAACGGATGCATGTAAA
$OlaC_{GH}53$	3,283	13	TTGTTTCATTTGT
$OlaC_{GH}56$	2,467	209	AAATTGCCTTTTTGTGGTGGCAAACAATACAAGCAGGATGCCGTA ATTTCAAAGCCCAATCCCGTGCATTTCCGCTCATCCCGGACAAAA ATGCCAGTTTTACGGCCCCTGTTGGAGCTCGGCTGTTGGTCTCAAA TGCAGACGAACAAAGCAAACTCGACTAACTGGCTAGACGTCTGG GCTAAATTACTTTATGGTTTTAATGGACG
$OlaC_{GH}59$	1,892	155	CTAGATGACGCTGTTGGTCCGTGTTTTGAGCGAAGCCTGACTGTTG GCGGACTTTTTTAACTGCTTTGTGGCTTGGCTGAAGAATGAAAAGT AACCTATCGATGACAAAGTGAAGGATTTGGTCAAGCATCCCATTT TGTGTAATCGGCCACAGA
$OlaC_{GH}60$	1,505	35	ACATTTTACACCTTAACCCCGGTTCTCATTGAAAT
$OlaC_{GH}61$	1,134	224	CTATGTTTTATAAAAATGCGTAAAAAAACGCAGGGAACCCCAGA AAAACCACGCAAGCCACCGCTTTTGCATGTCAGGAGCAAAGTTGT GCTGAGCAGGGAGCTCCGCAGACTGCAGCTAAACATCCAATGTTA CATTTTATGAACTCCCGTGTGGGAGATTTACTGCGTCTCCTCGCCG GTTTTACGGGGTCAGTTAGTGGCACACGGTATTCATAGGTCGTG
$OlaC_{GH}62$	459	451	TTATGATTAGTGAAGCTTTGCTCAAGGTGGTAATACGCAGGGCCC GCGTCTGACCGGTCATGGAAAGATGAAGGGGTGGGGAAGGTGGA TGAGGGTGGGGGCGGTTGGGTGGGAAGCTAAAAAAAAGGGGAAAA GAAAGAAAAAGAGCAGCGTCCGTGACTGTGCCGGTGGGACAATCC AGCTGTAATCGCAGCTTACGTACATGCAAACTCTCACGCCATGAG CTGCTTTTGAGACTGCAAAGCCCGAAGAAGAAC
$OlaC_{GH}63$	9,080	170	TTTTATTGTCCTTTTTATGGTTATATGGCTGAAAAGAGAAAAAGAA GGCTCCCATAAACAGAGCAAGAGTCACAAGGAGGCAGTGAGCGT CAACTTTATGACCACCCTTCCATTTTGTCAAGAGGAATCCCCCCTC CAAAAAAATAAATAAATAAATCCTGCTTCATAAA
$OlaC_{GH}64$	8,469	149	TCAGACATCAGAATTTTTTTCTAGGATTACCAGGGAAACGCAACA GTTTGAAATCCTTTGATGCTACCTAAACCTGGGGAAATTGCAATA GAGATTTCTGAATATAGAGATTTCATATGGTTTAAAACATTTTTTA GTGTTCTATATCA
$OlaC_{HJ}65$	8,272	370	TAAACACTTGTTGAAGACAATTTTTTGGAGTCTTGGGAGGACAAG TAGCTGAAAGCCAAGCGAACAATTGAGAGATTTTACCACGGCACT ACCTAATCTCAGAGCTTTGCCTTTGCGGAGTTTACCTTCCGCACAA GTGCGGCATCGCTCCGACGTGATTGCAAGCTATATTCCCCAACA ACTAAAACTGCCTATGACGCACACTTATTGCGTACCTGCTACAAT TACAACGTCTTTAAAGTGAGGTCAATTTA
$OlaC_{HJ}66$	6,174	174	GTGTGTATATACTTTTACTAGTGAGTTAACATTACCAATAAAGTTT AGTGTACCTGCAGTAAACTTTATTGGGGTGTAAAGACATGGGTCA CATGGCTCTGCGTCTTTTGTCTCCTTGTGGCTCTGCTGTTTGCGGGA AAAACGGCCTGCGAGGATTCCCCGCATTTCTAGTGT
$OlaC_{HJ}67_{abc}$	5,250	131	ACATTCCTGGTTGTTAAAGGAGAGAAATTTACAGCTGAGTAATAA AAGTTTACGACTGAATCCTTCCTACGATTGGCTGCGGCGAGCCAC GTGGCGCACGCTCTGTGAACATGAACTTTATGCTGTTGTCT
$OlaC_{HJ}68$	4,606	206	TGACCTAACCTGTAGCCTGTATATAGATATGCTGATGCTGGGAGC GGTTGTCATGGTTCCTCATTGTTTGGAGCACCTGCGTGCGGCTGGC ATCGTGGGGAAAACATTTCACAAGAGCTGTGCTAATCTGTTGTAG TAAACTGCACTCTAATCGTGAACTTTCTCATCACTCATAGGCCCTT GTGACCAGGCAGTGACGAATTACA
$OlaC_{HJ}69$	3,833	33	TTTATTGGAAATGTAAAACAACGCGACCGTTTT
$OlaC_{HJ}70$	3,146	349	TAAATCTCCTTTTAAGATTAAGAAGGGAAAATCGCGGGCCATTGT TGAGCTCTTTAGGATTTTAAAAAGCAGAGTTAAGCTGACATCAAA CTCCATGAAAGTCGTGAAGCTGTCTGCGTCACCACATTTAGAAAC TCACTTACAAACCCCGACAGGACAAAAATTTTCCCCAAACTAATG GTCATAATTTATGACGTGCGGATCAACATGTGCTCATTGGGCAGC AGCTTCCTGAGTGCTTAACGAGCATGAAGT
$OlaC_{HJ}71$	2,186	236	ATAAAACACTAAAACTCTGGGGAAAGAATGACCTTTCTGATGCTG CACACGGACACAGCCTCTGCTGCATTCCAGTCCACACATTGGCTT CACTGTTCAGCTTTTATTTACAGCCTCCACTTTGAGGACGAGCGGG TTTTCGTGCGTGTGTGTGTTTGAGAGAGAAACGAGCCTGAAAACA AAACACACATATTTGCTTTCAGTTTTCACACTTTTGATTGTATTATT TGTTTTAT
$OlaC_{HJ}72$	939	193	GCTTTCATCTGTGCGTCCGGATGATAGCCTTGAGCTGCTCCGCTGC AGAGATAGCTCTTGGTGAGCTGTGGAACGTTACTTGTTTTATTGCC CTCGCAGAAAAGAGACTATAAACGCATTCTTAACCCCCCCCCCCT GCCCCCAAAAAGCTGTGCTCACTTGACAGATAAATCATACACAA AGGACACATTAA
$OlaC_{HJ}73$	468	466	GTTAACAGCTCCGTGTCTTCCTCTAAAACAAATGCTTCTAAACTTG AAGTCCAAACTTCCCATTGCGTTTAGCTGGACGGTGCGCCTCTCC AAATCCACGTTGTGTTTATCGGCATTTAGCACTAATGTTCAAGCTG TGAGCTAAAAGCTACAGTTTCGCCTCAACTCCTGCGCCTCAATTG GCTGAGAAGGGTCAGCTGACACTGTCATATTGTCTCTCACCGAGC CAAGCCTCGGCTATAACACTCGGGTTTG
$OlaC_{JK}82$	1,607	150	TAACTACCTAAAAAAACTACCTAAAAATGTATGGACTTGAGTGCG ACTTCACTGCATGATGGCTTCCCCGCTGCTCTTTATCTCTTGTTGAC CACATGAGTGTGGAGCCACATATCCAAATAATCTACAGTTTTGTC CTTGTGACATTTG
$OlaC_{JK}83$	410	255	TGACGCTCTTGTCTAACTGTCCTGCCCTATTTGCGCCAAAGAACCC TGCCTGGCCCCTAACAACTAACAGCTCAAAGTTTACCGGACACGT TTCTCCTATTCATCAATATCCCCATTGAGCATCAAGCCAATTTATG ACTGGCCAACACGTGCACGTGATCACATACAAATACTCCATATTT GGGCAGCGCAAGCCGGATCAAGAATGCAAAAAAAAAAAAAAAA GATAACAAAGCTCACTCCACCTATAAATCC

$OlaC_{KL}84$	8,578	217	GCAGGAGTGCCGCCTGACCTTCAGATGAACCGCAAGGAGCAATG ACATCATGACACCGCTCTGTGTGCTTCATCACCAATTATGTGTCAC TTGCTGCAGTTTGGGACCTAATGTATTGAATATCATTCACTTTGAT CAATCGGTTTACATTTCTAAATTAATTTGCTGCCCATGCGATGATA ACCCCACTTTACTGCGTGTAAGGAACAGGAGGTTG
$OlaC_{KL}85$	8,134	215	AATGTGTGTGGACTGACCAGTGTGCCTGTAAGTGGATTCTCCGTCT GTGTCTGTCATTCCTCGTGGAAGTGAATCATTCTGTAAACTGTATT GACTGTATGTTAAAATATTGTAACGTGCAATGAAGTCGTTTCTCTC CTGACGGTGCCTGCTAAACACAAGGCCTCAAAAATCACTGAAAG ATAGCCTGTTTGTTGACTGTCATTAATAAAGTT
$OlaC_{KL}86$	7,734	146	GACAAACGCGCTGAGAAGTTTAAAGCCCCCCCATAAAACTTTATT GCCCCTTTTTCCATTACTCCGCAGGCACACTGCGCTTCCTGTTTTG CCAGGGAAGGAAGGGACCCACAATCCTGTAATAACAACACACAT GGACGCAAAAC
$OlaC_{KL}87_{ac}$	7,183	153	AGGTTACATATGCCAATTGCCCCAAGCAGGGCCTGTGAATGGTGC ATGGGGAGCACGTGGTGTCATTTAAGTGGGTTTTATGGCCTGGAA GAGCTGACAAACCTTCGATATATACACATCATATATAATCTTAAC TGTCCGGAATCGCAGCTG
$OlaC_{KL}92$	5,615	602	ATAGGGGGCACTGTGTGACGCGCTCCATGTGCGGGTCGGCAACGT GTTCTTGTCACAGCCTCTTCTGATTTTCAGACCTGCAAAATGTTTGT TTTTCCCGTTTTTAAATATGTCTTCGTCATTTTAGATGCAAATATTT GTCTATGGTCGTTAGACTGCTTTTCCTGCAGGATAGGAAGAAAAA TATTGCTTTGCGTAAACAATAAATACCTGCGTGGGTTTAAAAATTT CCTCTACTGTTATCAGATTCATAAA
$OlaC_{KL}94$	4,195	646	CCACAGCTCTGCACTAAAACACAGTCAGTGGTTTATGTGGCGAAT GCAGTTTTTAAAACCTCTGCTCCCCTAACTGCAGCTTCCAGAGAG CATTCAAACCCTACGAGTCTCCAGTCTGTCTGCACACTTCATTAAA AACTTTTTTTATTTTTTATTTTAACTGGACCATCCAGCTTTATTACA CCTGGGGAGGGTAAACGCGAAAAGGCTATTAGGTTAAAGGGTTTG ATTAAAAGCGCATCGTTACCATTTTTA
$OlaC_{KL}95$	3,256	311	TGTGCATCTCTACTCTTATCAATTCCATCGATTCACAATCTGAGCA TGTTTGCTGGCTTTAATAACTTCCAAACAGGTTGCATGTCCGATAT TACATTTTTCATCCGAGGCTTTTCTCATTAGCTTCCAGAGTGGCAG GAAAAAAGGTTACGGCGCGGTCAGTATGGTAATGGCAGTCAGGA GAAATGCTTGGGAACAAAAGAAACTTATTTTTCATCTGCGCTGTC ACGCATCCCTTTTGTATCCGCCGTGTTT
$OlaC_{KL}96$	2,857	51	TATATTTTAAAAACTATCAGTTTGGGTTTAAACGCGAATCTTGTGA TTTTT
$OlaC_{KL}97$	2,769	551	TTAGTTGCCTATATGTACCCTGTAGAACCGAATTTGTGTGGAGTAC AAGCATTCGCAAATACGTCTCTACAGGAATACATGGGGAACTGA AATACCACGCCAACAAAGGATGGCTCTGATTGTTCTTCATCAGCC CCTCCTCACTCGCCTCTAGCTGTTTTCCTTCGTCCCCTCCTTCCTCT GCATACAAGCATCAGTAACGATGCGTGCAGCTCACACCCAGGCA CCAGCAGTGCTATCTCTTTCCCCTGCAAA
$OlaC_{KL}99_{ab}$	3,711	162	ACATTATAACTAGTTATTGAACTAGGTGCACGATCTGAAAGCCAT TTGTGGGGAAAAGAATTCATTGCTGTCTCTCCATCAATAATCCTTG GCAGTGAACTATTGGAACCAGTCAAACGCGAGGGGTGAAACGCG GGTCAGGCTGTCTAACTAATATTAAAA
$OlaC_{KL}102$	151	151	GGTTCCTTATCCGGGAACTACCTCTTTCTCTGCTATTGGGCCATTG GGTCACGTGGTTAAAGTAACTTTACAGGGCTGCTCGCAAGTAGGA GGGCTTTATGAAGCAGAAAAAACGACAAAGCTAGAAAAATTATTTT CCACTCCAGAAATTA
$OlaC_{LM}103$	7,707	301	TTATTTATAGGACAATTCTACATTTTGTTTTCGTAGCTATCTTGTGC GGTTTCCTTTAAGTCCAAAGTTATATAAAATGCATGTTATATGGCA TGGATTACTGCGAATACCGATGAATTATTTTTACGTGACACAGGGT TTAATTTATTTGAAGTTCAATTATGATGTTTTTTTTATTAAAAACGT TTTGAATGAAGGTGGAAAAACAGTAAAAAAAAAAAAAAAGACATT TTATCAAATTGTTATTGTGGGCCAT
$OlaC_{LM}104$	7,229	133	GTTGTGTTTAGACCGTCAAAAGCGGATTATTGTGAAATGCAATAA ACGGACTTTAGTGTATTTGTGCTAATCATATTGGTCAATAAAACAG TGAGTGTCTACTAGTTTTAAACACGTTTTGTCGGAGTTTATT
$OlaC_{LM}105$	5,763	398	TTATGTTGCTGGACAGATTGGAAAAATTAATGTGCTCGCCACAGT CGCAGCATCCCTTTAGCCCCCTTTCTAGGGACTGTCGCCTCCATGT CGCGCAGCTGAAAAGGTAGTTCATTGAAGTTTGGGGGCGAAATGG GAGAAGGCCGAGTTCACCTCGAGGACACATTTTCTGTCCCATTAG GTTCATTTCAGTCTGGATGGCCGACCTCTCCAACCCTGTGACCTGA CTCACTCTCTATTCGTGTACCTTCTGTG

Tetraodon nigroviridis spotted green pufferfish Hox Ca

PFC	Dist to 3' gene	Length	Sequence
$TniC_{CD}1$	6,217	177	TCCATCAACTCCATAGACGCTGACACGCCGCCTATCGTTCCCAGTT CAGAGATGATGCTCTAAAACTGTGAAAAACAATAAAAAAAATAC CCTGTCACATTTGTTTTATATAAATGTACATATATAAATATATATAA TATTTAAAACTGTCCGTATTTCATTGACAAGCATTGCGTG
$TniC_{CD}2$	4,204	167	TACCCTTGACCATGACTATGCTGTCGCTTTGACCTTGACATCACAG CAGATTGTGAATAATAAATAGGGGGCTGGACCCCAACACCGACGG TTTGTTTGCAGGCAAACAAGGTGCTATTTCCCTGTCTAGACAGTTA GGCTTAATGGCACGCCATAAATACATAAAA
$TniC_{CD}3$	3,035	166	GGCTGATAACCTCTGCTGGTGACCAGAGGAAAATATCAGAAGGGC TACCAGTTGGTTATTGTAAGGGATTGGCTTTTGTTTCATTCATTTAA ATCCTGATTTCTAATCTTATTTCATTAATTTATTTTTCTGAATGAATT TTTTTAAATGATTTCTCTGGTCAGAC
$TniC_{CD}5$	2,051	96	CTTTGTTTGGCCGATTGCCCTGTCTGGACTGGACCGCCTGAATCCA GCGGGGTCATTTCAATATGTCTCAGACGTCGGACTCGGCAGTGACC TTGA
$TniC_{CD}6$	1,593	191	GCTTCTCCAGTTTAAGCATTTGATATTTGTTCTGTCTATTTAAAGCT GAAGGGAACCCCTGTGTCTTTGGGTTAAGTAGGATGGTCGCTTTAA GTGACAGATACGCCTCTCAAATATCGACTGGGTAGCAAATCTTTAC CAATGTTTACATAAATGTTAATGTTCTCACATAACACACGACCAGC ACATTT
$TniC_{CD}7$	1,163	41	AATCCATCTTTTCTTCTACAGGGATGACTACACAAATATTT
$TniC_{CD}8_{abcde}$	681	681	CCTCTGATATCTTGTAAAAGCTCATATAAAAGAACTGTAAGAAGT GTAAAAGTCCCACCAGAGTTTGGCACAAAATGTTTATGGGAGATA TCTGTTATAGTTGTTAATGTAAGGTGAACTGAAGATTATACGTGAA ATAGGCTCTCTTAAAGTCACCTTGTAGAATATCTCGGGCTTTTACA GACCTGGCGGTTTCAACCTGCTGACCCTGATATTGTAACTATGAGA GGAGCTCGGCAATGAGCATCATTCCTT
$TniC_{DE}13$	10,463	14	AATGCATGCACTCA
$TniC_{DE}14$	9,924	19	GTTCACGGTCACGGGCAGG
$TniC_{DE}15$	9,707	16	AAAACGGGTATGAGCT
$TniC_{DE}16$	9,450	106	AGCATGTTACTAATGTGAAATACAGTAGCAACATATGTTTTGCTGG CCACTTTCCTGTAAGTTTTTTTTTTTTTAACTTATGCGTAAAAAGAC GAATTCAAAAAGC
$TniC_{DE}17$	9,206	119	GTATAACCTCGTCAAATGAACTTCTGTATGCTACTATGTGCATCCC ATTTGAGGATATGTTATCACTATGTAATAGTGTTTTATCTGTTCACC GTCTTGTAAGTCATTATCATTCCATA
$TniC_{DE}18$	8,947	78	TTTGCGCTTGTAAGGATTGCTCTGTTGTCTTGTAAAAGTAAAAAAA AATGAATGTATATGCGTAAGGTCTTCAATAAA
$TniC_{DE}19$	8,563	188	TTATTTATAACACTGTGTAACCTGTAGGGGCAGTTAAATAGTATTTT TATAATGGAACAACACTGTCGCAGTTTCTTTTTCTTGAGTTTTGTGG GTTCTCTTGAAGTTCAACCTAATCATTGTCACCAGTTTGTTACAAGT CTATAGCGTGGAAGCTAAACAGCTCTGTAAAGTATTAATAGGGTTA C
$TniC_{DE}20$	8,170	31	CCTGCTAGAGAGGAAGTGTAAAGCACGCAGC
$TniC_{DE}21$	4,661	281	AGCTCATTCAGTCAAGTTTAACTGGCCGACATCCATCCAATTAAAC TTATCTGGGCCACGAGGGGCCAGTAACGTGATTACAGGCGCCCCA CATTTAGCTTTGATTCCAGCAAATCGCGAAGCCTGATCCTGTTGAA CGGCGTCCATTGTGTGTGTCATGAGTGTTGTATAATTTTGAGCAGAT TTATTTTCCAGAGACAAAAGCACTAAGCAATAGTTAGAAGAGTCG GTTAGAGTTCATCTTAGCACGAGGAT
$TniC_{DE}26_{bcd}$	169	139	CTCCTACACATGCAACTCTTCCAAATTGATATATGACAATATCTAC TTTCGATCACGTGTTTGCGTGGCGACTTAGACGGATTGCGCGTCAT CTCGCCGTCCCAAATTTTTCCCTTCTGCAGCAGCTCGAATCCAAAA C
$TniC_{EF}27$	6,343	50	CTCACTCAAGTCTAATCCTGGGCGTTTTCCAAAGGAGCCCACATCA CCAA
$TniC_{EF}28$	6,030	265	TTTTCACCACTACAAAATGTGACGTCCAATCTGTGAAAGTACTCTC CCCCCTGTACAGAGCAGCGTATGCAGTGTGCGTTCATGAGTGTAAA TAGGCATATGCTGAATTTTCTTTTCCCTTTTTCCTTGCTGCTATTTTC ACTACACTACTTCAACAGCTTTAACGCCAACATTCGTTAAATTATT TGCTTAGACCATTCGGATTCCAGAGAGGAAGAACAGATTGTTGGC CTGTGTATCATGAATATAACAAAT
$TniC_{EF}29$	5,390	146	CTTTTAGTGCGCACTAAATGGAATCGCCACTCTTGTCGTCTAGACA CTGCCATAAAGGGTTAGGCACTGAATGGCTATTTTTTTCCCTTAATT GCACCGTCTGTTCGGGGGAATGGCCTTCACTTCCGCAGTAAATAGT AAACAAG
$TniC_{EF}30$	4,654	255	TATAATCAAAAGTAGTCTTTGAACTTCATTAATGTCAAGGTCTTCA CCTTTAACCCGCTACAATAAAGCTAAATCAAAGAGCCTCTTAAGT CCGGAAATGTGCGTGGCGACAAAACGAGCCAATGAGCTCGAGTGG GGAACAGGACGTGGAATGTGTACAGAGAGCAACTTTGACCCCTCG CATTAAACTACAGCAGGTAAGTTTGCAGCTTTAGCACGGCCAGTA AATACACTTTATTGTCCCTGTAAAATGTA
$TniC_{EF}31$	4,002	66	CAAGGCTCCGGGAGCCTTGACACACCCACCAACCTTGGCCCTGGG ACGGCGGTGAAGGGTAATTTG
$TniC_{EF}35_{ac}$	324	240	CTTTGTTTCCTCTGCTGCAGGCCCGCTGTCCTGACCCACAGACGGT CGCAAGGTGGCGCTCTTGCTCCGTCTAAAAGGCAGAGGACTTTCTA ACTTGACCGCGCTGACGTCATGTCGGTCTGGAGCGCCAAGGCCAC TTCCACGATGCCATTGGCCTGGAAAAATCACATGACACGTCTCCC AGCATCCATAATTATGTTACTTATATTATCTGTGCGCCCCCTTCCAA AAGATGTCAGC
$TniC_{FG}37$	6,183	65	ACACACGTACACACGTACACACACCTCGTGCCTTCCTGTATAAAC TCTAAAGCGTTTCACTGCAC
$TniC_{FG}38_{ab}$	5,772	162	ATAACGTTGATTTAAATATTATCCAGGTGACCACAGTAAGTCAAGG TCATAAAATCCTAATGTCAGGTTCGCCCTGGAAAGCGTTGGGGGTG GATTTTATGATCTGCAAATATAATGTGCTGCAGCAGTAAAGATGCG TTTAAAGGTGTGTGTGGAGGAGGG
$TniC_{FG}39$	5,381	20	AGGTAGGACTGAAATTTTAT
$TniC_{FG}40_{bcd}$	4,941	157	GTGGTTTAGGTAGTTTCATGTTGTTGGGGATGGCTTCCTGGCTCGGC AACAAGAAACTGCCTTGATTACGTCAGTTCGTCTTCATCAAGGGCA CAATTTCCGCTGAATTACCCGATAGCCACCCGAGGTTGTCGCAAAT GGCATTGTAAATGAACTG
$TniC_{FG}41$	4,035	343	CCTTGTCTCCCCCACCCCAGCCACACAACTGGCAGAGACAGTCTG GAATCCAAAGTCACTTTCATTCAGCTGGGTGAAAGAACGTTTTCCG TGGAATGCACCGGTGACAAAGACTTTACAGCAGTGGAAACAGCCC CGAGAACAGCCCGAATTCACCCGAAAACCCTATACACATAAAAC TATGTCCAGGAACTGGATCTTCATCGTGGGGGAGCTGCGTTAGTGC CACTGACAGGATGAGACACGAACCAAAAT
$TniC_{FG}42$	842	111	TATTAACGGCACATTTATGAACAATCAAATGGCCCTCGTTAAAATT TATTGACGGGCATAAAATCACGAAGGCCCACAAGCTGAAAAAAA GCTCCATGATTTGGGCTGCTT
$TniC_{FG}52_{abc}$	192	172	GCGCATTGATCCACTGTACAGTTTTTGCAGGAGTAAATATAATCAC GTGTCACAGAGGCAGCCAATAGCAGCCGCGAAAGCTCCTAGAAAT AATTACCTGCCCTGATTGTTCCATGGCTAGATAAAAAAAAAAGTA CACATACAGCCCATATAATAATCCGATGCATGTAAA
$TniC_{FG}44$	843	64	TTATTAACGGCACATTTATGAACAATCAAATGGCCCTCGTTAAAAT TTATTGACGGGCATAAAA
$TniC_{FG}45$	433	29	GACAGAGATGGATTACACCATCAAATAGT
$TniC_{FG}50_{ab}$	829	18	TTTATGAACAATCAAATG
$TniC_{GH}53$	3,322	13	TTGTTTTATTTGT
$TniC_{GH}56$	2,480	211	AAATTGCCTTTTTGTGGTGCCAAACAATACGGACAGGATGCTGTGA TTTAGAAGCTAAATCCAGCGCGTTTCCGCTCTGCTGGGACAAAAAT GCCAGTTTTACAGCCCCTCTTGGAGTTCGGCTGTTTGTCTCAAATGC AATGACGAACAAGCAAACTAGACTAACTGGCTAGACGTCTGGGC TAAATGACTTTATGGTTTTAATGGACG
$TniC_{GH}59$	1,834	149	CTAGATGACGCTGTTGCTCCACGTTTTGAGGCGAGTTACCCAAGGG TGCTGCTGCAGCTCTGTGGCTTGGCTGAAAAATGAAGAACAACAT ATCGATGACAAAATGACAGATATGGACACCGACCACATTTTGTGT AACAGGCCACAGA
$TniC_{GH}60$	1,485	27	ACATTTTTTTAATTTGTAAATTGAAAT
$TniC_{GH}61$	1,138	224	CTATGTGTGTAAAACGCTAGCAGCACAAAAAAAAAAAACCCAAAA AAAACCTTACTGTGTATTTTCTGCAGCCTTGGAGCTGCGTGAGTGT ATAGTGCGCGATGTCATTGCAGCAAACCATCAGATGTTAAATTTTA CGAACTCCTGCGTCTCTGAGATTTACTGCGTTTCCACATCGGTTTTA CGGGGGTCACTATAAATGGGAACCCTGTATTCACACGTCGTG
$TniC_{GH}62$	471	463	TTATGAATAATACAGCTCCTCTCGAGGTGCACATGCGTAGGCCCCA TACGTCTGAGCGCTCTCTGAAAGATGAGGGGTATGGGTAGGGGGT GGGGGCGCCCCTGGCGGGGGCGGTTGGGTGGGAAGCTTGAAAAAA AAGGAAAAAGAAAGAAAAAGAGCAGCGTCTGTGACTGTGCCGGTC GAGCAAGCCCAGCTAACGCGCAAAAAAAAAAAGCACATGCCATG AGCCGGGTTTGAGCCTGCACAAAGCTCGGGG
$TniC_{GH}63$	9,638	167	TTTTATGGTCCTTTTTATGGCTATATGGCTGAAAGGGAGGAAAGAA AGAGTGGCCCCATAAACAGATGAAGAGTCACAGGGAGTCAGAGC GCGTTAACTTTATGACCATCCTTTCATTTTTTTTGTCAAGAGGGAAC CCAAAATCCCTCCAATTTTGCTGTCATAAA
$TniC_{GH}64$	9,027	146	TCAGACCCTGTTTTTTTTTCTATAGGATGACCAGGGAAATGCAATC GTTGAAATCCTTTGATGCTACCTAAACCTGGGGAAATTGCAATAGA GATTCCTGAGCGAGGAGATTTCAGATGTTTTAAATTTTAAAGTGTC CTATATCA
$TniC_{HJ}66$	6,305	174	GTGTGTCTGTATTTTTACCACACGGTTGCTATTACCAATAAAGTTGA AAGTACCTGCTGTAAACTTTATTGGCGGTGTTGTATATAACCAGCA TTGGTCCCATCCGCATCTTTTGTTCTGGGATCCCAGCCTGCGCAGG CAAAGCACCTCTCCCCCAGCCCTAAAAACTAGTGT
$TniC_{HJ}67_{abc}$	5,360	129	ACATTCCTGGTTGTTAAAGGAGAAATTTACAGCTGAGTAATAAAA GTTTACGACTAGATCCTCTCCGCCATTGGCTGCAGCTGACCACGTG GCGCACGTTCTATGAACATGAACTTTATGCTGTTGTCT
$TniC_{HJ}68$	4,704	215	TGACCTGACCTACAGCCCATAGCCACATTTGTTGCGCAGCCATTGT TTCTGCACTTTTTGGATGGTGTGGCAGCACCTGTGTGCTCCACTGCA GCTCCAGACAGCATTAGCCCAGCATTTCGGAGCCGCTCTGCCAAT CGGTTGCAATAAATTTACTACTGATAATAAAGTTTGTCATCAGTCA CAGGCCTCTGACCAGTCAGTGACAAATTACA
$TniC_{HJ}69$	3,956	34	TTTATTTGAATTGTAAGCAAAAAAAAAACGTTTT
$TniC_{HJ}70$	3,335	402	TAAATCCTTGTTTTAAATGCGAGGAGAAAACCGCGGTCTATTGTTT TGCCCTGAAGTGTTTTAAAAGGCAGATAGCTCTGCATTAAACTCCA TAAAGTCATGAATCTGTCCGCATCACCACATTTAGAACCTCTCTTA CAAATCCCGTCAGGACCAAAAAAATTTTGCCCACTAATGGTTCTA ATTTATGACGTGCAGATCAACATGTGCTCATTGAGCAACAGCTTCC TGAGTGCTTAACGAGCATAAAGTCGG
$TniC_{HJ}71$	2,393	262	ATAAAACACTAAAACTCAAGGGGAAACAATGACCTTTCTAATTCT GCACGCTGGCTTAGCGTCGGCTGCATTCCAGCCCTCTCCTGTCTGCTT TCTGGCTTTAGTGTTCAGCCTTTATTTACAGCCTGAACTTTTGCAGC CTTTTCGTTTTCAGAGGAGGCGACAAATAGTTTAGACAAATCTCTG CCTCAAATATTATTATTCTTTAACTTGCTGTTCAAGACTGTTTTTAG GCTTTTGAATTATGTGGGGTTTT
$TniC_{HJ}72$	975	210	GCTTTCATTTCTGAAATGTGCGAGATTTTTTTTTTTTTTTTTTTTTGG CTACAGAAATATTTGCTGGTAAACTGCGGAACTTGTTTTATTGCCC TCGCAGAAGAGACAACATAAAGGCATTCTTACTACCTACACCCTC CCCCCAGCCCACTTCTAAAACACAGATATTGCTCAAGGGCCAGCC AATTTATACTCAGTCGAAACATTAA
$TniC_{HJ}73$	473	471	GTTAACATTCTACATCGCATTCCTCCGCGAATACCACATTTTTTAAT ACTTTAGTGGAAAAAACCCCGTCTCGTCCAAGCGAGCAGTGCGCC ACTCCAACTCTGCGTGGTGTTTATCGGCATTTAGCTGTAATGTTCAA GCTCCAAGCTAAAAGCTACAGTTCCCTCTCCACTCCTGCCTCTCAA TTGGTGGAGAGTAGTCAGCTGACATTGTCATATTGTCTCTCACCGA GCCAAGCCTCAGCTATAACGCTCG
$TniC_{JK}82$	1,693	141	TAACTACCTAAAAACGAATTACAGACTTGAAAGCGACTCACTGCA TGATGGCTACCCCGCAACTTTATCACTTATATTCCACGTTAACCAC AGGAACGTTCCAAAATATTTCAGGAGCCACTATGGCCACCGTGAC ATTTG
$TniC_{JK}83$	432	237	TGACGCTCTTGTCTTACTGTCCTGCCCTTTTTTGCGCTAAAGCTTCTT GCCTGGTCCTCACCAAAGTTTACCGGACACGTTTCTCCTATTCATC AATATCCCCATTCAGTATCAACGCCAATTTATGACTGGCCAACATG CGCACGTGATCTCAAACAAATACCCCATATTTGGGCAGCGCCCAT CGGATCAAGAATTTTTAAAAAAGACAGCAAAGCCTACCCCACCTA TAAATCC

$TniC_{KL}84$	7,689	217	GCAGGAGTGCTGCCTGACCTCCGGATGAACCGCCGGGCAGCAATG ACATCATGACACCGCTCTGTGCGCTTCATCACCAATTATGTGTCAC TCTCTGCAGTTTGGAACCTAATGTATTGAATATCACTCACTTTGATC TATTGGTTTACATTTCTAAATGAATTTGCCGCTGTGCGATGATATCC TTACTTTGCGACGTGCAGAGGACACGAGGTTG
$TniC_{KL}85$	7,250	154	AATGTGTTAAATGTTTAGTGAAGTAATACTTTTATGTTTGTAGTTGT CATTTTCTGGTAAATCCTCTGACTGTATATGGTAATATTGTAATGCG CAAATAACTCATTTCCACCTTCCAGAAATGACTTGATTGTTGATTT GAAAGAATAAAGTT
$TniC_{KL}86$	7,031	126	GACAAACGCTTTCTGGAGTTTAAGACCCTCATAAAACTTTATTGCC CCGTTTCCACAGCGCGCTAGTTTCCTGTTTGGTCTGGGAAGGAAGC GACCCAACCTGTGACAACGCTCACCCACCAAAAC
$TniC_{KL}87_{ac}$	6,559	153	AGGTTACATATGCCAATTGCCCTAAGCAGGGCCTGTGAATGGTGC ATAGGAAGCACGTGGTGTCATTTAAGTGGGTTTTATGGCCTGGAAG AGCTGACAAACCTTCGATATATACACATCATATATAATCTTAACTG TCCGGAATCGCAGCTG
$TniC_{KL}92$	5,286	510	ATAGGGGGCGGTGTGTGACACACTGTTGGTTGGGGAAAACTAAAG CAAAATGTGTTTATTTTATTTTAAGTATTGATTTTGTAGGCTTTCTTT TACATATATATCATTTTATTCACTGTCATTACTGCGTAAATTGCTGT GAAGACCTCCTGAAAGAATGCAATGGGTTTGGTTTTACTGTAAAAG TAAATAAATTCAGATGACAATTCAGAATAGGGCCATTACATTTTCA GAACGATAATAATTTTAAGAAAA
$TniC_{KL}93$	2,948	16	AAAGAAACTTATTTTT
$TniC_{KL}94$	4,022	663	CCACAGCATCGTTAAGACCAACTTGGTCTTTGCATGACAGCAAGTT TGCTAAAACCGAGTCCTCTGTCTCCCTGCTGCTTTAAACTGTCCAT GCCCAATTAGTCTCCAGTCTGTCAGCATTTCATTAATACCATTTTTT TTTCTAGCGGACGTTATCCACCTTAACTTTATTACACCTGGGGAGA TTTACCAGGAAACGGCTATTAAGTTAACGCTTTGATTAAACGTGCG CCGAAGCAGTTTCCATGGAGCAGA
$TniC_{KL}95$	3,149	312	TGTGCACATCTGCAATGTAACCAATTTCAATGAATCACAAATCCA AGCGTGTTTGGTGTTTTTGATAACCTCGGAACAGGTCACATGTTTG ATACTGCATTTTTCATTCTGGGCTTTTCTCATTAGCTTCTCTAGTGGC AGGAAGAAAGGTTACGGGGCGGTCAGTATGGTAATAGCAGACAG AAGAAATGGCTGGGAACAAAAGAAACTTATTTTTCATCAGTGCTG TCACGCCATCCCTTTGTACCCTGTGTG
$TniC_{KL}97$	2,679	492	TTAGTTGTCTATATGTACCCTGTAGAACCGAATTTGTGTGAGTTCAG ACAGTCACAAGTACGTCTCTACAGGAATACATGGGCAACTGAAAA TCCACAGCGGCCACTGGACACTCCTGCTTCTCCTCCTCTGCACACC TTTCCTCCTTTTTCACTCCAAGCATCTGTGTTTTACACTCAGTCTGC ATCCAGTCGCTGTAGCATAATCTTGTCACAAAGTGTATTCATGGTC CAGACAGTCGGGGCTTGTGTGTAG
$TniC_{KL}98$	1,114	31	ATTAAAACCCTGTTAAGATCCCATTTCATTC
$TniC_{KL}99_{ab}$	3,522	163	ACATTATAACTAGTTATTGAACTAGGTGCGCGATCTGAAAGCCATT TGTGTGGATAAGGAATTCATTGCTGTCCCTCCATCAATAACCCTTG GCAGTGAACTATTGGAACCAGTCAAACGCGAGGGGTGAAACGCG GGTCAGCCTGTCTAACTAATATTAAAA
$TniC_{KL}102$	151	151	GGTTCCTTATCCGGGAACTACCTCTAAGCCCACTATTGGCCCATTG TGTCACGTGGTAAAAGTAACTTTACAGGGCTGCTCGCAAGTAGGA GGGCTTTATGGAGCAGAAAAACGACAAAGCTAGAAAAATTATTTT CCACTCCAGAAATTA

Danio rerio zebrafish Hox Cb

PFC	Dist to 3' gene	Length	Sequence
$DreCB_{DE}23$	1,295	21	AGAAAGCAGTTTTAAAAATAA
$DreCB_{DE}26_{b}$	168	59	CTCCTACGTTTGCAACCGTTCCAAATTGATATATGAGAATATCTACT TTCGATCACGTG
$DreCB_{EJ}32$	14,468	20	GGAAAATTAAGCCTTTTTAT
$DreCB_{EJ}47$	12,556	17	TCTCTGGAGCTCATTTC
$DreCB_{EJ}54$	12,259	43	TGTTATGATTATGTACACCATGGGGATATGTGGTGACGTCACT
$DreCB_{FG}40_{abc}$	11,859	72	CAGCTGATGCGTGGTTTAGGTAGTTTGATGTTGTTGGGGTTGACTTC CTGGCTCGACAACAAGAAACTGCCT
$DreCB_{EJ}76$	7,169	32	TATTAAATCTACAGTCCATTGAGGGCATTGAA
$DreCB_{EJ}67_{b}$	3,950	30	GAAAATTTACAGCTATGTAATAAAAGTTTA

PFC	Hornshark D	Frog D	Coelacanth D	Human D	Bichir D	Zebrafish D	Medaka Da	Tetraodon Da	Medaka Db	Tetradodon Db
1	$HfrD_{AC}1_{ab}$	$XtrD_{AC}1_{ab}$	$LmeD_{AD}1_{ab}$	$HsaD_{AC}1_{ab}$		$DreD_{AC}1_{b}$	$OlaDA_{AD}1_{b}$	$TniDA_{AD}1_{b}$		
2						$DreD_{CD}2$	$OlaDA_{AD}2$			
3				$HsaD_{AC}3$			$OlaDA_{AD}3$			
4						$DreD_{CD}4$		$TniDA_{AD}4$		
5		$XtrD_{CE}5$		$HsaD_{CD}5$						
6		$XtrD_{CE}6$		$HsaD_{CD}6$						
7						$DreD_{DE}7$		$TniDA_{DE}7$		
8				$HsaD_{DE}8$			$OlaDA_{DE}8$			
9	$HfrD_{DE}9_{abcd}$		$LmeD_{DE}9_{bcd}$			$DreD_{DE}9_{abcde}$	$OlaDA_{DE}9_{abce}$	$TniDA_{DE}9_{abe}$		
10	$HfrD_{DE}10_{bcde}$	$XtrD_{CE}10_{abcde}$	$LmeD_{DE}10_{bcde}$	$HsaD_{DE}10_{abcde}$		$DreD_{DE}10_{cd}$	$OlaDA_{DE}10_{d}$	$TniDA_{DE}10_{cd}$		
11	$HfrD_{EF}11_{ab}$	$XtrD_{EF}11_{ab}$	$LmeD_{EF}11_{ab}$	$HsaD_{EF}11_{ab}$		$DreD_{EF}11_{b}$	$OlaDA_{EF}11_{ab}$	$TniDA_{EF}11_{ab}$		
12	$HfrD_{EF}12$			$HsaD_{EF}12$						
13				$HsaD_{EF}13$		$DreD_{EF}13$				
14		$XtrD_{EF}14$		$HsaD_{EF}14$						
15		$XtrD_{EF}15$		$HsaD_{EF}15$						
16				$HsaD_{EF}16$		$DreD_{EF}16$				
17				$HsaD_{EF}17$			$OlaDA_{EF}17$			
18		$XtrD_{EF}18$		$HsaD_{EF}18$						
19	$HfrD_{EF}19_{ac}$	$XtrD_{EF}19_{bc}$	$LmeD_{EF}19_{abc}$	$HsaD_{EF}19_{abc}$		$DreD_{EF}19_{abc}$	$OlaDA_{EF}19_{ac}$	$TniDA_{EF}19_{ac}$		
20				$HsaD_{FG}20$			$OlaDA_{FG}20$			
21	$HfrD_{FG}21_{abc}$	$XtrD_{FG}21_{b}$	$LmeD_{FG}21_{ab}$	$HsaD_{FG}21_{ab}$		$DreD_{FG}21_{abc}$		$TniDA_{FG}21_{abc}$		
22						$DreD_{FG}22$	$OlaDA_{FG}22$	$TniDA_{FG}22$		
23	$HfrD_{FG}23_{abce}$	$XtrD_{FG}23_{abce}$	$LmeD_{FG}23_{abce}$	$HsaD_{FG}23_{abe}$		$DreD_{FG}23_{bcd}$	$OlaDA_{FG}23_{bcd}$	$TniDA_{FG}23_{bcd}$		
24				$HsaD_{GH}24$						$TniDB_{GL}24$
25			$LmeD_{GH}25$	$HsaD_{GH}25$						
26				$HsaD_{GH}26$						$TniDB_{GL}26$
27				$HsaD_{GH}27$			$OlaDA_{GL}27$			
28			$LmeD_{GH}28$	$HsaD_{GH}28$						
29						$DreD_{GL}29$	$OlaDA_{GL}29$			
30				$HsaD_{HL}30$		$DreD_{GL}30$				
31				$HsaD_{HL}31$		$DreD_{GL}31$				
32							$OlaDA_{GL}32$	$TniDA_{GL}32$		
33		$XtrD_{HL}33$		$HsaD_{HL}33$						
34	$HfrD_{HK}34$			$HsaD_{HL}34$						
35	$HfrD_{HK}35_{abc}$	$XtrD_{HL}35_{abc}$	$LmeD_{HL}35_{ab}$	$HsaD_{HL}35_{abcd}$		$DreD_{GL}35_{bcd}$	$OlaDA_{GL}35_{bcd}$	$TniDA_{GL}35_{bc}$		
36				$HsaD_{HL}36$		$DreD_{GL}36$				
37	$HfrD_{KL}37_{bc}$	$XtrD_{HL}37_{bc}$	$LmeD_{HL}37_{b}$	$HsaD_{HL}37_{bc}$		$DreD_{GL}37_{abcd}$	$OlaDA_{GL}37_{abcd}$	$TniDA_{GL}37_{abcd}$		
38						$DreD_{GL}38$	$OlaDA_{GL}38$	$TniDA_{GL}38$		
39				$HsaD_{HL}39$						$TniDB_{GL}39$
40		$XtrD_{HL}40_{a}$		$HsaD_{HL}40_{ab}$		$DreD_{GL}40_{ab}$	$OlaDA_{GL}40_{ab}$	$TniDA_{GL}40_{ab}$		
41				$HsaD_{HL}41$						$TniDB_{GL}41$
42				$HsaD_{HL}42$					$OlaDB_{GL}42$	
43									$OlaDB_{GL}43$	$TniDB_{GL}43$
44									$OlaDB_{GL}44$	$TniDB_{GL}44$
45									$OlaDB_{GL}45$	$TniDB_{GL}45$
46		$XtrD_{HL}46_{abcd}$	$LmeD_{HL}46_{abcde}$	$HsaD_{HL}46_{abcde}$		$DreD_{GL}46_{bcd}$	$OlaDA_{GL}46_{cde}$	$TniDA_{GL}46_{b}$		
47		$XtrD_{HL}47_{abc}$	$LmeD_{HL}47_{abc}$	$HsaD_{HL}47_{abc}$		$DreD_{GL}47_{b}$				
48		$XtrD_{HL}48_{abcde}$	$LmeD_{HL}48_{abcde}$	$HsaD_{HL}48_{abcde}$		$DreD_{GL}48_{abcde}$	$OlaDA_{GL}48_{abcde}$	$TniDA_{GL}48_{abcd}$	$OlaDB_{GL}48_{bc}$	$TniDB_{GL}48_{ab}$
49		$XtrD_{HL}49_{abc}$	$LmeD_{HL}49_{abc}$	$HsaD_{HL}49_{abc}$		$DreD_{GL}49_{b}$	$OlaDA_{GL}49_{bc}$	$TniDA_{GL}49_{b}$		
50					$PseD_{LM}50$	$DreD_{LM}50$				
51					$PseD_{LM}51$	$DreD_{LM}51$				
52							$OlaDA_{LM}52$	$TniDA_{LM}52$		
53							$OlaDA_{LM}53$	$TniDA_{LM}53$		
54					$PseD_{LM}54$		$OlaDA_{LM}54$			
55							$OlaDA_{LM}55$	$TniDA_{LM}55$		
56		$XtrD_{LM}56_{abcde}$	$LmeD_{LM}56_{abcde}$	$HsaD_{LM}56_{abcde}$	$PseD_{LM}56_{bcd}$	$DreD_{LM}56_{c}$				
57		$XtrD_{LM}57$		$HsaD_{LM}57$						
58		$XtrD_{LM}58_{bc}$	$LmeD_{LM}58_{bcd}$	$HsaD_{LM}58_{bd}$	$PseD_{LM}58_{bcd}$	$DreD_{LM}58_{abcd}$	$OlaDA_{LM}58_{abcd}$	$TniDA_{LM}58_{abcd}$		
59						$DreD_{LM}59$	$OlaDA_{LM}59$			
60						$DreD_{LM}60$	$OlaDA_{LM}60$	$TniDA_{LM}60$		
61		$XtrD_{LM}61_{ab}$	$LmeD_{LM}61_{ab}$	$HsaD_{LM}61_{abc}$	$PseD_{LM}61_{ab}$	$DreD_{LM}61_{abcd}$		$TniDA_{LM}61_{acd}$		
62		$XtrD_{LM}62_{a}$	$LmeD_{LM}62_{ab}$	$HsaD_{LM}62_{ab}$	$PseD_{LM}62_{ab}$	$DreD_{LM}62_{ab}$	$OlaDA_{LM}62_{a}$			
63					$PseD_{LM}63$		$OlaDA_{LM}63$			
64			$LmeD_{LM}64$	$HsaD_{LM}64$						
65			$LmeD_{LM}65$	$HsaD_{LM}65$						
66				$HsaD_{LM}66$				$TniDA_{LM}66$		
67		$XtrD_{LM}67_{ab}$	$LmeD_{LM}67_{abc}$	$HsaD_{LM}67_{abc}$	$PseD_{LM}67_{bc}$	$DreD_{LM}67_{bc}$	$OlaDA_{LM}67_{bc}$	$TniDA_{LM}67_{bc}$		
68			$LmeD_{LM}68$	$HsaD_{LM}68$						
69			$LmeD_{LM}69_{abcd}$	$HsaD_{LM}69_{abd}$	$PseD_{LM}69_{abcd}$	$DreD_{LM}69_{abcd}$	$OlaDA_{LM}69_{bc}$	$TniDA_{LM}69_{bc}$		
70			$LmeD_{LM}70_{ab}$	$HsaD_{LM}70_{a}$	$PseD_{LM}70_{ab}$	$DreD_{LM}70_{ab}$	$OlaDA_{LM}70_{ab}$	$TniDA_{LM}70_{ab}$		
71			$LmeD_{MO}71$	$HsaD_{MO}71$						
72			$LmeD_{MO}72$	$HsaD_{MO}72$						
73			$LmeD_{MO}73$	$HsaD_{MO}73$						
74				$HsaD_{MO}74$	$PseD_{NO}74$					

Heterodontus francisci horn shark Hox D

PFC	Dist to 3' gene	Length	Sequence
$HfrD_{AC}1_{ab}$	13,773	317	CTGGTCAAAATGACCCGTGCTTCCTTTCTCCCGATGTGTCATTCATCAA ACTAGTGCTCGTCATTAAGGTACGAATGACGCTGTTCGAATAATCATT TATTGTAACAGGTTTATAAGCAAATAAATACAAGCGTCTGTCAGAGGA TGATGAAGGCGGCCTAGCGACAGGCGAATACATCCAGGTAGAGAAGA ATCAGAGGACAGCCCCTGTCCTCAGCCGATGATTGTATGCCCGGATGC TCTGGGCTTTGGCCT
$HfrD_{DE}10_{bcde}$	3,033	77	GAGCTGCCAGACGGTGTCTGTACATGTTCTCCAGATACTGGGGGCGCC ATAACAAAGTTAAGGTCAAGTTAGTGTCT
$HfrD_{DE}9_{abcd}$	250	88	TAATGACCAAAACTGATATATGGTAATTTCTACATTGGATCACATGAC TCGTTTACCCTTAGAATCAATCAAGATGAATTGCACGTCA
$HfrD_{EF}11_{ab}$	7,328	28	TGTTTACAAAACCTTGAACTGTCTAGAC
$HfrD_{EF}12$	6,741	31	CCCAAAGTCTCTTTGCACAGCATAAATAAAC
$HfrD_{EF}19_{ac}$	218	178	GTCGCTGTTGTCCACAGTTAAGCTCAAATTCAACTGACTTTGGATCATG TGATGTGCGGGGAGGTAGTGCTTCAAGGCCATTTTCAAATTCCATTGG CTCAGTAGTCATGTGGTTGTACAGAGGCATCCACAATTACAAAGGGA ATGTTTTGTTAGAGATGTCAGCTTACAAAGGACA
$HfrD_{FG}21_{abc}$	807	57	TTTTATTAAAGACACATTAATGTGTAATCAAATGCACCTCATAAAAAT TTTATTGAT
$HfrD_{FG}23_{abce}$	158	129	AGCGGATTGATTTACTCGCTGTATTGGTAAATACAATCACGTGAGCCC CGCAACCAATAGCTGCAACTGTAGTCGGCAAAATACTATGATTGTTCT GAGAGAAGGTATCAGATACACTTCCCAGTGTAA
$HfrD_{HK}34$	6,259	27	AGCTGTCGGATGCACAACTGGTACATT
$HfrD_{HK}35_{abc}$	5,928	151	AAATCCGCAAGGAATTGCTGCGAATTCCTTTTTGTTTGAAGAAAATTT ACAACTTTGTAATGGACCTTTTACGACGGAGACTGCGTTCCTATTGGCT GACTTTGGTCATGTGAATCTAGTAACTACGTTCATGGCCTTTTCATGAT TTCCC
$HfrD_{KL}37_{bc}$	-1,572	61	CACGTGATGTCATTAAACTTTGTTTTATGGTTAGGGGAGTTGACAAGCC AAAATATAATTC

Xenopus tropicalis western clawed frog Hox D

PFC	Dist to 3' gene	Length	Sequence
$XtrD_{AC}1_{ab}$	7,140	336	CTGGTCAAAATGACCCATACATCCTCCCAGGCAGCAGTGTCATTC ATCAAAAACTGGGGCTCGTCATTAAGGTCTGAATGACGCTGTTTG AATAATCATTTATTGTAACAGTTTTTATAAGCAAATAAATAGAGGCT CCTGTCAGTGGCTAATGAAGGCGGATTCCTGGCAGACAAATAGGG CCAGGTAGGGAGAGACACAACGAATGGACTGGCCGCTCTCTCTCC CCTCCAGCTACATCTCAGCTCAGCTCCAA
$XtrD_{CE}5$	3,800	27	TTGAGGTAGGGTAACAAATGGCAGAGA
$XtrD_{CE}6$	1,551	39	TTGGCCAAAAGCCTCTTGAGGGAGATTGCAGTTCCCTGG
$XtrD_{CE}10_{abcde}$	3,493	142	CTTGCAGAACTGTGGTGGCAATAAATGAAATGACTCACAATCTCT TCCCCAGTCAGGTTTTACGAGAGCTGCCAGACAGCGTCTGTTCCC GTTCCGCAGATACTAGGGGCGCCATAACAAAGTTAAGGTCAAGTT AGTGTCT
$XtrD_{EF}14$	5,242	34	GGAACAAGGACTTGAATTTGATGTCTCAAAAGCA
$XtrD_{EF}11_{ab}$	5,033	28	TGTTTAGAAAACCTTGAACTGTCTAGAC
$XtrD_{EF}15$	4,801	26	TGAGAACAAAGGGTGGGGGGAGAAAA
$XtrD_{EF}18$	1,098	408	ATTTGTGTTAAAACAAAAAGCAAAAAAAAAGAAGAGTGACCCCT TTTTTTGGTGGCAGTAAAGTGTAAAAATCAATGGGGCGATTTTCAA TAATTCATATTGCAAATGCCATTATGCAAAGTTTTACCTATTAATC TTCCCCCTTTACTAAGTTTCGATTGATCTGCTCGTAAAAGATCTGC CCGTGTGTGGCAACAGAGCGTTATAACTCTTTAAAAATCATCAGA AGGTAGAAATTGACCAAGTATTTCAAGA
$XtrD_{EF}19_{bc}$	162	132	CACGTGACCCACAGGGAGGCAGCAGCTCAAGGCCATTTTCAAATC TCATTGGCTCGCTTGTCATGTGGTTGTGCAGAGGCTTCACAATTAC ACAGGGAATGTTTTGCTAGAGATGTCAGCCTACAAAGGACA
$XtrD_{FG}21_{b}$	679	22	TAATCAAATGCACCTCATAAAA
$XtrD_{FG}23_{abce}$	158	128	AGCGGATTGATTTACTCAGTATTGGTAAATATGATCACGTGGACTC CGCAACCAATGGACGGAGGTTGCAGTCTGCAAAATACTATGATTG TTCTCAGGGAGGGTATTGCATACAGTTAAGAGTGTAA
$XtrD_{HL}35_{abc}$	21,956	154	AAATCCGCAAGGAATTGCAGTAAATTCCTTTTTTGTTTGAAGAAAA TTTACAACTTGGTAATAGACCTTTTTATGACCTCTGCGGGGCTGTG ATTGGCTCCTTCTGGTCACATGCAGGCTGGGATATTCTTCATGGCC TTTTGCCTGATTTCCC
$XtrD_{HL}40_{a}$	19,262	53	GAACAAAGAAAGTATTTCAGGCTATTTGGCAGACAGCTGGGAGAG GTATTTAC
$XtrD_{HL}33$	16,579	34	GTATTTCAAGCAGCTAAAATAAAAAAAACCTAAAA
$XtrD_{HL}37_{bc}$	4,957	61	CACGTGATTCCAATAAACTTTGTTTTATGGCCTGAGAGTTGACAAG CTAAAATATAATTC
$XtrD_{HL}46_{abcd}$	2,790	74	AATAACCGTGTGGCTTTGACCTGTCTGAGCGAGTCGTGCGATAAG GTGAAATTCAGGTCACTCAGTCTAACAAA
$XtrD_{HL}47_{abc}$	2,212	86	TGAACTTTTGTACTGCACTGTGTGACTGTCGGCCACGTAAAAATAA TGGAACTTTTTGATATGTTTGCAAATGATTTCCAATGACC
$XtrD_{HL}48_{abcde}$	1,232	80	GTCTATATGTACCCTGTAGAACCGAATTTGTGTGGTTCGTACAGTC ACAGATTCGATTCTAGGGGGATATATGGTCGATG
$XtrD_{HL}49_{abc}$	143	143	TGATAAACTTTCTGCTCTGTGATTGGCTGCAGCGGTCACATGGCCA CCTAACTTTATTCAGTTGACAGCAAGTAGGAGGGCTTTATGGAGG GAGAAAAAAAGACAACACGAGAAAAATTAGTATTTTCTATCCTCA GAAATTA
$XtrD_{LM}56_{abcde}$	11,397	187	ATTGTTTCTATGGCCAAAAGCTCATAGTTCCTAAAGAGCAGAGGG TTCTTTTTCTAGAAGCTTTCACGTTGAGCAGTTGAAAAGGTATTTTA CTGCAGAAACAGGTTCAACCGAGGGGCAAATTTTCTGTCCGATTA GTTGTATTTCAGCCGGGAGTGCTGACCTCTAAACCCTTGACCCCTT GGAC
$XtrD_{LM}57$	10,334	12	TATTATTTATAC
$XtrD_{LM}58_{bc}$	8,690	105	CACGTGATCCACCAAATAATTAATTCAGCTCGTCCCTTAAGAAAC ACAGCGTCGTCATTAATTTGCCAAGCAAAGGACTCTATCAGACTT GAAAACTGAAGAGAT
$XtrD_{LM}67_{ab}$	7,419	103	TGACAATGGCCCAAGTTGTTGGTATAAATCATTGTAAGTAATTCCT GAAAGGGTGTGAATGTGATGGCGAATGTCCGAGGACTGTAAATCT CCCAGTTTTATT
$XtrD_{LM}61_{ab}$	37,166	42	GTCATAAATTTACCTCCCAAGCAAAATGACAGGTGC
$XtrD_{LM}62_{a}$	35,469	36	GTCATAAATTTACCTCCCAAGCAAAATGACAGGTGC

Latimeria menadoensis coelacanth Hox D

PFC	Dist to 3' gene	Length	Sequence
$LmeD_{AD}1_{ab}$	59,957	331	CTGGTCAAAATGACCCATGCATCCTTTCTAGCCCGAAATGTCAT TCATCAAAAAGTTGTGCTCGTCATTAAGGTAGGAATGACGCTGT TGGAATAATCATTTATTGTAACAGCTTTATAAGCAAATAAATAC AAGCGGCTGTCAGTGGATGATAAAAGCGGCCTTTGTGCAGACA AATACATCCAGGTAGAGACAGAATAGAAGTGAATACTAGATCT CTTGTCCTAAGCTTGATCACTGCTTGGCTGGATTGTT
$LmeD_{DE}10_{bcde}$	2,950	145	CTTGCAACACTGGCGGTAATAAATGGAATGACTCAAGACACTC TTCTCTTGCCTGCTATTTTTTACGAGAGCTGTCAGACAGTGTCTG TTCATGTTCTCCAGATACTGGGGGCGCCACAACAAAGTTAAGGT CAAGTTAGTGTCT
$LmeD_{DE}9_{bcd}$	191	70	TATGGTAATTTCTACATTGAGACACGTGACGCAATTACTCCTAG AATCGATCAAGATGCATTACACGTCA
$LmeD_{EF}11_{ab}$	5,450	28	TGTTTATAAAACCTTGAACTGTCTAGAC
$LmeD_{EF}19_{abc}$	207	177	GTCGCTGTTCACCATATTTGAGGTGAAAGTCAACCATCAACAAC CACGTGACTCCTAACGAGGTAGTGTCTCAAGGCCATTTTCAAAT TTCATTGGCTTCTCTGTCATGTGGGTCTGTAGAGACACTCACAAT TATACAGAGAATATTTTCCTAGAGATGTCAGCCTACAAAGGAC A
$LmeD_{FG}21_{ab}$	726	46	TTTTATTAGGGAGGCATTATAGTATAATCAAATGCACCTCATAA AA
$LmeD_{FG}23_{abce}$	160	130	AGCGGATTGGTTTACTCCTCATATCGGTAAATATAACCACGTGA GCTCCGTAACCAATGGTTGAAGTCGCTATCTGCAAAATACTATG ATTGTTGAGAGTGGAACGTATCTTTTACTCTTAACAGTGTAA
$LmeD_{GH}25$	5,468	235	GAAGGTGTTGGTGCTGTATCACTATTTTCACGCGTTTACCTATAA TGCAATTAAAGCAGGAAGTAACGGGTACAATAAACCCATAGA ATTGGCTAGACGTCTGGAGCTAATGAGTTTATGAGTTGTAACTT GGAATTGCACTGTGCTTTGAAACCGTCTTATTTGGATGTTTGTGT CTGCTGGCGACAAGCAACGATGCAAACACAAATTATTGCAAGG AAATCTAGACAATAAA
$LmeD_{GH}28$	3,362	33	AAAATCCATTGGTCTTCCATTAAGGTACAAACA
$LmeD_{HL}35_{ab}$	19,854	52	AAATCCACAAGGAATTGCAGTAAATTCCTTTTTTGTTTGAAGAA AATTTACA
$LmeD_{HL}37_{b}$	6,869	46	CACGTGATCGCAATAAAACTTGTTTTATGACAAGGGAGTTGACA AG
$LmeD_{HL}46_{abcde}$	2,687	81	AATAACTCTGGCTTTGACCTGTCTGAGCAAGTCGCACAATAAGG TGAAATTCAGGTCACAACGTCTAACAAATTTGAAAAT
$LmeD_{HL}47_{abc}$	2,053	86	TGAACTTTTGTACTTCTTTGCGTGGTTGTCGGCAAAGTAAAAATA ATGAAACTTTGTGATATGTTTGTAAATGATTTAGTATGACC
$LmeD_{HL}48_{abcde}$	1,286	82	GTCTATATATACCCTGTAGAACCGAATTTGTGTGATGTTATCAG AATCACAGATTCAATTCTAGGGGAGTATATGGTCGATG
$LmeD_{HL}49_{abc}$	153	147	TGATAAACTCCTTGGATTTTTATTGGCTATCTCTGTCACATGGGT ACCTAACTTTATTCAGTTGACAGCAAGTAGGAGGGCTCTATGTA GGGAGAGAAAAAAAAAGACAACTCGAGAAAAATTAGTATTTT CTACCTTCAGAAATTA
$LmeD_{LM}56_{abcde}$	12,504	203	ATTGTTTGTATGACAAAAAGCACAGAATTCACGCATAGCAAAG ACGACTTCTTTTTGACGCCTTCACGTTGAGAAGCTGAAAAGGTA TTTTACTGAAGTTCGGCTAAACTGCAGAAACAGGTTCAACTAGG GGACAAATTTTCTATTCGATTAGTTGTATTTCAGCCGGGAGAGC TGACCTCTAAACCCTTGACCTTTTGGAC
$LmeD_{LM}58_{bcd}$	10,151	198	CACGTGATTTACTAAATAATTAATTCAGTACGTCCCCTAAGAAA CACGGCGTCGTCATTAATCGACCAAGCAAAGGACTCTATCAGA CTTGAAAACTGAAGAGATCCCAAGAATATAATATACAAACGGT GCAGTATCCTAATTCTACTCTTGAAGCTCTTTGGTAACATTGGG ACAAATGGAAGCTTGGTAGGTAAA
$LmeD_{LM}64$	9,650	149	GAACAATGCCGACTGTCAGTGCATCTGCTCCTATTCTTAAAGAC GGTGAGAAAAGGGCCTGGACTTCTTTTCAATCACGGGTAGTAA ATTTTTCTTTGCGTCATTAGAGAAAGGCTATAAAACTGAGTCGA ATGTTTCCCAAGGCAGGT
$LmeD_{LM}65$	9,074	143	AAAGTTTCCTTTTTAGAGCTACATTAGGATAATAAACCATTAAA TTAGAGGTTCAAAGATGACATATACTTTCCAGTTTCAGTAGTGA TTGGGGAAATAATTCATTGGTTGCTCATAGCTGCTCATTTTGTTT GACCAGAATC
$LmeD_{LM}67_{abc}$	8,770	116	TGACAATAGCCAGAATTGTTATAAATCATTCTAAGTAATTCATG AAACAGTGCGAGGCTGTTGGGGGCCGGGCGAAGATTGTAAATC TTTCAGTTTTATTGCCCCGTGAACATATG
$LmeD_{LM}68$	8,090	21	ATATTTTACTGTAAAATATAA
$LmeD_{LM}69_{abcd}$	7,246	92	CTCAATCAGAACAATCTGGTATACAGATCACGTGAACAAATAT GCTTGTATTTTAAGGCAGCGCCTATATTTGTGATTATAAAAGGTT TCCG
$LmeD_{LM}70_{ab}$	7,026	233	TTTATTGGTAGTTGAGCCTGAGACTGTTTCCATTCTATCGGGAAT ACTGTCTGCACTGGTATATGGAAATGTCTGTAAAACTGCAAGAT CAGGTTTAGGACAGTATTGTCCGCAGACAAAGGGTGAAGGATA TACACAAGCAACCGCAGGCGAAGAAACTGAGTCTGAGAAGCG CAGGTTTATGATAAATTGATATACAGGTAAGCTATAGCATGACA AAATGGAACCTTTAA
$LmeD_{LM}61_{ab}$	4,621	42	GTTCATTAAGGGGTGAGTTATTGCTGTATAAGCCAAAGGTCA
$LmeD_{LM}62_{ab}$	2,302	45	GTCATAAATTTTGTTGCAAACCCCAATGACAGGTGCATTGATAT G
$LmeD_{MO}71$	19,338	474	TTAATTTATGTACTTTTTCAATAATGTCCGAATTGCATTAAACTC GATATGTATTTATTATTTTAAATAAGAGCATTTTGTATCACTTAT TTATCCTTGTCTTAATGTATTTATGTGAACATTTGTAGAACTTCT ACAGCATGACGTGGACGGGTTGATTCAGATCCATGGTCATTAA ATTTTTCCTCTTTAGTAAACTTGGTTTGAAATAGTTCCGAAAGTA GTTCTGAACAATTGTATCAGTGGCTGGTGTTT
$LmeD_{MO}72$	9,621	30	AAAAATGAAAGAACTGAGCATGTTCACAGT
$LmeD_{MO}73$	7,631	27	AAAAAAAAAAGAAAAAAAAACACTATT

Homo sapiens human Hox D

PFC	Dist to 3' gene	Length	Sequence
$HsaD_{AC}1_{ab}$*	8,977	339	CTGGTCAAAATGACCCATACATCCTTCCGATCCCGAGATGTCA TTCATCAAAAAGTAGCGCCCGCTCGTCATTAAGGTACGAATGA CGCTGTTCGAATAATCATTTATTGTAACAGGTTTATAAGCAAAT AAATACACCCTCCTGTCAGTGGGTAATGAAGGCAGCTTCAGAG CAGACAAATAGATCCAGGTAGGAGGCGAGAAGAGACAAGTG AGGAGGAAGGCTCCGGTCCTTTGCCCGCTCCGAGCCAGTTC
$HsaD_{AC}3$	3,006	55	TAATGCACCCACACACAAACACAAGGCTTGGTCTGTGTTCCTG GCCACCTAAAGA
$HsaD_{CD}5$	2,973	26	TTGAGGGGGAGGGACAGAGACAGAGA
$HsaD_{CD}6$	783	39	TTGGCCACGATTGGTGGATCGAGCTGATGGGTTCCCTGG
$HsaD_{DE}8$	4,835	25	ATGTTTTGGATTCCACCTTAATGTT
$HsaD_{DE}10_{abcde}$*	3,031	142	CTTGCAGAACTGTAGTGGCAATAAATGAAATGACTCAGAATCC CTTCCCCAGTCAGCTTTTACGAGAGCTGCCAGACAGTGTCTGTT CACGTTCTCCAGATACCAGGGGCGCCCTGACAAAGTTAAGGTC AAGTTAGTGTCT
$HsaD_{EF}11_{ab}$*	6,554	28	TGTTTACAAAACCTTGAACTGTCTAGAC
$HsaD_{EF}12$	6,463	32	CCCAAAAGAGATTTCCGCAATGTGCAATAAAC
$HsaD_{EF}13$	6,226	49	TTTGAACTTCCAAAATGTCAAGGTCATCACCTTTAACCTTTCTG AATAA
$HsaD_{EF}14$	5,739	39	GGAACAAAAGGAAATATAGTTATAATGCTTTTAAAAGCA
$HsaD_{EF}15$	5,278	25	TGAGAAGTTGTATAGAGACAGAAAA
$HsaD_{EF}16$	5,360	43	TGCAAAAGAATACCTTGGTATAACCAGATGGTGTGGAGCAGA G
$HsaD_{EF}17$	3,266	29	TTTATTCCATTTGTCTCACCTCAGCCTTA
$HsaD_{EF}18$	1,464	387	ATTTGTATTAAAACAAAAATGACCTTCTTGGTTAGCAGCGAAG GGAAAAAAATCAATAGTATAATTTTCAATAATTCATAAAGCGA ACGCCATTATGGTAAATTTAACTATTAATCTTATCCTTTACAAA GTCTCGATTGATTTGTTAATAAAGTATCTTCCCGTGTGTGGCAA CAGCGGGCATAACTCTTTAAAAACCTTCAGAAGCAAGAAAAT TACCAAGTAGCCCAAGAATGTAGATGAGAATTTTATTGA
$HsaD_{EF}19_{abc}$*	208	178	GTCGCTGTTGTCCGTGCTTACCCGGCCGGCCGGCCAGGCTCTG GAGCACGTGACCCGAGAGGAGGCTGCGGCTCAAGGCCATTTT CAAATCTCATTGGCTTGGTTGTCATGTGGTCGGCAGAGGCATCC ACAATTACACGGGGAATGTTTTCCTAGAGATGTCAGCCTACAA AGGACA
$HsaD_{FG}20$*	2,906	33	TGCATGTTCTGGTCGCATGTATAATGCAATAAA
$HsaD_{FG}21_{ab}$*	691	46	TTTTATTAGGGACACATTAATCTATAATCAAATACACCTCATA AAA
$HsaD_{FG}23_{abe}$*	162	133	AGCGGACTGATTTACTCCCGGTATTGGTAAATATGATCACGTG GGCCGCGCGACCAATGGTGGAGGCTGCAGCCTGCGAACTAGT CGGTGGCTCGGGCGCCGGCGGGGGAGCTGCTCGGCGGCGGACA GTGTAA
$HsaD_{GH}24$	6,120	23	TTTGGTGCTTGATTTCCAGAAAC
$HsaD_{GH}25$	4,591	228	GAAGGTGCTGCGCACCCCTGCCACTGTGGCCAGCTGGGTACCT AGGGCTCTTTGAAAACAGGAAGAGCCGAGGTGTCATAAAGCC ATCTAGCGGGCCAGACGTCTGGAGGTAATGAGTTTACGACAGG CCCAGTGCTTTGCTTTGAAACCATCTCATTTTGATGTTTGTGTTT GTGGGAGAAAAGGAAAAATGCAGACAAAACTGGGTCTTAAAA TCTAGACAATAAA
$HsaD_{GH}26$	4,157	17	TAGAAACAAACAAACAA
$HsaD_{GH}27$	3,022	19	CTATTAATTTCTGTGTTTG
$HsaD_{GH}28$	2,631	31	AAAATCCACTCATTTATCCAGGGTTCAAACA
$HsaD_{HL}30$	17,402	29	ATGTGTCAACATCCAAGCATTGAGGCAAC
$HsaD_{HL}31$	16,788	34	TAATTTAAAATATTGTTTTATTCTGTAACACAGA
$HsaD_{HL}33$	15,895	31	GTATTTCGGAGAGGTTGACAAGTACCTAAAA
$HsaD_{HL}34$	15,588	40	AGCTGTGGGGATCGCTTCCCCAGAGCGTGAGCAGTACATT
$HsaD_{HL}35_{abcd}$*	15,223	177	AAATACCGCAAGGAATTGCAGTAAATTCCTTTTTGTTTGAAGAA AATTTACAACTTGGTAATAGACCTTTTTATGACCTATTTGCGGT CGTCATTGGCTGCGGCTGGTCATGTGCAGGCGCCGCTCGCCTTC ACGGCCTTTTTCCTGATTTCCCCGGCGAATTTCCCCCCGCATTC TG
$HsaD_{HL}36$	13,285	33	AATAAGAAACTTGAGTGATTGTTATTTTATTTG
$HsaD_{HL}37_{bc}$*	5,380	60	CACGTGATTCCAATAAACTTTGTTTTATGGCTTGAGAGTTGACA AGCCAAAATATAATTC
$HsaD_{HL}39$	12,554	27	GAGAAAGACGAAGCTCGAAATTATAAA
$HsaD_{HL}40_{ab}$*	12,419	73	GAACAAAGACAGTATTTCACCCTGGGCTGGCAGGCAGCTGTTA GAGGTATTTACGGCCTTGCCGCAGTGCGGC
$HsaD_{HL}41$	9,512	18	AAAAGACCCCTGGTTCAC
$HsaD_{HL}42$	6,519	22	CGTTTTTACACTGTTTGGTGGC
$HsaD_{HL}46_{abcde}$*	2,614	83	AATAACTCGTTGGCTTTGACCTGTCTGAACAAGTCGAGCAATA AGGTGAAATGCAGGTCACAGCGTCTAACAAATATGAAAAT
$HsaD_{HL}47_{abc}$*	2,005	86	TGAACTTTTGTACTCTTGTGTGCTGCTGTCGGCAAAGCAAAAAT AATGAAAGTTTGTGATATGTTTGTAAATGATTTCGAATGACC
$HsaD_{HL}48_{abcde}$	1,314	82	GTCTATATATACCCTGTAGAACCGAATTTGTGTGGTATCCGTAT AGTCACAGATTCGATTCTAGGGGAATATATGGTCGATG
$HsaD_{HL}49_{abc}$*	143	143	TGATAAACTTGCTCCCTCGCCATTGGCTGGCCTGGTCACATGGC TGCCCAACTTTATTCAGTTGACAGCAAGTAGGAGGGCCCTATG GAAGGAGAAAAAAAGACAACACGAGAAAAATTAGTATTTTCT ACCTTCTGAAATTA
$HsaD_{LM}56_{abcde}$*	14,261	203	ATTGTTTCTATGACAAAAAGCACAGAGTTCATACATAGTCAAG ACGTCTTTTTTCTGACGCCCTCACGTTGAGAAGCTGAAAAGGT ATTTTACCGAAGTTCGGGTAAATTACAGAATCAGGTTCATCCA GAGGACAAATTTTCTATTTGATTAGCTGTATTTCAGCCGGGAGG ACTGACCTCTAAACCCCTAACCTTTTGGAC
$HsaD_{LM}57$	13,480	372	CCACTTAGGGCAGGGAAGGAAATGGGTTTCCATCTGAGAACGT GCTTTGGAGAAAGCTAGGTGTGGAAAAGCTCCAATGCCCATTT GCTATTATTTGTTTCCAGTTTGTTCCTTTAAATATGAGCCAGAA GTGTTTGTGTTGGTGTTTTAAAAACAAAAACAAAAACCGTGTT GGGGTCCTGACTGGGGGAGGGGGAGAGTGAAGTGTTTGCTGAG GACATTGCTCCTCTGACTCCCATCTCACTTTGTCCATCG
$HsaD_{LM}58_{bd}$*	11,855	198	CACGTGATTGGCGAAATAATTAATTCAGCACGTCCCTTAAGAA ACACGGAGTCGTCATTAATCTGCCACGCAAAGGGCTCTCTCCG ACTTGGAAAGTGCAGGGATCCCAAGAATATCACCCGTCCAGG GGGGCCGCGCGGTGCCCCCGGCCCTCCACCCCCGGCCCCCGG CGGGCGCGGGAGCGCGGCCGCAGGTAAA
$HsaD_{LM}61_{abc}$*	5,973	57	GTTCATTAAGGGGTGAGTTATTGCGGTGCGAGCCAAAGGTCAC TTCAAAGGCTTATG
$HsaD_{LM}62_{ab}$*	3,899	45	GTCATAAATTTTGCTACAAACCACAATGACAGGTGCATTGATA TG
$HsaD_{LM}64$	11,225	149	GAACAATGGTCGCTGTCACGGCATCTGCCGCCTATTCTTAAAC CGGTGAGAAAAGGCCCTGGCCCTCTTTTCAAGCGAGGGTCGTA AATTTTTCTTTGCGTCATAATAGAAGGCTATAAAATCGAGTTGA AATTTTACCCCAGGCAGGT
$HsaD_{LM}65$	10,461	151	AAAGTTTCCCCCATTATGAATTATACATTCAAACAATAACACA TTAATTCAATTATTCAAAGATGACAAATGTTTATGTGCTTTGCG AGTGACTCGGGCGCAGATTCCAGGCGCTTTCTCTGAGCTGCTT GCATTTTTCTCAATGAGAATC
$HsaD_{LM}66$	10,176	24	CGGCTTGGCGGGCACTGGCCCGGG
$HsaD_{LM}67_{abc}$*	10,138	119	TGACAATTGCCCGGGTTGGTGTGATAAATCATCGTAAGTAATT CCTGAAAGGGTGCGAGACTGTTGGGGGCCGGGCGAGGACTGT AAATCTTTCCGGTTTATTGCTCTATGAACATATG
$HsaD_{LM}68$	9,398	17	ATATTTCTAAAATATAA
$HsaD_{LM}69_{abd}$	8,541	94	CTCAATCAGTACAAGCTTCCCTCGGGGTCACGTGAACAAATAT GCTTGCATTTGAAGGCAGCGTCTGTATTTCCCGACTATGAGGG GGTTTCCG
$HsaD_{LM}70_{a}$*	8,233	204	TTTATTGGTAGTTGAACCTCAGCCTGGTTCCGTTCTACCCGGGAA TTCCGTGTGCTCGAGTATATGGCCGTGTCTGCGAGCGGCGCAAG ACCAGGGTTGGGACAGTGTTGTCTGCAGACAAAGGGGGAAGG CTAGCTCTGCCCCCCACTGGCGCCCACTCTGAGGCCGAGGACA CCAGGTTTATGATAAATTGGGATCCAGGTAAG
$HsaD_{MO}71$	16,143	491	TTAATTTATATTCTCCTTCCTGTGCCGTAAGGATTGCATCGGAC TAAACTATCTGTATTTATTATTTGAAGCGAGTCATTTCGTTCCCT GATTATTTATCCTTGTCTGAATGTATTTATGTGTATATTTGTAGA TTTATCCAGCCGAGCTTAGGAATTCGCTTCCAGGCCGTGGGGG CCACATTTCACCTCCTTAGTCCCCCTGGTCTGAACTAGTTGAGA GAGTAGTTTTGAACAGTCGTAACCGTGGCTGGTG
$HsaD_{MO}72$	6,777	25	AAAAATAAAAGAAGTCGGGCACAGT
$HsaD_{MO}73$	4,912	27	AAAAAAAAATTCCTGAGTCAGACTATT
$HsaD_{MO}74$	2,912	18	AAAACAAAACAAAACAAA

Polypterus senegalus bichir Hox D

PFC	Dist to 3' gene	Length	Sequence
$PseD_{LM}56_{bcd}$	21,864	128	TCACGTTGAGCAGATGAAAAGGTATTTTTTACTGAAGTTCGGCCCTGAAACAGAAACAGGTTCATCCAGAGGACAAATTTTCTATTCGATTAGCTGTATTTCAGCCGGGAGGACTGACCTATAAACCC
$PseD_{LM}50$	21,808	87	CAGGTTCATCCAGAGGACAAATTTTCTATTCGATTAGCTGTATTTCAGCCGGGAGGACTGACCTATAAACCCTTGACCTGATGGACC
$PseD_{LM}51$	20,249	15	TATAATTGCAAAATA
$PseD_{LM}54$	18,323	17	TTTTTTATTAGTTGTAT
$PseD_{LM}58_{bcd}$	17,931	199	CACGTGGTTTACTAAATAATTAATGCAGCACGTCCCCTAGAAACACGACATCGTCATTAATCGGCCAAGCAAAGGACTTTATCAGCCTTGAAAACTGAAGAGATCCCAAGAATATAATATACTAACCGTCCGGTATCCTACGCTACACTAACCAGCTCTTTTGTATAGCATCTGAACAAATGGAGGCTTGGTAGGTAAA
$PseD_{LM}67_{bc}$	16,624	95	TAAATCATCCGATGCAATCCTTGAAAGGGTGTTAGGCTGCTGGGGGCCGGGCGAAGACTGTAAATCTTTCAGTTTTATTGCCCTATGACCATATG
$PseD_{LM}61_{ab}$	12,717	70	GTTCATTAAACGGTGAGTTATTGCCTAACATGCCAAAGGTCATGTAAAAGGCTTATGACTGCTAAATATT
$PseD_{LM}62$	10,006	45	GTCATAAATTTTACCGCAGCCCACAATGACAGGTGCATTGATATG
$PseD_{LM}63$	13,811	22	CATTTTGTGATCAATACATATA
$PseD_{LM}69_{abcd}$	15,091	91	CTCAATCAGTACAATCTGGTACACCGATCACGTGAACAAATATGCTTGTTTCTGAGGCATCGCCTTTATTTGTCATTATAAAAGGTTTCCG
$PseD_{LM}70_{ab}$	14,873	233	TTTATTGGTAGTTGAGTGTGAGTAGGTTTCCATTCTATCGGGAATACTGTCTGCATTGGTACATGGAAATGTCTGTGAAACCGCAAGATCAGGTTTAGGACGTTATTGTCTGCAGACAAAGGGTGAAGGATATATTCTAGAGACTACAGGCGAAGAAATTGAGACCGAGGAGCACAGGTTCATGATAAATTGATACGCAGGTAAGCAATTGCATGGCAAAATGCAACCTTTAA
$PseD_{NO}74$	2,645	19	AAAACACACATAAAACAAA

Danio rerio zebrafish Hox Da

PFC	Dist to 3' gene	Length	Sequence
$DreD_{AC}1_{b}$	12,618	296	ATTCATCAAAAAGTGGTTTGCGCCTTCGTCGTTAAAGGTGCGCGTGACGCGGCTCGAATGATCATTTATTGTAACAGGTTTATAAGCAAATAAATAGGAGAGAGACTGTCATCGGTTGATAGAGTCGGCCCTTGATCGGCCAGACGCGTATATCCAGGTGGAGAGAGAGAGGAGAAGTGAGGAGAGAGCAGCTTCTGATCTGAGTCTGTTTTAGATGGCTCCTGGGGTTGGCCTCTGGGGTGAAATTGACAGTCT
$DreD_{CD}2$	2,475	27	ATTTTAGACTCGGTGTCTGAATAAGTT
$DreD_{CD}4$	111	75	TACTTTGCATATCACGTGATGGTGCATTAACAAATCAACAATTACCTTGCTCGTATTCTTTAGGGGTAACTAAAA
$DreD_{DE}7$	3,509	53	ATTAATTGCAATAATCTAAAACAAGTATATGGAACAGGACAAACGGGAGAAGA
$DreD_{DE}9_{abcde}$	249	152	TAATGACTGATATTGATATATGGTAATTTTTTTACCGGATCACATGACACAATTACCTCAAGAATCGATCAAGATGAATTGCACGTCAGCGTACGTTTCCGATTTTTTCTTCCTCGCGAGTCCTATCCGCACGCATATGGTGCAATAGATGG
$DreD_{DE}10_{cd}$	2,061	28	GGGGCGCCAATACAAAGTTAAGGTCAAG
$DreD_{EF}11_{b}$	5,617	20	AAACCTTGAACAGTCTAGAC
$DreD_{EF}13$	5,259	49	TTTGAACTTCTGTAATGTCAAGGTCGTCACCCTTAACCTTTTTGAATAA
$DreD_{EF}16$	4,948	46	TGCAAAACAAACAAAGCCGTAATATGCTTATAGTGATCGAGCAGAG
$DreD_{EF}19_{abc}$	226	179	GTCGCTGTTGACCACGTCTGAAACTTCAAATCATGTGTCAATCCGCCCACGTGACGGGGAAGAGCCTGCGTCTCAAGGCCATTTTCAAATTTCATTGGTGAGAGTGTCATGTGGTTGCAGAGAGACTCGCTGGGTTATACAGGGATTGTTTTGCTAGATATGTCAGCTTACAGAGGACA
$DreD_{FG}21_{abc}$	713	58	TTTTATTAAGGGAACATTAATACATAATCAAATGCACCTCATAAAATGTTTTATTGAT
$DreD_{FG}22$	476	33	TTTACAGCATGTTATTACATTTATTTTACGAGG
$DreD_{FG}23_{bcd}$	127	111	GTAAATATGATCACGTGATTCATGTAACCAATCACTGAAGGTGAAGGCAGCAAAAATACTACGATTGTTCGGAGGCAAGGTTTCGGAAACAGAGTACCGTTTTATATGAGT
$DreD_{GL}29$	12,865	81	CAGGAAGAAACGGGGTCCAATAAAGCTATAGAAAGAGCTAGACGTCTGGACTAAATGAGTTTATGGTACACGACTGTAATT
$DreD_{GL}30$	11,317	29	ATGTGTGTGTTTTAAAACAGCGAGGCAAC
$DreD_{GL}31$	10,723	34	TAATTTTATCTGTTTGGGAAATCTGCGGCACAGA
$DreD_{GL}35_{bcd}$	8,951	135	AATTTACAGCTTCGTAATAGATCTTTTTATGAGCCTATTTCGTCTGTCATTGGATGCCACTGGTCATGTGCAGCACGCAAACGTCTTCATGACCCCTTTTCCTGATTTCCCAAGCGATTTTTCCACTACATTCTG
$DreD_{GL}36$	7,359	31	AATAAGTTATGCATGCTTTAATATATATTTG
$DreD_{GL}37_{abcd}$	5,107	136	TATTGAATAAGTGCAACTTTCAGGATTATTTATGGCACCCGCGCGCTGTCATGAATGGCTGTGGGGAAGCACGTGATACCATTAAACTTTGTTTTATGGCCAGGGAGTTGACAAGCCAAAATATAATTCACATTGT
$DreD_{GL}38$	615	191	TTATTGATTGGCACGGCGCATTGATGGCGAACCAGGGCGCCCTATACAGCGAGCTGAAAGGATGAGAAGCGCTGGAACGATGAGACTCCAGTAACTTCTGTGTGACCTCTACATGACAAAGACGGTCTTCATCAATTGTCTGCAGGCTTTATGCGTCTCTTTTTGTGCCCTGGACCGTTTTTATCTTTTGT
$DreD_{GL}40_{ab}$	7,758	72	GAACAAAGACAGTATTTCACAGTCAGCTGACAGGCCGCTGCAGAGGTATTTACAACCTCTCTGCAATGCGGC
$DreD_{GL}46_{bcd}$	3,372	63	GGCTTTGACCCGCCTGAACAAGTCGCAATTCAAGGTGAAACACAGGTCACGCTGTCTAACAAA
$DreD_{GL}47_{b}$	2,771	39	AAAAATAATGAAACATTGTGATGTGTGTTGTAAATGATT
$DreD_{GL}48_{abcde}$	2,257	82	GTCTATATATACCCTGTAGAACCGAATTTGTGTGAAAAAATAACATTCACAGATTCGATTCTAGGGGAGTATATGGTCGATG
$DreD_{GL}49_{b}$	63	60	ATTGGCCAAGCTGGTCACATGGTAGGCTAACTTTATTCAGTTGACAGCAAGTAGGAGGGC
$DreD_{LM}50$	9,100	89	CAGGTTCATCCAGAGGACACAGTTTCTGTTGCATTACAGACGCTTTTCAGCCGCGAGCACTGACCTCTACACTCGTGACCCGATGGACC
$DreD_{LM}51$	9,979	15	TATAATAAAAAAATA
$DreD_{LM}56_{c}$	9,100	65	CAGGTTCATCCAGAGGACACAGTTTCTGTTGCATTACAGACGCTTTTCAGCCGCGAGCACTGACC
$DreD_{LM}58_{abcd}$	6,965	432	TGACCTGGATGTCTGAACAGAATAAATGGGAGCCATACGCAGTTCCATTTGCAATGACACTGCACCGACTACCACCGCCGCTCTCACAATGCCTCTCAAACGGCCATTTTTGATTTCCAGCTGTTCGCTGAAAATGTGTTCGCCCATCGAGACCAGCCGTTAAGATAAATCAAACTGCGCTGTCGACATCCCGAGACAGAATTCGGTATATGGTTCCGTTCTATATCTTTTTTTTGTTTGACCACGTGATTGTCTA
$DreD_{LM}59$	5,325	19	GTCATTGCCTAAACCATAA
$DreD_{LM}60$	3,333	29	CAAACTTACCGTGGGAGTGCGTAGGCGAA
$DreD_{LM}61_{abcd}$	2,922	70	GTTCATTAATCAGTGAGTTATTGGAGAGCAAGCCAAAGGTCACCCAAAAGGCTTATGAGTGCTAAATATT
$DreD_{LM}62_{ab}$	1,724	45	GTCATAAATTTTGTCGGGCTCCACAATGACAGGTGCATTGATATG
$DreD_{LM}67_{bc}$	5,816	95	TAAATCATGCTCAGTAATTCCTGAAAGGGTGAGAGGCTGTTGGGGGCCGGGCGTGGACTGTAAATCTTTCACTTTTATTAGCCCGTGAACATATG
$DreD_{LM}69_{abcd}$	4,802	90	CTCAATCAGGACAATCTGCGATCCGGATCACGTGAACAAATATGCTTGTATTTTAAGGCAGCGCCTTTATTTGTCATGATAAGGTTTCCG
$DreD_{LM}70_{ab}$	4,600	231	TTTATTGGTAGCTGAGTGAGTGGCTCCCATTCTTTCGGGAATACTGTCTGCATTGGTATATGGAAATGCCTGGAAAAGCGCGAGATCAGTTTTAAGACAGTACTGTCTATAGACAAAGGGTGAAGGATATATCCGAGCAGCTGAACTGAGAGATATTGAGACTAAGAAGCACGGCTATATAATAAATTGACATCACAGGTAAGCAATTGCATTGCAAAACGTAACCTTTAA

Oryzias latipes medaka Hox Da

PFC	Dist to 3' gene	Length	Sequence
$OlaDA_{AD}1_b$	9,694	299	ATTCATCAAAAGTGCCGGCGCTTCGCTGGTAAAACGCAGGCTG ACGTTGCTCCAATGATCATTTATTGTAACAGGTTTATAAGCAAA TAAATAGGAGAGGGCTGTCACTTGATGATGGAGGCGGCCTTCG GTCAGACGAATATTATCCAAGTGGAGAGGAGGAGAGGAGGTGA GGAGCGCGCTGCTCGTCCCTGAGTTGATATCTGAGTCTGTTTTA GATTGCTCTTGGGTTTGGCCTCATGGGCGAAATTGACAG
$OlaDA_{AD}2$	4,918	27	ATTTTAAGTTCATTAAGTTCATAAGTT
$OlaDA_{AD}3$	3,333	57	TAATGCATTACACAAACAAGTCACATTGGTTTAAATATAATGA TTATCAATTAAAGA
$OlaDA_{DE}8$	3,353	24	ATGTTTGCATGTTTATTAAATGTT
$OlaDA_{DE}10_d$	1,910	16	CAAAGTTAAGGTCAAG
$OlaDA_{DE}9_{abce}$	218	153	TAATGACTGATATTGATGTATGGTAATTCCTTAGCTCGATCACA TGACACAATTACCTCAAGAATCGATCAAGATGTATTGCAGGTC TGCCTGCGTTTCCTATTTTTTTCTCCCTCGCTGGGTCTTTCCACA CGCGCGCCGTGCTATAGATGG
$OlaDA_{EF}11_{ab}$	2,657	28	TGTTTACTAAACCTTGAACCGTCTAGAC
$OlaDA_{EF}17$	1,738	30	TTTATTGTCTTCTAAATTGCAATAGCCTTA
$OlaDA_{EF}19_{ac}$	193	162	GTCGCTGTTGACCACGTCTGCATCCGCAGTGGGAACGCGTGCG TGGCGAGGAGGCTGCGTCTCAAGGCCATTTTCAAATCTCATTG GTGGGTTTGTCATGTGGTCTGGAGGCATCCTGACTTACAGATTG TTTTTTCTAGATATGTCAGCTTACAAAGGACA
$OlaDA_{FG}20$	1,409	33	TGCATGTTTGTTGCATGTATATAATGCAATAAA
$OlaDA_{FG}22$	434	33	TTTACAGCATGTTATTACATCTCATTTACGAGG
$OlaDA_{FG}23_{bcd}$	128	113	GTAAATATGATCACGTGATCCACGTAACCAATCCCTGTAGATG CAGGCCAGCAAAAATACTATGATTGTTCACAGAGGGAAGCTTC CCGTACGAGTGCCTGTATTTTATGAGT
$OlaDA_{GL}40_{ab}$	5,098	71	GAACAAAGACAGTATTTCACTGGTGCCTGACAGGCAGCTGCG AAAGTATTTACAGCCAACTGCAATGCGGC
$OlaDA_{GL}27$	4,754	16	CTATTAATTTTGTTTG
$OlaDA_{GL}29$	11,025	80	CAGGAAGGAGCGTGGTCCAATAAAGCTAAAGAAATGGCCAGA CGTCTGGTCTAAATTAGTTTATGACTCTTTCAGTAATT
$OlaDA_{GL}32$	3,631	24	GGATTGGCCTAAAAGAAGGAATGA
$OlaDA_{GL}35_{bcd}$	7,177	136	AATTTACAACTTGGCAATAGAGGTTTTTATGTGCCTCCATCGCC TGTCATTGGATGCCACTGGTCATGTGTGAGAGGCAAACGTCTT CATGGCCCTTTTCCTGATTTCCCAGGCGATTTCCCCCCACTGCA TTCTG
$OlaDA_{GL}37_{abcd}$	3,308	135	TATTGAATAAGTGCAACTTCGGGGATTATTTATGGGCCAGTGC GGGGTTGTGAATGGCTGCGAGGAAACACGTGACGCCATTAAA GTTTGTTTTATGGCTTGGGACTTGACAAGCCAAAATATAATTCT CATTGT
$OlaDA_{GL}38$	641	205	TTATTGATTCCCAGAAGCTGTTGCCACTGACCAATGGGCTCCC GAAACACTTTTCTTCTCCTTTTTTTCAAAGGGACAGGGGCAGAT GAAGAGATGAGAGTCATGTAACTTTTGGATGACCCCATCTTGA CAAAGACAGTGTCCATCACTTGTCCTCAGTCATTCTTCCCTGCT TTTATGCCAAAGGGCTTTTTGTGTCTTTTGT
$OlaDA_{GL}46_{cde}$	2,326	41	AAGGTGAAAGACAGGTCAGGGCGTCTAACAAATATCAAAAT
$OlaDA_{GL}48_{abcde}$	980	82	GTCTATATATACCCTGTAGAACCGAATTTGTGTGAAGAACGCT CAATCACAGATTCGATTCTAGGGGAGTATATGGTCGATG
$OlaDA_{GL}49_{bc}$	123	123	ATTGGCCGATCTGGTCACATGGTTCGCTAACTTTATTCAGTTGA CACCAAGTAGGAGGGCTTTATGGAGGGAGGAAAAAAAGACAA CTCGAGAAAAATTAGTATTTTCTACCTTCAGAAATTA
$OlaDA_{LM}52$	7,599	307	TGTTTTAACTTATTTTGTACAAAAATAACTAGCTTGGAAATCTC CTTAACCCAAATTTCTATTCCCACACGGAAGACATTCCTGCGC GCATTCTGGTCTGAGCGCCCTTATAAATCTCCTCTGACATCTCT GTAAACAGTGTACATTTTCAGTGGGCTCTCTTATTCTATATGAA ACTCACATGAATGATCATTCGATCCTGCATGGCTCAAATTGATT TTGGCCGTTTTTACCCCCATCGGACATAATCGAATA
$OlaDA_{LM}53$	7,101	336	ATATTTTTTGTTTTATATGGTCATTGTGCACGAATAGAAAAAGC CCAGAAGCTTTTTTAGTACAAGGAGGTCTCTACTCTAAAGTGA TGAGAGGGTTGAAAAGGTATTTCATCCGAAGGTCGACAAAGG CAGCTGAGGCAGGTTCGTCCCGGGGACACATTTCAGCCGCGAG GACTGACCCCAGCACCCCTGACCCGCCGGACAGGCAGCATTTT CTCTCCCAGGCGCTACTCGCTTTGGTTTGGAAGACTGTTG
$OlaDA_{LM}54$	6,678	15	TTTTTTATCTTGTAT
$OlaDA_{LM}55$	5,731	60	AGACAAAGTGATGCTTCCAGTTCAGCGTCCTGTCAAAAGTGAT CCGGATATTGATGCTGA
$OlaDA_{LM}58_{abcd}$	5,413	425	TGACCTACATGTCTGAACAGCAATAAATGAGAGCCAAAGGCA GTGGCATTTTCAATGACACTGCACCCAGTGCATGAACACAACA GCCTTGCTCTCCATCCGGCCATGTTTGATTTCCAGCTGTTTTTTC GCAGAAACCTCTCCGAGCCAGACAACCCCCTAAGATATATTA AAGTGCGCTGTCTGCATTCAGAGATATAATCCAGAGACTATGC CACTCCGCTCTCTATTCTCGACCACGTGATTGTCTAAATA
$OlaDA_{LM}59$	4,045	16	GTCATTTTCCCCATAA
$OlaDA_{LM}60$	2,310	29	CAAACTTACCGTAACAGCGCGTAGGCGAA
$OlaDA_{LM}62_a$	916	36	GTCATAAATTTTGCCACGGTCCACACTGACAGGTGC
$OlaDA_{LM}63$	2,423	22	CATTTTTGGACAAAAGCATATA
$OlaDA_{LM}67_{bc}$	4,439	93	TAAATCAGTGTAAGTAATCCTTAAGTGTGCGAGGCTGTTGGGG GCCGGGCGAAAACTGTAAATCTTTCACTTTTATTACCCTCCGA ACATATG
$OlaDA_{LM}69_{bc}$	3,614	47	TCACGTGAACAAATATGCTTGTATCTAAAGGCAGCGCCTTTAT TTGT
$OlaDA_{LM}70_{ab}$	3,430	235	TTTATTGGTAGTTGAATGCAAGTCCTTCCATTCTTTCGGGAATA CTGTCTCCACTGGTGTATGGAAATGTCTGAAAAACAGCAAGAT CAGGTTTAGGACAACTCTGTCTGCAGACAAAGGGTGAAGGATT TATTCTAGAGACTGCGTGGAAAGGAACCCATGCTCAGAGGAA ACTTGCTTTATATGATGAATTGATTCCCAGGTAAGAAACTGCA CGGCGGAAAGCAACCTTTAA

Tetraodon nigroviridis spotted green pufferfish Hox Da

PFC	Dist to 3' gene	Length	Sequence
$TniDA_{AD}1_b$	5,394	297	ATTCATCAGAAGTGCTCGCCGCTCGTTGTTAAAATGCTCGCTGAC GTCGCTCCAATGATCATTTATTGTAACAGGTTTATCAGCAAATAA ATAGGAGAGGGCTGTCACTTGATGATGGAGGCGGCCTTCGGTCA GACGAATAATATCCAGGAGGAGAGGAGGAGGGAGTGAGGAG AGATGCTCGTCCCTGGGTTGATCTCTCCGTCTGCTCCACATTGCT CTTGGGTTTGGCCTCTTGGGTGAAATTGACAGTC
$TniDA_{AD}4$	111	75	TACTTTGCATATCACGTGACGCTGTATTAACAAATCAACAATTA CCTACCCTCGATTCTTCAGGAGTTGCTAAAA
$TniDA_{DE}7$	3,451	52	ATTAATTCTCCTAAATAAGACAATTATACTGTTTTTAAAAAAAA GAAGAAGA
$TniDA_{DE}10_{cd}$	2,186	28	GGGGCGCCAAAGCAAAGTTAAGGTCAAG
$TniDA_{DE}9_{abe}$	241	155	TAATGACCGATATTGATGTATGGTAATTTCTTGGCCGGATCACAT GACGCGATCACCTCCAGAATCGATCCGGTTGTGCTGCAGGTCTG CCCACGACTCCGATTTGTTTCCCTCCCTGCCTGAGCTTTCCCACA CGCTTGCGCGGCTACAGATGG
$TniDA_{EF}11_{ab}$	2,387	28	TGTTTAGCAAACCTTGAACCGTCTAGAC
$TniDA_{EF}19_{ac}$	214	177	GTCGCTGTTGACCACGTCTGTATCCGCAGTGCGCGGACGCAGAC GTGGGAGGGGGAGAGAAAAGGCTGCGTCTCAAGGCCATTTTCA AATCTCATTGGTGGGCTTGTCATGTGGTCGGGAGACATTCGGACT GATCCGTAGATTGTTTTTCCCAGATATGTCGGCGTACAAAGGAC A
$TniDA_{FG}21_{abc}$	646	51	TTTTATTGACAGCACGTATAATCACACGCACCTAATAAAACATT TATTGAT
$TniDA_{FG}22$	437	32	TTTACACCATATCATTACAGCATTTTACGAGG
$TniDA_{FG}23_{bcd}$	137	122	GTAAATATGACCACGTGATCTATGCAACCAATCCCTGTAGATGC AGGCCACCAGAAAAGGAAATACTATGATTGTTCATAGAGGGAA GCTTCCCATAACAGCGCGACCCTTATTATATGAGT
$TniDA_{GL}32$	3,309	24	GGATTGGCCAGGATTAATGAATGA
$TniDA_{GL}35_{bc}$	5,149	111	AATTTACAACTTAGTAATAAACCTTTTTTATGTGCCTCCATCGCC TGTCATTGGATGCCACTGGTCATGTGTGAGAGGGAAACGTCTTC ATGGCTTTTCTCCTGATTTCCC
$TniDA_{GL}37_{abcd}$	2,995	136	TATTGAGTAAGTGCAAGTTTGCGGATTATTTATGGTCTTTCGCGG TGTTGTGAGTGGCTGTGAGGAAACACGTGACACCATTAAAGTTT GTTTTATGACCCGGGAGTTGACAAGGCAAAAATATAATTCTCAT TGT
$TniDA_{GL}46_b$	1,941	18	AAGGTGAAAGACAGGTCA
$TniDA_{GL}38$	575	184	TTATTGATTCCCAGAGGCGGCTGCTGCCGACCAATAGGCTCCCG AGACAGTTTTTGGTGAAGGGGGGCAGCAGATGAAGTGATGAGAT TCGTGTAACTTTTGGATGACCCCTCCTTGACAAAGACAGTCTTCA TCAATTGTCCCCAGTCATTCTTCGCTGCTTTTATGTGATGGGGGCT TTTGT
$TniDA_{GL}40_{ab}$	4,229	72	GAACAAAGACAGTATTTCACTGCGGCCTGACAAGCAGCTGCGA AAGTATTTACAGCCTTGCTGCAATGCGGC
$TniDA_{GL}48_{abcd}$	930	76	GTCTATATATACCCTGTAGAACCGAATTTGTGTGATCAAGTCAC AGTCACAGATTCGATTCTAGGGGAGTATATGG
$TniDA_{GL}49_b$	63	60	ATTGGCCGATCTGGTCACATGGTTCGCTAACTTTATTCAGTTGAC ACCAAGTAGGAGGGC
$TniDA_{LM}52$	10,660	305	TGTTTTAACTTATTTTGTACAATAAAGCTTTGAAATATTCTTAGCC TAAATATTATTCCCAACAGCAAGACCTTCCTGCGTTCAAACTCTT CCGAGTGCCCTAATAAATCTTTGCTGACATCTATCTAAACAGTGT ACATTTTCAGAGGGTCTTTTGTGCCATTTGAAAATTAAATGAAAG ATCGTTTCTTCCTGCATGGCATTAAATAATTGATGGTCCTTTTTGG TCAGAAATAATCAAAGCCAGATTATTTA
$TniDA_{LM}53$	10,159	347	ATATTTCTTGTTTGATACCGTCATTGTCCTGAAACAAAAAAGCAC AGAATCTCTCGCAGTCCAAGGAGGTCCTCTTTCTGAAGTCTTCAC CATGCCGAGCCTGAAAAGGTATTTCAGCCGAAGTCCGACAGAA CAGCGGAGGCAGGTTCATCCAGGGGACACGTTTCAGCCGCAAG GAGCTGACCCCAGCACCCATGACCCGCTGGACCGAGCGGCATTTT CTCTCCCAGGCGCTTTTGTCTTTCAGTTTGGACCC
$TniDA_{LM}55$	9,047	56	AGACAAAGTGATGCTATTTTCACCATGCTGCCAGAAATGATCTG CACATATGCTGA
$TniDA_{LM}58_{abcd}$	8,782	423	TGACCTACATGTCTGAACAGCCATAAATGGGAAGCAAAGGCGC TGCCATTTTCAATGACACTGCACCCGGTGCAGGAATACGGCGTC GTCGCTCTCCATCCGGCCATATTTGATTTCCAGCTGTGTTCGCAG AAGCTGCTGCGCGCCAGACAGCCCCCCCAAAGATACATGAGAG CGCGCTGTCTGCATTCAGAGATATAATACTGAGACTATACCACT CCGCTCTCTATTCTGGACCACGTGATTGTCTAAATA
$TniDA_{LM}60$	5,798	29	CAAACTTACCGTAACAGCGCGTAGGCGAA
$TniDA_{LM}61_{acd}$	5,434	70	GTTCATTAAACTTCGGCTTAACGCAAAATGACCCGGGGGTCACT AAAAAAGCTTATGACTGCTAAATATT
$TniDA_{LM}66$	5,421	24	CGGCTTAACGCAAAATGACCCGGG
$TniDA_{LM}67_{bc}$	7,791	96	TAAATCACTCTCGGTAATTCCTAAAAGGGTGCAAGGCTGTTGGG GGCCGGGCGAAAACTGTAAATCTTTCACTTTTATTGCCCTCCTGA ACATATG
$TniDA_{LM}69_{bc}$	7,004	41	TCACGTGAACAAATATCTTATAACGCTGCGCCATTATTTGT
$TniDA_{LM}70_{ab}$	6,797	225	TTTATTGGTAGCTGAACGCGAGTCCTTCCATTCTTTCGGGAATAC TGTCTCCGTTGGTTTATGGAAATGTCTGAAAAACAGCAAGATCA GGTTTAGGACAGCACTGTCTGCAGACAAAGGGTGAAGGATTTAT CCCCGTGCACTGCGTTTAAGGGAGCTGAGGCTGAGGAGGGAGA GGCGCTCTCATGACAAATTGATGCCCAGGTAAGAAACTGCAAC CTTTAA

Oryzias latipes medaka Hox Db

PFC	Dist to 3' gene	Length	Sequence
$OlaDB_{GL}42$	6,176	22	CGTTTTACGACCACTAGGTGGC
$OlaDB_{GL}43$	5,693	107	GACGCGCAGCGTGTTTTCACAGGGTCGCTTTATGGCTTTATTGCTCC CCATGTACGGAAATCAAAGCAGGAAACAAATGCTCCAAACAGCA TTCCAGACACATCCGG
$OlaDB_{GL}44$	5,351	160	CCTAATAATTTATGAACCCTTGTGGGAGTGGGGGGGGGGGGGGTC AGAACACGTGACAGTAATAAAGCGTATTTTATTGCCGCCGTTTGAC AAGCCCCAAAATATAACTCAGACAGAAAAGCCGACAGGCAGCCA GATGGGATCCGCGCTGGATATTTG
$OlaDB_{GL}45$	218	218	ATGAACATTTTTTGTCCTCCAAATTATCCCATTACATCTGACATAAA CAGCCCAACGTGCTGCCTCTCATTGGCTGGCCGCAGGGTCACATGA CCCACACGACCGTCCGTCTATTTGACAGCCGCAGGATGGCTTGATG CCAGAGGGGGAAAAGGGACAGAGAGAGGGGAGAGAAAGAAAAA TTAGTATTCTCCTACAAGCCTCTCTATAAATCAATC
$OlaDB_{GL}48_{bc}$	1,498	58	TACCCTGTAGAACCGAATGTGTGTGGAGCTGCCTCAGTCACAGATT GGGTTCTAGGGG

Tetraodon nigroviridis spotted green pufferfish Hox Db

PFC	Dist to 3' gene	Length	Sequence
$TniDB_{GL}24$	6,494	23	TTTGGTCATCGTGGTCTAGAAAC
$TniDB_{GL}26$	4,430	17	TAGAAATCGGTGAACAA
$TniDB_{GL}39$	11,470	24	GAGAAAACGGGCCGAATTTATAAA
$TniDB_{GL}41$	8,478	18	AAAAGAAAGATCGTTCAC
$TniDB_{GL}43$	5,598	107	GACGCGCGGTATGTTTTCACAGACTTGGTTTATGGTTTTATGGCTCC CCTTTTACGGAAATCAAAGCAGGAAACAAATGCTTCAAACAGCAT TCCTGAAACATCCGG
$TniDB_{GL}44$	5,178	160	CCTAATAATTTATGAACCCGGGTGGGAGCAGAGAGGGGGGAAAAG TGAGCACGTGACTGTAATAAAGTGTGTTTTATTGCCGGCGGTTTGAC AAGCCCCAAAATATAACTCAGAGAGAAAAGCCGACAGGCGGCCA GATAGGATTGGTAGCCGATATTTG
$TniDB_{GL}45$	224	224	ATGAACACTTTTTTCTCGCTTTAAACGATCCCATTACCTTTGACATA AACAAACCCCCAAAAGTGCCGGCTCTTATTGGTCGGCCGCGGGGTC ACGTGACCCGCGTGGCCGTCCGTCTATTTGACAGCCGCAGGATGGC TTGATGCCAGAGGGGGAAAAGAGACAGAGAGAGGGGAGAGAAAG AAAAATTAGTATTCTCCTACCAGTCTCGCTATAAATCAATC
$TniDB_{GL}48_{ab}$	1,921	27	GTCTATACATACCCTGTAGAACCGAAT

Table S2-Conserved PFC Data Supplement to Figure 9

Data in this table corresponds to Fig. 9. For each hox cluster the order in this file corresponds to the order in the figure from left to right. Each PFC is also color coordinated with the lines representing them in Fig. 9. The groups are coded as such 1) horn shark, 2) >2 lobe-finned fish, 3) bichir, 4) >2 teleost fish. In this analysis the PFC was color coded according to the follwing code: Red lines represent PFCs that are found within untranslated regions (UTRs) of Hox gene mRNAs. Blue lines represent PFCs that are found within 500 nucleotides of a Hox gene proper region but have not been identified as UTRs. Green lines represent PFCs that have high or identical sequence identity to sequences in the EST library. Yellow lines represent PFCs that are not found in any databases or published papers. Black lines are sequences with known function and labeled accordingly. If a PFC fits into more than one category then colors are used in this decreasing order: Black, Red, Blue, Green, then Yellow. In the description block Fantom3 refers to a non-coding database of Mus musculus that was submitted to the RNAdb (Pang et al. 2005). RNAz refers to non-coding RNA predicted to have function as assessed by Washietl et al. 2005. Evofold also refers to non-coding RNA predicted to have function as assessed by Pedersen et al. 2006.

Hox A Cluster

PFC	Group	Color	Identity	Accession #	Brief description
$A30_{BC}$	1,2,3,4	Yellow	unknown function		
$A31_{AC}$	1,3,4	Red	DreAa 88/89-103 DreAa 91/91-103	#CK142034.1 #XM_693696.2	Dre whole body cDNA Dre HoxA13a mRNA
$A43_{CE}$	1,2,3	Green	HsaA 101/101-101 HsaA 69/74-101	#BG259921.1 #AK020562	Hsa cancer cell line Fantom3
$A57_{CE}$	2,3,4	Green	HsaA 175/175-175 HsaA 70/70-175	#BM742665.1 #AF071165.1	Hsa stomach cDNA Hsa A11 antisense
$A60_{CE}$	1,2,3,4	Red	HsaA 169/169-169 HsaA 70/70-169 HsaA 94/94-169	#DB289082.1 #BC033706 #U20366	Hsa uterus cDNA Hox A11 mRNA Hox A11 antisense
$A62_{EF}$	1,2,3	Yellow	unknown function		
$A75_{EF}$	1,2,3,4	Red	HsaA 106/106-106 HsaA 68/68-106	#BQ215499.1 #NM_018951	Hsa cancer cell cDNA Hox A10 mRNA
$A78_{FG}$	1,2,3	Green	HsaA 164/167-170 HsaA 70/70-170	#CO601426.1 #16273071	Cfa testis cDNA RNAz
$A79_{FG}$	1,2,3,4	Black	HsaA 54/54-54 HsaA 54/54-54	#BQ923616.1 #15105502	Hsa eye cDNA **mir-196b**[6]
$A92_{FG}$	1,2,3	Red	HsaA 198/209-209 HsaA 155/162-209	#CF957153 #NM_010456	rice leaf mRNA, also rat Mmu HoxA9 mRNA
$A98_{GI}$	1,4	Yellow	unknown function		
$A103_{GI}$	1,2,3,4	Green	XtrA 94/94-129	#CR440650.1	Xtr tailbud cDNA
$A111_{GI}$	1,2,4	Red	HsaA 169/169-169 HsaA 166/169-169 HsaA 64/64-169	#BP263590.1 #NM_010455 #16628248	Hsa intenstine cDNA Mmu HoxA7 mRNA Evofold
$A119_{IJ}$	1,2,3,4	Green	HsaA 160/172-192 HsaA 187/192-192	#BY737288.1 #AK146012	Mmu placenta cDNA Fantom3
$A130_{IJ}$	1,2,3	Blue	unknown function		
$A132_{JK}$	1,2,3,4	Red	HsaA 285/292-292 HsaA 215/217-292	#DN125225.1 #NM_010453	Scr early devo cDNA Mmu HoxA3 mRNA
$A136_{KL}$	1,2,3,4	Black	HsaA 201/201-201 HsaA 52/52-201 HsaA 40/40-201	#DB026292.1 #AK028207 #16628248	Hsa testis cDNA Fantom3 Evofold **A5 mesoderm enhancer**[1]
$A140_{KL}$	1,2,3,4	Yellow	unknown function		
$A143_{KL}$	1,2,3,4	Yellow	unknown function		
$A149_{KL}$	1,2,3,4	Black	HsaA 132/138-152	#BU205782.1	Gga embryo cDNA **A4 RARE**[7]
$A157_{KL}$	1,2,3,4	Blue	HsaA 49/49-71	#DT837650.1	Bta liver cDNA
$A179_{LM}$	1,2,3,4	Yellow	unknown function		
$A181_{LM}$	1,2,3,4	Yellow	unknown function		
$A193_{LM}$	1,2,3,4	Yellow	unknown function		
$A194_{LM}$	1,2,3,4	Green	HsaA 28/28-151 DreAa 67/67-151	#DB001242.1 #CO923716.1	Hsa tumor cDNA Dre larvae cDNA
$A198_{LM}$	1,2,3	Green	HsaA 114/117,28/28-166 HsaA 114/117-166	#BB660130 #AK142386	Mmu embryo cDNA Fantom3
$A199_{LM}$	1,2,3,4	Red	HsaA 197/199-197 92/92-197	#DA631881 #NM_153631	Hsa kidney cDNA Hsa HoxA3 mRNA
$A204_{MN}$	2,3,4	Black	HsaA 76/82-84	#BY717422.1	Mmu embryo cDNA **A2 enhancer**[2]
$A210_{MN}$	1,2,3	Green	HsaA 155/162-162	#AI046800.1	Mmu stem cell cDNA
$A211_{MN}$	1,2,3,4	Red	HsaA 207/207-207 HsaA 131/131-207	#CR452916.1 #NM_006735	Bta extraembryonic cDNA Hsa HoxA2 mRNA

Hox B Cluster

PFC	Group	Color	Identity	Accession #	Brief description
$B50_{GH}$	2,3	Red	HsaB 203/221-342 HsaB 163/165-342	#BU450737.1 #NM_010461.2	Gga ovary cDNA Mmu HoxB8 mRNA
$B51_{GH}$	2,3	Red	HsaB 61/63-71 HsaB 69/71-71	#BU450737.1 #NM_010461.2	Gga ovary cDNA Mmu HoxB8 mRNA
$B52_{GH}$	2,3	Red	HsaB 153/153-153 HsaB 148/153-153	#DA691980.1 #NM_010461.2	Hsa NT2 cDNA Mmu HoxB8 mRNA
$B60_{HI}$		Yellow	unknown function		
$B63_{HI}$	2,3	Red	HsaB 154/161-238 HsaB 220/239-238 HsaB 54/54-238	#CB805852.1 X06762 #16628248	Rno prostate cDNA Mmu Hox2.3 mRNA Evofold
$B65_{IJ}$	2,4	Yellow	unknown function		
$B70_{IJ}$	2,4	Blue	HsaB 61/61-61	#DA689465.1	Hsa NT2 cDNA
$B79_{JK}$	2,3,4	Red	HsaB 110/110, 107/107-246 HsaB 108/110, 107/107-246 HsaB 106/106-246	#DN996030.1 #BC040755.1 #16628248	Hsa cancer cDNA Mmu HoxB5 mRNA Evofold
$B87_{KL}$	2,3,4	Green	HsaB 75/75-204 HsaB 87/87-204	#AA373970.1 #16628248	Hsa stem cells cDNA Evofold
$B90_{KL}$	2,3,4	Yellow	unknown function		
$B102_{KL}$	2,3,4	Green	HsaB 188/200-200	#CR522341.1	Mmu cDNA clone
$B104_{KL}$	3,4	Yellow	unknown function		
$B108_{KL}$	2,3	Blue	unknown function		
$B110_{LM}$	2,4	Red	DreBa 71/71-71 XtrB 71/71-71	#CT631859.2 #BC090114	Dre myoblast cDNA Xtr HoxB4 mRNA
$B113_{LM}$	2,4	Green	HsaB 152/156-156 HsaB 122/122-156	#BG835561 #16628248	Ssc pooled cDNA evofold
$B118_{LM}$	2,4	Green	HsaB 184/184,81/82-393 HsaB 82/82-393	#BF996393.1 #16273071	Hsa placenta RNAz
$B139_{MN}$	2,4	Yellow	unknown function		
$B140_{MN}$	2,4	Red	HsaB 104/104,28/28,24/24-232 HsaB 104/104,28/28,24/24-232 HsaB 81/81-232 HsaB 33/33-232	#BG114205.1 #NM_002146.4 #16628248 #16273071	osteosarcoma cDNA Hsa HoxB3 mRNA Evofold RNAz
$B161_{NO}$	2,4	Blue	unknown function		

Hox C Cluster

PFC	Group	Color	Identity	Accession #	Brief description
$C8_{CD}$	2,4	Red	HsaC 91/97-172	#BC120847.1	Mmu HoxC12 mRNA
$C26_{DE}$	2,4	Red	HsaC 95/95-257 HsaC 95/95-257 HsaC 28/28-257	#DA906037.1 #BC001543.1 #16628248	Hsa neuroepi cDNA Hsa HoxC11 mRNA Evofold
$C35_{EF}$	2,4	Blue	HsaC 123/123-204	#AI830498.1	Hsa kidney cDNA
$C38_{FG}$	2,4	Green	HsaC 146/161-161	#CK024405.1	Dre whole cDNA
$C40_{FG}$	2,4	Black			**mir196-2**[3]
$C50_{FG}$	2,4	Yellow	unknown function		
$C52_{FG}$	2,4	Red	HsaC 159/164-164 HsaC 159/164-164 HsaC 68/68-164 HsaC 33/33-164	#BU514229.1 #NM_008272.3 #16628248 #16628248	Mmu limb cDNA Mmu HoxC9 mRNA Evofold Evofold
$C55_{GH}$ $C56_{GH}$	2 4	Black			**C8 early enhancer**[4]
$C67_{HJ}$	2,4	Green	DreC 130/130-130	#CT615268.1	Dre myoblast cDNA
$C87_{KL}$	2,4	Green	HsaC 101/101-206	#BP872195.1	Hsa kidney cDNA
$C99_{KL}$	2,4	Green	HsaC 142/164-164 HsaC 108/108-164	#BJ083098.1 #16628248	Xla tailbud cDNA Evofold
$C102_{KL}$	2,4	Red	HsaC 149/151-151 HsaC 148/151-151	#AA792111.1 #NM_013553.1	Mmu myotubes cDNA Mmu HoxC4 mRNA

Hox D Cluster

PFC	Group	Color	Identity	Accession #	Brief description
$D1_{AC}$	1,2,4	Green	HsaD 266/280-339 HsaD 46/46-339 HsaD 31/31-339 HsaD 65/65-339	#CA327595.1 #16628248 #16628248 #16628248	Mmu brain cDNA Evofold Evofold Evofold
$D10_{DE}$	1,2,4	Green	HsaD 142/142-142	#AW785383.1	Ssc cDNA
$D11_{EF}$	1,2,4	Yellow	unknown function		
$D19_{EF}$	1,2,4	Red	HsaD 175/178-178 HsaD 40/40-178	#CV397708.1 #NM_002148.3	Hsa kidney cDNA Hsa D10 mRNA
$D21_{FG}$	1,2,4	Yellow	unknown function		
$D23_{FG}$	1,2,4	Red	HsaD 133/133-133	#NM_014213.3	Hsa HoxD9 mRNA
$D35_{HL}$	1,2,4	Green	HsaD 177/178-177 HsaD 28/28-177	#DA384584.1 #1628248	Hsa thalamus cDNA Evofold
$D37_{HL}$	1,2,4	Yellow	unknown function		
$D46_{HL}$	2,4	Black			**D4 RARE**[5]
$D48_{HL}$	2,4	Black	HsaD 79/82-82 HsaD 82/82-82	#CJ057604.1 #X17360.1	Mmu spinal cord cDNA Hsa Hox 5.1 mRNA **mir-10b**[3]
$D49_{HL}$	2,4	Red	HsaD 143/143-143 HsaD 143/143-143	#DR006918.1 #NM_014621.2	Hsa spleen cDNA Hsa HoxD4 mRNA
$D58_{LM}$	2,3,4	Green	DreD 243/255-432	#DT228767.1	Ppr(fish) brain cDNA
$D61_{LM}$	2,3,4	Yellow	unknown function		
$D62_{LM}$	2,3,4	Yellow	unknown function		
$D67_{LM}$	2,3,4	Yellow	unknown function		
$D69_{LM}$	2,3,4	Yellow	unknown function		
$D70_{LM}$	2,3,4	Green	DreD 171/172-172	#AI883324.1	Dre liver cDNA

10. Bibliography

Aboobaker A, Blaxter M. 2003. Hox gene evolution in nematodes: novelty conserved. Curr Opin Genet Dev 13:593-598.

Acampora D, D'Esposito M, Faiella A, Pannese M, Migliaccio E, Morelli F, Stornaiuolo A, Nigro V, Simeone A, and Boncinelli E. 1989. The human HOX gene family. Nucleic Acids Res 17:10385-10402.

Akam M. 1989. Hox and HOM: homologous gene clusters in insects and vertebrates. Cell 57:347-349.

Albertson RC, Streelman JT, and Kocher TD. 2003. Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. Proc Natl Acad Sci U S A 100:5252-5257.

Albertson RC, Streelman JT, Kocher TD, and Yelick PC. 2005. Integration and evolution of the cichlid mandible: the molecular basis of alternate feeding strategies. Proc Natl Acad Sci USA 102:16287-16292.

Alfaro ME, Brock CD, Banbury BL, and Wainwright PC. 2009. Does evolutionary innovation in pharyngeal jaws lead to rapid lineage diversification in labrid fishes? BMC Evol Biol 9:255.

Alfaro ME, Santini F, and Brock CD. 2007. Do reefs drive diversification in marine teleosts? Evidence from the pufferfishes and their allies (Order Tetraodontiformes). Evolution 61:2104-2126.

Allendorf FW, Thorgaard GH 1984. Tetraploidy and the evolution of salmonid fish. In: Turner JB (ed) Evolutionary genetics of fish. Plenum Press, New York, pp 1–53.

Altekar G, Dwarkadas S, Huelsenbeck JP, and Ronquist F. 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 20:407-415.

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

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

Badets M, Verneau O. 2009. Hox genes from the Polystomatidae (Platyhelminthes, Monogenea). Int J Parasitol . 39(13):1517-23.

Baugh TM. 1981. Notes on the reproductive behavior of five species of the genus *Fundulus* in aquaria. Journ. of Aquariculture. 2(4): 86-9.

Bellwood DR and Wainwright PC. 2002. The history and biogeography of fishes on coral reefs. pp. 5-32. In: "Coral Reef Fishes. Dynamics and diversity in a complex ecosystem" (P.F. Sale, ed.), Academic Press, San Diego.

Bemis WE, Simons AM. 1997. Coelacanth catches. Science 278:370.

Benton MJ (2005) Vertebrate paleontology, 3rd edn. Blackwell, Malden.

Berezikov E, Guryev V, van de Belt J, Wienholds E, Plasterk RH, and Cuppen E. 2005. Phylogenetic shadowing and computational identification of human microRNA genes. Cell 120:21-24.

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, and Hirschhorn JN. 2004. Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet 74:1111-1120.

Bianchi ME, Beltrame M. 2000. Upwardly mobile proteins. Workshop: the role of HMG proteins in chromatin structure, gene expression and neoplasia. EMBO Rep 1:109-114.

Biemont C, Vieira C. 2006. Genetics: junk DNA as an evolutionary force. Nature 443:521-524.

Birstein, VJ and E Bemis, 1997 How many species are there within the genus Acipenser? pp. 157–163 in Sturgeon Biodiversity and Conservation, edited by VJ Birstein, JR Waldman, and WE Bemis. Kluwer Academic Publishing, Dordrecht, The Netherlands.

Blomme T, Vandepoele K, De Bodt S, Simillion C, Maere S, and Van de Peer Y. 2006. The gain and loss of genes during 600 million years of vertebrate evolution. Genome Biol 7:R43.

Braasch I, Salzburger W, and Meyer A. 2006. Asymmetric evolution in two fish-specifically duplicated receptor tyrosine kinase paralogons involved in teleost coloration. Mol Biol Evol 23:1192-1202.

Braasch I, Brunet F, Volff JN, and Schartl M. 2009. Pigmentation pathway evolution after whole-genome duplication in fish. Genome Biol Evol 1:479-493.

Bradley LC, Snape A, Bhatt S, and Wilkinson DG. 1993. The structure and expression of the Xenopus Krox-20 gene: conserved and divergent patterns of expression in rhombomeres and neural crest. Mech Dev 40:73-84.

Brainerd EL, Liem KF, and Samper CT. 1989. Air Ventilation by Recoil Aspiration in Polypterid Fishes. Science 246:1593-1595.

Britten RJ, Davidson EH. 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Q Rev Biol 46:111-138.

Brunet FG, Crollius HR, Paris M, Aury JM, Gibert P, Jaillon O, Laudet V, and Robinson-Rechavi M. 2006. Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. Mol Biol Evol 23:1808-1816.

Burglin TR, Ruvkun G. 1993. The Caenorhabditis elegans homeobox gene cluster. Curr Opin Genet Dev 3:615-620.

Cameron RA, Rowen L, Nesbitt R, Bloom S, Rast JP, Berney K, Arenas-Mena C, Martinez P, Lucas S, Richardson PM, Davidson EH, Peterson KJ, and Hood L. 2006. Unusual gene order and organization of the sea urchin hox cluster. J Exp Zool B Mol Dev Evol 306:45-58.

Carlson CM, Dupuy AJ, Fritz S, Roberg-Perez KJ, Fletcher CF, and Largaespada DA. 2003. Transposon mutagenesis of the mouse germline. Genetics 165:243-256.

Carrasco AE, McGinnis W, Gehring WJ, and De Robertis EM. 1984. Cloning of an X. laevis gene expressed during early embryogenesis coding for a peptide region homologous to Drosophila homeotic genes. Cell 37:409-414.

Chambers KE, McDaniell R, Raincrow JD, Deshmukh M, Stadler PF, and Chiu CH. 2009. Hox cluster duplication in the basal teleost Hiodon alosoides (Osteoglossomorpha). Theory Biosci 128:109-120.

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

Chiu CH, Amemiya C, Dewar K, Kim CB, Ruddle FH, and Wagner GP. 2002. Molecular evolution of the HoxA cluster in the three major gnathostome lineages. Proc Natl Acad Sci U S A 99:5492-5497.

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

Chopra VS, Mishra RK. 2006. "Mir"acles in hox gene regulation. Bioessays 28:445-448.

Christoffels A, Koh EG, Chia JM, Brenner S, Aparicio S, and Venkatesh B. 2004. Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. Mol Biol Evol 21:1146-1151.

Cook CE, Jimenez E, Akam M, and Salo E. 2004. The Hox gene complement of acoel flatworms, a basal bilaterian clade. Evol Dev 6:154-163.

Crow KD, Stadler PF, Lynch VJ, Amemiya C, and Wagner GP. 2006. The "fish-specific" Hox cluster duplication is coincident with the origin of teleosts. Mol Biol Evol 23:121-136.

Danke J, Miyake T, Powers T, Schein J, Shin H, Bosdet I, Erdmann M, Caldwell R, and Amemiya CT. 2004. Genome resource for the Indonesian coelacanth, Latimeria menadoensis. J Exp Zoolog A Comp Exp Biol 301:228-234.

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

Degnan BM, Morse DE. 1993. Identification of eight homeobox-containing transcripts expressed during larval development and at metamorphosis in the gastropod mollusc Haliotis rufescens. Mol Mar Biol Biotechnol 2:1-9.

DeVries AL, Wohlschlag DE. 1969. Freezing resistance in some Antarctic fishes. Science 163:1073-1075.

Dingerkus G, Howell WM. 1976. Karyotypic analysis and evidence of tetraploidy in the North American paddlefish, Polyodon spathula. Science 194:842-844.

Di-Poi N, Montoya-Burgos JI, and Duboule D. 2009. Atypical relaxation of structural constraints in Hox gene clusters of the green anole lizard. Genome Res 19:602-610.

Di-Poi N, Montoya-Burgos JI, Miller H, Pourquie O, Milinkovitch MC, and Duboule D. 2010. Changes in Hox genes' structure and function during the evolution of the squamate body plan. Nature 464:99-103.

Doak TG, Doerder FP, Jahn CL, and Herrick G. 1994. A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. Proc Natl Acad Sci U S A 91:942-946.

Doerksen LF, Bhattacharya A, Kannan P, Pratt D, and Tainsky MA. 1996. Functional interaction between a RARE and an AP-2 binding site in the regulation of the human HOX A4 gene promoter. Nucleic Acids Res 24:2849-2856.

Dolecki GJ, Wang G, and Humphreys T. 1988. Stage- and tissue-specific expression of two homeo box genes in sea urchin embryos and adults. Nucleic Acids Res 16:11543-11558.

Donoghue PC, Purnell MA. 2005. Genome duplication, extinction and vertebrate evolution. Trends Ecol Evol 20:312-319.

Duboule D. 2007. The rise and fall of Hox gene clusters. Development 134:2549-2560.

Duboule D, Dolle P. 1989. The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes. EMBO J 8:1497-1505.

Duret L, Marais G, and Biemont C. 2000. Transposons but not retrotransposons are located preferentially in regions of high recombination rate in Caenorhabditis elegans. Genetics 156:1661-1669.

Eide D, Anderson P. 1988. Insertion and excision of Caenorhabditis elegans transposable element Tc1. Mol Cell Biol 8:737-746.

Eide D, Anderson P. 1985. Transposition of Tc1 in the nematode Caenorhabditis elegans. Proc Natl Acad Sci U S A 82:1756-1760.

Emmons SW, Roberts S, and Ruan KS. 1986. Evidence in a nematode for regulation of transposon excision by tissue-specific factors. Mol Gen Genet 202:410-415.

Emmons SW, Yesner L, Ruan KS, and Katzenberg D. 1983. Evidence for a transposon in Caenorhabditis elegans. Cell 32:55-65.

Fayet O, Ramond P, Polard P, Prere MF, and Chandler M. 1990. Functional similarities between retroviruses and the IS3 family of bacterial insertion sequences? Mol Microbiol 4:1771-1777.

Ferrier DE, Holland PW. 2002. Ciona intestinalis ParaHox genes: evolution of Hox/ParaHox cluster integrity, developmental mode, and temporal colinearity. Mol Phylogenet Evol 24:412-417.

Ferrier DE, Minguillon C. 2003. Evolution of the Hox/ParaHox gene clusters. Int J Dev Biol 47:605-611.

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

Fischer SE, van Luenen HG, and Plasterk RH. 1999. Cis requirements for transposition of Tc1-like transposons in C. elegans. Mol Gen Genet 262:268-274.

Fischer SEJ, Wienholds E, and Plasterk RHA. 2001. Regulated transposition of a fish transposon in the mouse germ line. Proceedings of the National Academy of Sciences of the United States of America 98:6759-6764.

Fisher S, Grice EA, Vinton RM, Bessling SL, and McCallion AS. 2006. Conservation of RET regulatory function from human to zebrafish without sequence similarity. Science 312:276-279.

Fontana F, Colombo G. 1974. The chromosomes of Italian sturgeons. Experientia 30:739-742.

Fontana F, Congiu L, Mudrak VA, Quattro JM, Smith TI, Ware K, and Doroshov SI. 2008. Evidence of hexaploid karyotype in shortnose sturgeon. Genome 51:113-119.

Freckleton G, Lippman SI, Broach JR, and Tavazoie S. 2009. Microarray profiling of phage-display selections for rapid mapping of transcription factor-DNA interactions. PLoS Genet 5:e1000449.

Fried C, Prohaska SJ, and Stadler PF. 2004. Exclusion of repetitive DNA elements from gnathostome Hox clusters. J Exp Zool B Mol Dev Evol 302:165-173.

Friedman M. 2010. Explosive morphological diversification of spiny-finned teleost fishes in the aftermath of the end-Cretaceous extinction. Proc Biol Sci .

Galas DJ, Schmitz A. 1978. DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res 5:3157-3170.

Garcia-Fernandez J. 2005. The genesis and evolution of homeobox gene clusters. Nat Rev Genet 6:881-892.

Garcia-Fernandez J, Holland PW. 1994. Archetypal organization of the amphioxus Hox gene cluster. Nature 370:563-566.

Gaunt SJ, Krumlauf R, and Duboule D. 1989. Mouse homeo-genes within a subfamily, Hox-1.4, -2.6 and -5.1, display similar anteroposterior domains of expression in the embryo, but show stage- and tissue-dependent differences in their regulation. Development 107:131-141.

Gaut BS, Weir BS. 1994. Detecting substitution-rate heterogeneity among regions of a nucleotide sequence. Mol Biol Evol 11:620-629.

Giraldo R, Diaz-Orejas R. 2001. Similarities between the DNA replication initiators of Gram-negative bacteria plasmids (RepA) and eukaryotes (Orc4p)/archaea (Cdc6p). Proc Natl Acad Sci U S A 98:4938-4943.

Goda M, Fujii R. 1995. Blue Chromatophores in Two Species of Callionymid Fish. Zool Sci 12:811-813.

Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, and Carroll SB. 2005. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature 433:481-487.

Grenier JK, Garber TL, Warren R, Whitington PM, and Carroll S. 1997. Evolution of the entire arthropod Hox gene set predated the origin and radiation of the onychophoran/arthropod clade. Curr Biol 7:547-553.

Hacker U, Kaufmann E, Hartmann C, Jurgens G, Knochel W, and Jackle H. 1995. The Drosophila fork head domain protein crocodile is required for the establishment of head structures. EMBO J 14:5306-5317.

Hafen E, Levine M, and Gehring WJ. 1984. Regulation of Antennapedia transcript distribution by the bithorax complex in Drosophila. Nature 307:287-289.

Hahn MW. 2007. Detecting natural selection on cis-regulatory DNA. Genetica 129:7-18.

Hahn MW, Rockman MV, Soranzo N, Goldstein DB, and Wray GA. 2004. Population genetic and phylogenetic evidence for positive selection on regulatory mutations at the factor VII locus in humans. Genetics 167:867-877.

Hallstrom BM, Janke A. 2009. Gnathostome phylogenomics utilizing lungfish EST sequences. Mol Biol Evol 26:463-471.

Hardison RC. 2000. Conserved noncoding sequences are reliable guides to regulatory elements. Trends Genet 16:369-372.

Harper RW. 1992. Behavior of *Fundulus chrysotus* and *Fundulus olivaceus* in a Captive Environment. American Currents.

Harris LJ, Baillie DL, and Rose AM. 1988. Sequence identity between an inverted repeat family of transposable elements in Drosophila and Caenorhabditis. Nucleic Acids Res 16:5991-5998.

Hart CP, Awgulewitsch A, Fainsod A, McGinnis W, and Ruddle FH. 1985. Homeo box gene complex on mouse chromosome 11: molecular cloning, expression in embryogenesis, and homology to a human homeo box locus. Cell 43:9-18.

Hasegawa M, Kishino H, and Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160-174.

Helfman GS, Collette BB and Facey DE. 1997. The Diversity of Fishes. Blackwell Science, Inc. (Blackwell Publishing company), Massachusetts.

Hewetson A, Hendrix EC, Mansharamani M, Lee VH, and Chilton BS. 2002. Identification of the RUSH consensus-binding site by cyclic amplification and selection of targets: demonstration that RUSH mediates the ability of prolactin to augment progesterone-dependent gene expression. Mol Endocrinol 16:2101-2112.

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

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

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

Holland PW, Holland LZ, Williams NA, and Holland ND. 1992. An amphioxus homeobox gene: sequence conservation, spatial expression during development and insights into vertebrate evolution. Development 116:653-661.

Hughes AL, Nei M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167-170.

Huh JW, Kim YH, Lee SR, Kim H, Kim DS, Kim HS, Kang HS, and Chang KT. 2009. Gain of new exons and promoters by lineage-specific transposable elements-integration and conservation event on CHRM3 gene. Mol Cells 28:111-117.

Hulsey CD, Garcia de Leon FJ, and Rodiles-Hernandez R. 2006. Micro- and macroevolutionary decoupling of cichlid jaws: a test of Liem's key innovation hypothesis. Evolution 60:2096-2109.

Hunt P, Gulisano M, Cook M, Sham MH, Faiella A, Wilkinson D, Boncinelli E, and Krumlauf R. 1991. A distinct Hox code for the branchial region of the vertebrate head. Nature 353:861-864.

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

Inoue JG, Miya M, Tsukamoto K, and Nishida M. 2003. Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the "ancient fish". Mol Phylogenet Evol 26:110-120.

Ivics Z, Izsvak Z, Minter A, and Hackett PB. 1996. Identification of functional domains and evolution of Tc1-like transposable elements. Proc Natl Acad Sci U S A 93:5008-5013.

Jacob F, Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3:318-356.

Jacobson JW, Medhora MM, and Hartl DL. 1986. Molecular structure of a somatically unstable transposable element in Drosophila. Proc Natl Acad Sci U S A 83:8684-8688.

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

Jamieson AJ, Fujii T, Solan M, Matsumoto AK, Bagley PM, and Priede IG. 2009. Liparid and macrourid fishes of the hadal zone: in situ observations of activity and feeding behaviour. Proc Biol Sci 276:1037-1045.

Janvier P. 1996. Early Vertebrates. Clarendon (Oxford University Press), New York.

Jimenez-Guri E, Paps J, Garcia-Fernandez J, and Salo E. 2006. Hox and ParaHox genes in Nemertodermatida, a basal bilaterian clade. Int J Dev Biol 50:675-679.

Johnston CR, O'Dushlaine C, Fitzpatrick DA, Edwards RJ, and Shields DC. 2007. Evaluation of Whether Accelerated Protein Evolution in Chordates Has Occurred before, after, or Simultaneously with Gene Duplication. Mol Biol Evol 24:315-323.

Kaufman TC, Lewis R, and Wakimoto B. 1980. Cytogenetic Analysis of Chromosome 3 in DROSOPHILA MELANOGASTER: The Homoeotic Gene Complex in Polytene Chromosome Interval 84a-B. Genetics 94:115-133.

Keightley PD, Kryukov GV, Sunyaev S, Halligan DL, and Gaffney DJ. 2005. Evolutionary constraints in conserved nongenic sequences of mammals. Genome Res 15:1373-1378.

Keynes R, Krumlauf R. 1994. Hox genes and regionalization of the nervous system. Annu Rev Neurosci 17:109-132.

Kikugawa K, Katoh K, Kuraku S, Sakurai H, Ishida O, Iwabe N, and Miyata T. 2004. Basal jawed vertebrate phylogeny inferred from multiple nuclear DNA-coded genes. BMC Biol 2:3.

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

Kim DS, Kim TH, Huh JW, Kim IC, Kim SW, Park HS, and Kim HS. 2006. LINE FUSION GENES: a database of LINE expression in human genes. BMC Genomics 7:139.

Kmita M, Duboule D. 2003. Organizing axes in time and space; 25 years of colinear tinkering. Science 301:331-333.

Kmita-Cunisse M, Loosli F, Bierne J, and Gehring WJ. 1998. Homeobox genes in the ribbonworm Lineus sanguineus: evolutionary implications. Proc Natl Acad Sci U S A 95:3030-3035.

Koblmuller S, Schliewen UK, Duftner N, Sefc KM, Katongo C, and Sturmbauer C. 2008. Age and spread of the haplochromine cichlid fishes in Africa. Mol Phylogenet Evol 49:153-169.

Koh EG, Lam K, Christoffels A, Erdmann MV, Brenner S, and Venkatesh B. 2003. Hox gene clusters in the Indonesian coelacanth, Latimeria menadoensis. Proc Natl Acad Sci U S A 100:1084-1088.

Kottelat M, Britz R, Hui TH, and Witte KE. 2006. Paedocypris, a new genus of Southeast Asian cyprinid fish with a remarkable sexual dimorphism, comprises the world's smallest vertebrate. Proc Biol Sci 273:895-899.

Krasnov A, Koskinen H, Afanasyev S, and Molsa H. 2005. Transcribed Tc1-like transposons in salmonid fish. BMC Genomics 6:107.

Kristie TM, Roizman B. 1986. Alpha 4, the major regulatory protein of herpes simplex virus type 1, is stably and specifically associated with promoter-regulatory domains of alpha genes and of selected other viral genes. Proc Natl Acad Sci U S A 83:3218-3222.

Krumlauf R. 1993. Hox genes and pattern formation in the branchial region of the vertebrate head. Trends Genet 9:106-112.

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

Kurosawa G, Yamada K, Ishiguro H, and Hori H. 1999. Hox gene complexity in medaka fish may Be similar to that in pufferfish rather than zebrafish. Biochem Biophys Res Commun 260:66-70.

Kuziora MA, McGinnis W. 1988. Autoregulation of a Drosophila homeotic selector gene. Cell 55:477-485.

Lam WL, Seo P, Robison K, Virk S, and Gilbert W. 1996. Discovery of amphibian Tc1-like transposon families. J Mol Biol 257:359-366.

Lanave C, Preparata G, Saccone C, and Serio G. 1984. A new method for calculating evolutionary substitution rates. J Mol Evol 20:86-93.

Landan G, Graur D. 2007. Heads or tails: a simple reliability check for multiple sequence alignments. Mol Biol Evol 24:1380-1383.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, and International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. Nature 409:860-921.

Langin T, Capy P, and Daboussi MJ. 1995. The transposable element impala, a fungal member of the Tc1-mariner superfamily. Mol Gen Genet 246:19-28.

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

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

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

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

Li H, Zhang Y, Zhang YP, and Fu YX. 2003. Neutrality tests using DNA polymorphism from multiple samples. Genetics 163:1147-1151.

Liao LW, Rosenzweig B, and Hirsh D. 1983. Analysis of a transposable element in Caenorhabditis elegans. Proc Natl Acad Sci U S A 80:3585-3589.

Liem KF. 1973. Evolutionary Strategies and Morphological Innovations: Cichlid Pharyngeal Jaws. Syst Biol 22:425-441.

Liu G, Geurts AM, Yae K, Srinivasan AR, Fahrenkrug SC, Largaespada DA, Takeda J, Horie K, Olson WK, and Hackett PB. 2005. Target-site preferences of Sleeping Beauty transposons. J Mol Biol 346:161-173.

Loots GG, Ovcharenko I. 2004. rVISTA 2.0: evolutionary analysis of transcription factor binding sites. Nucleic Acids Res 32:W217-21.

Ludwig A, Belfiore NM, Pitra C, Svirsky V, and Jenneckens I. 2001. Genome duplication events and functional reduction of ploidy levels in sturgeon (Acipenser, Huso and Scaphirhynchus). Genetics 158:1203-1215.

Lufkin T. 1996. Transcriptional control of Hox genes in the vertebrate nervous system. Curr Opin Genet Dev 6:575-580.

Luo G, Ivics Z, Izsvak Z, and Bradley A. 1998. Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. Proc Natl Acad Sci U S A 95:10769-10773.

Lynch M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. Proc Natl Acad Sci U S A 104 Suppl 1:8597-8604.

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

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

Maddison WP, Maddison DR. 1989. Interactive analysis of phylogeny and character evolution using the computer program MacClade. Folia Primatol (Basel) 53:190-202.

Malicki J, Bogarad LD, Martin MM, Ruddle FH, and McGinnis W. 1993. Functional analysis of the mouse homeobox gene HoxB9 in Drosophila development. Mech Dev 42:139-150.

Mansharamani M, Hewetson A, and Chilton BS. 2001. Cloning and characterization of an atypical Type IV P-type ATPase that binds to the RING motif of RUSH transcription factors. J Biol Chem 276:3641-3649.

Martin AP, Palumbi SR. 1993. Body size, metabolic rate, generation time, and the molecular clock. Proceedings of the National Academy of Sciences of the United States of America 90:4087-4091.

Martin A. 2001. The phylogenetic placement of chondrichthyes: inferences from analysis of multiple genes and implications for comparative studies. Genetica 111:349-357.

Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Munch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, and Wingender E. 2003. TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res 31:374-378.

McCann, C. 1961. "The Sunfish (Mola mola) in New Zealand Waters." Records of the Dominion Museum. Vol. 4 (2).

McClinctock B. 1956. Controlling elements and the gene. Cold Spring Harb Symp Quant Biol 21:197-216.

McGinnis N, Kuziora MA, and McGinnis W. 1990. Human Hox-4.2 and Drosophila deformed encode similar regulatory specificities in Drosophila embryos and larvae. Cell 63:969-976.

McGinnis W, Garber RL, Wirz J, Kuroiwa A, and Gehring WJ. 1984. A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. Cell 37:403-408.

McGinnis W, Hart CP, Gehring WJ, and Ruddle FH. 1984. Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of Drosophila. Cell 38:675-680.
McGinnis W, Krumlauf R. 1992. Homeobox genes and axial patterning. Cell 68:283-302.
McGinnis W, Levine MS, Hafen E, Kuroiwa A, and Gehring WJ. 1984. A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes. Nature 308:428-433.
Meyer A, Malaga-Trillo E. 1999. Vertebrate genomics: More fishy tales about Hox genes. Curr Biol 9:R210-3.
Meyer A, Schartl M. 1999. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. Curr Opin Cell Biol 11:699-704.
Meyer A, Van de Peer Y. 2005. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). Bioessays 27:937-945.
Mitchell PJ, Timmons PM, Hebert JM, Rigby PW, and Tjian R. 1991. Transcription factor AP-2 is expressed in neural crest cell lineages during mouse embryogenesis. Genes Dev 5:105-119.
Mitsouras K, Wong B, Arayata C, Johnson RC, and Carey M. 2002. The DNA architectural protein HMGB1 displays two distinct modes of action that promote enhanceosome assembly. Mol Cell Biol 22:4390-4401.
Moghadam HK, Ferguson MM, and Danzmann RG. 2005. Evolution of Hox clusters in Salmonidae: a comparative analysis between Atlantic salmon (Salmo salar) and rainbow trout (Oncorhynchus mykiss). J Mol Evol 61:636-649.
Monod J, Jacob F. 1961. Teleonomic mechanisms in cellular metabolism, growth, and differentiation. Cold Spring Harb Symp Quant Biol 26:389-401.
Monteiro AS, Ferrier DE. 2006. Hox genes are not always Colinear. Int J Biol Sci 2:95-103.
Mori I, Benian GM, Moerman DG, and Waterston RH. 1988. Transposable element Tc1 of Caenorhabditis elegans recognizes specific target sequences for integration. Proc Natl Acad Sci U S A 85:861-864.
Morrison A, Moroni MC, Ariza-McNaughton L, Krumlauf R, and Mavilio F. 1996. In vitro and transgenic analysis of a human HOXD4 retinoid-responsive enhancer. Development 122:1895-1907.
Moses AM, Chiang DY, Pollard DA, Iyer VN, and Eisen MB. 2004. MONKEY: identifying conserved transcription-factor binding sites in multiple alignments using a binding site-specific evolutionary model. Genome Biol 5:R98.
Mulley JF, Chiu CH, and Holland PW. 2006. Breakup of a homeobox cluster after genome duplication in teleosts. Proc Natl Acad Sci U S A 103:10369-10372.
Murtha MT, Leckman JF, and Ruddle FH. 1991. Detection of homeobox genes in development and evolution. Proc Natl Acad Sci U S A 88:10711-10715.
Naruse K, Tanaka M, Mita K, Shima A, Postlethwait J, and Mitani H. 2004. A medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping. Genome Res 14:820-828.
Neduva V, Russell RB. 2005. Linear motifs: evolutionary interaction switches. FEBS Lett 579:3342-3345.
Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418-426.
Nieto MA, Bradley LC, and Wilkinson DG. 1991. Conserved segmental expression of Krox-20 in the vertebrate hindbrain and its relationship to lineage restriction. Development Suppl 2:59-62.
Nelson GJ, 1969. Gill arches and the phylogeny of fishes, with notes on the classification of vertebrates. Bull. Am. Mus. Nat. Hist. 141 (1969), pp. 475–552.
Nelson, JS 1994. Fishes of the World, 3rd edition. New York: John Wiley & Sons
Nelson, JS Fishes of the World (Wiley, 2006)
Nonchev S, Vesque C, Maconochie M, Seitanidou T, Ariza-McNaughton L, Frain M, Marshall H, Sham MH, Krumlauf R, and Charnay P. 1996. Segmental expression of Hoxa-2 in the hindbrain is directly regulated by Krox-20. Development 122:543-554.
Nylander, J. A. A. 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.
Ohno S, Muramoto J, Stenius C, Christian L, Kittrell WA, and Atkin NB. 1969. Microchromosomes in holocephalian, chondrostean and holostean fishes. Chromosoma 26:35-40.
Ohno, S. 1970. Evolution by gene duplication. Springer-Verlag, New York.
Oxtoby E, Jowett T. 1993. Cloning of the zebrafish krox-20 gene (krx-20) and its expression during hindbrain development. Nucleic Acids Res 21:1087-1095.

Pang KC, Stephen S, Engstrom PG, Tajul-Arifin K, Chen W, Wahlestedt C, Lenhard B, Hayashizaki Y, and Mattick JS. 2005. RNAdb--a comprehensive mammalian noncoding RNA database. Nucleic Acids Res 33:D125-30.

Papillon D, Perez Y, Fasano L, Le Parco Y, and Caubit X. 2003. Hox gene survey in the chaetognath Spadella cephaloptera: evolutionary implications. Dev Genes Evol 213:142-148.

Passamaneck YJ, Halanych KM. 2004. Evidence from Hox genes that bryozoans are lophotrochozoans. Evol Dev 6:275-281.

Patterson C. (1982) Morphology and interrelationships of primitive actinopterygian fishes. American Zoologist 22, 241–259.

Pedersen JS, Bejerano G, Siepel A, Rosenbloom K, Lindblad-Toh K, Lander ES, Kent J, Miller W, and Haussler D. 2006. Identification and classification of conserved RNA secondary structures in the human genome. PLoS Comput Biol 2:e33.

Peifer M, Karch F, and Bender W. 1987. The bithorax complex: control of segmental identity. Genes & Dev 1:891-898.

Peterson KJ. 2004. Isolation of Hox and Parahox genes in the hemichordate Ptychodera flava and the evolution of deuterostome Hox genes. Mol Phylogenet Evol 31:1208-1215.

Petifils, L. E. 1986. Ecological relationships among members of the *F. notatus* species complex. Master's thesis, Univ. New Orleans, New Orleans, LA.

Pierce RJ, Wu W, Hirai H, Ivens A, Murphy LD, Noel C, Johnston DA, Artiguenave F, Adams M, Cornette J, Viscogliosi E, Capron M, and Balavoine G. 2005. Evidence for a dispersed Hox gene cluster in the platyhelminth parasite Schistosoma mansoni. Mol Biol Evol 22:2491-2503.

Plasterk RH, Izsvak Z, and Ivics Z. 1999. Resident aliens: the Tc1/mariner superfamily of transposable elements. Trends Genet 15:326-332.

Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817-818.

Postlethwait J, Amores A, Cresko W, Singer A, and Yan YL. 2004. Subfunction partitioning, the teleost radiation and the annotation of the human genome. Trends Genet 20:481-490.

Powers TP, Hogan J, Ke Z, Dymbrowski K, Wang X, Collins FH, and Kaufman TC. 2000. Characterization of the Hox cluster from the mosquito Anopheles gambiae (Diptera: Culicidae). Evol Dev 2:311-325.

Prince V, Lumsden A. 1994. Hoxa-2 expression in normal and transposed rhombomeres: independent regulation in the neural tube and neural crest. Development 120:911-923.

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

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

Prohaska SJ, Fried C, Amemiya CT, Ruddle FH, Wagner GP, and Stadler PF. 2004. The shark HoxN cluster is homologous to the human HoxD cluster. J Mol Evol 58:212-217.

Prohaska SJ, Fried C, Flamm C, Wagner GP, and Stadler PF. 2004. Surveying phylogenetic footprints in large gene clusters: applications to Hox cluster duplications. Mol Phylogenet Evol 31:581-604.

Radice AD, Bugaj B, Fitch DH, and Emmons SW. 1994. Widespread occurrence of the Tc1 transposon family: Tc1-like transposons from teleost fish. Mol Gen Genet 244:606-612.

Raftery AE. 1986a. Choosing models for cross-classification. Amer. Socio. Rev. 51:145-146.

Raftery AE. 1986b. A note on Bayes factors for log-linear contingency table models with vague prior information. J. Royal Stat. Soc. Series B. 48: 249-250.

Ravi V, Venkatesh B. 2008. Rapidly evolving fish genomes and teleost diversity. Curr Opin Genet Dev 18:544-550.

Richardson MK, Crooijmans RP, and Groenen MA. 2007. Sequencing and genomic annotation of the chicken (Gallus gallus) Hox clusters, and mapping of evolutionarily conserved regions. Cytogenet Genome Res 117:110-119.

Riddle DL, Blumenthal T, Meyer BJ, and Priess JR. 1997. C. Elegans II. CSHL press. 106.

Rizzon C, Martin E, Marais G, Duret L, Segalat L, and Biemont C. 2003. Patterns of selection against transposons inferred from the distribution of Tc1, Tc3 and Tc5 insertions in the mut-7 line of the nematode Caenorhabditis elegans. Genetics 165:1127-1135.

Robinson-Rechavi M, Laudet V. 2001. Evolutionary rates of duplicate genes in fish and mammals. Mol Biol Evol 18:681-683.

Robison BH, Reisenbichler KR. 2008. Macropinna microstoma and the Paradox of Its Tubular Eyes. Copeia 2008:780-784.
Rockman MV, Hahn MW, Soranzo N, Zimprich F, Goldstein DB, and Wray GA. 2005. Ancient and recent positive selection transformed opioid cis-regulation in humans. PLoS Biol 3:e387.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.
Rose AM, Snutch TP. 1984. Isolation of the closed circular form of the transposable element Tc1 in Caenorhabditis elegans. Nature 311:485-486.
Rosenzweig B, Liao LW, and Hirsh D. 1983. Target sequences for the C. elegans transposable element Tc1. Nucleic Acids Res 11:7137-7140.
Rozas J, Sanchez-DelBarrio JC, Messeguer X, and Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496-2497.
Ruan KS, Emmons SW. 1987. Precise and imprecise somatic excision of the transposon Tc1 in the nematode C. elegans. Nucleic Acids Res 15:6875-6881.
Ryynanen HJ, Primmer CR. 2006. Varying signals of the effects of natural selection during teleost growth hormone gene evolution. Genome 49:42-53.
Sadl VS, Sing A, Mar L, Jin F, and Cordes SP. 2003. Analysis of hindbrain patterning defects caused by the kreisler(enu) mutation reveals multiple roles of Kreisler in hindbrain segmentation. Dev Dyn 227:134-142.
Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425.
Sanchez-Herrero E, Vernos I, Marco R, and Morata G. 1985. Genetic organization of Drosophila bithorax complex. Nature 313:108-113.
Sandelin A, Alkema W, Engstrom P, Wasserman WW, and Lenhard B. 2004. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res 32:D91-4.
Santini F, Harmon LJ, Carnevale G, and Alfaro ME. 2009. Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. BMC Evol Biol 9:194.
Santini S, Bernardi G. 2005. Organization and base composition of tilapia Hox genes: implications for the evolution of Hox clusters in fish. Gene 346:51-61.
Santini S, Boore JL, and Meyer A. 2003. Evolutionary conservation of regulatory elements in vertebrate Hox gene clusters. Genome Res 13:1111-1122.
Sato Y, Hashiguchi Y, and Nishida M. 2009. Temporal pattern of loss/persistence of duplicate genes involved in signal transduction and metabolic pathways after teleost-specific genome duplication. BMC Evol Biol 9:127.
Scemama JL, Vernon JL, and Stellwag EJ. 2006. Differential expression of hoxa2a and hoxa2b genes during striped bass embryonic development. Gene Expr Patterns 6:843-848.
Schaeffer B, Rosen De. 1961. Major adaptive levels in the evolution of the actinopterygian feeding mechanism. Amer Zool 1:187-204.
Scott MP. 1993. A rational nomenclature for vertebrate homeobox (HOX) genes. Nucleic Acids Res 21:1687-1688.
Seo HC, Edvardsen RB, Maeland AD, Bjordal M, Jensen MF, Hansen A, Flaat M, Weissenbach J, Lehrach H, Wincker P, Reinhardt R, and Chourrout D. 2004. Hox cluster disintegration with persistent anteroposterior order of expression in Oikopleura dioica. Nature 431:67-71.
Shao H, Tu Z. 2001. Expanding the diversity of the IS630-Tc1-mariner superfamily: discovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons. Genetics 159:1103-1115.
Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jonsson B, Schluter D, and Kingsley DM. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature 428:717-723.
Shashikant CS, Bieberich CJ, Belting HG, Wang JC, Borbely MA, and Ruddle FH. 1995. Regulation of Hoxc-8 during mouse embryonic development: identification and characterization of critical elements involved in early neural tube expression. Development 121:4339-4347.
Shepherd JC, McGinnis W, Carrasco AE, De Robertis EM, and Gehring WJ. 1984. Fly and frog homoeo domains show homologies with yeast mating type regulatory proteins. Nature 310:70-71.

Shippy TD, Ronshaugen M, Cande J, He J, Beeman RW, Levine M, Brown SJ, and Denell RE. 2008. Analysis of the Tribolium homeotic complex: insights into mechanisms constraining insect Hox clusters. Dev Genes Evol 218:127-139.

Sidow A. 1996. Gen(om)e duplications in the evolution of early vertebrates. Curr Opin Genet Dev 6:715-722.

Siegel N, Hoegg S, Salzburger W, Braasch I, and Meyer A. 2007. Comparative genomics of ParaHox clusters of teleost fishes: gene cluster breakup and the retention of gene sets following whole genome duplications. BMC Genomics 8:312.

Sinzelle L, Pollet N, Bigot Y, and Mazabraud A. 2005. Characterization of multiple lineages of Tc1-like elements within the genome of the amphibian Xenopus tropicalis. Gene 349:187-196.

Smith AG, Smith DG, Funnell BM. Atlas of Cenozoic and Mesozoic Coastlines. Cambridge, UK: Cambridge Univ Press; 1994.

Smith TM, Wang X, Zhang W, Kulyk W, and Nazarali AJ. 2009. Hoxa2 plays a direct role in murine palate development. Dev Dyn 238:2364-2373.

Steinke D, Salzburger W, Braasch I, and Meyer A. 2006. Many genes in fish have species-specific asymmetric rates of molecular evolution. BMC Genomics 7:20.

Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tagle DA, Koop BF, Goodman M, Slightom JL, Hess DL, and Jones RT. 1988. Embryonic epsilon and gamma globin genes of a prosimian primate (Galago crassicaudatus). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. J Mol Biol 203:439-455.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512-526.

Thisse C, Thisse B. 2008. High-resolution in situ hybridization to whole-mount zebrafish embryos. Nat Protoc 3:59-69.

Thomerson JE. 1966. A comparative biosystematic study of *Fundulus notatus* and *Fundulus olivaceus* in Southern Illinois. Trans. II. Acad. Sci. 60: 375-9.

Thomerson JE and Wooldridge DP. 1970. Food habits of allotopic and syntopic populations of the topminnows *Fundulus olivaceus* and *Fundulus notatus*. American Midland Naturalist. 84(2): 573-6.

Tournamille C, Colin Y, Cartron JP, and Le Van Kim C. 1995. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. Nat Genet 10:224-228.

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

van Luenen HG, Colloms SD, and Plasterk RH. 1994. The mechanism of transposition of Tc3 in C. elegans. Cell 79:293-301.

Vavouri T, Lehner B. 2009. Conserved noncoding elements and the evolution of animal body plans. Bioessays 31:727-735.

Venkatesh B, Kirkness EF, Loh YH, Halpern AL, Lee AP, Johnson J, Dandona N, Viswanathan LD, Tay A, Venter JC, Strausberg RL, and Brenner S. 2007. Survey sequencing and comparative analysis of the elephant shark (Callorhinchus milii) genome. PLoS Biol 5:e101.

Vogel G. 1998. Doubled genes may explain fish diversity. Science 281:1119,1121.

Volff JN. 2005. Genome evolution and biodiversity in teleost fish. Heredity 94:280-294.

Vos JC, De Baere I, and Plasterk RH. 1996. Transposase is the only nematode protein required for in vitro transposition of Tc1. Genes Dev 10:755-761.

Vos JC, Plasterk RH. 1994. Tc1 transposase of Caenorhabditis elegans is an endonuclease with a bipartite DNA binding domain. EMBO J 13:6125-6132.

Vos JC, van Luenen HG, and Plasterk RH. 1993. Characterization of the Caenorhabditis elegans Tc1 transposase in vivo and in vitro. Genes Dev 7:1244-1253.

Washietl S, Hofacker IL, and Stadler PF. 2005. Fast and reliable prediction of noncoding RNAs. Proc Natl Acad Sci U S A 102:2454-2459.

Wilkinson DG, Bhatt S, Chavrier P, Bravo R, and Charnay P. 1989. Segment-specific expression of a zinc-finger gene in the developing nervous system of the mouse. Nature 337:461-464.

Wittkopp PJ, True JR, and Carroll SB. 2002. Reciprocal functions of the Drosophila yellow and ebony proteins in the development and evolution of pigment patterns. Development 129:1849-1858.

Woods IG, Wilson C, Friedlander B, Chang P, Reyes DK, Nix R, Kelly PD, Chu F, Postlethwait JH, and Talbot WS. 2005. The zebrafish gene map defines ancestral vertebrate chromosomes. Genome Res 15:1307-1314.

Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. Nat Rev Genet 8:206-216.

Yasukochi Y, Ashakumary LA, Wu C, Yoshido A, Nohata J, Mita K, and Sahara K. 2004. Organization of the Hox gene cluster of the silkworm, Bombyx mori: a split of the Hox cluster in a non-Drosophila insect. Dev Genes Evol 214:606-614.

Yekta S, Shih IH, and Bartel DP. 2004. MicroRNA-directed cleavage of HOXB8 mRNA. Science 304:594-596.

Yu W, Rajasegaran V, Yew K, Loh W, Tay B, Amemiya CT, Brenner S, and Venkatesh B. 2008. Elephant shark sequence reveals unique insights into the evolutionary history of vertebrate genes: A comparative analysis of the protocadherin cluster. Proceedings of the National Academy of Sciences 105:3819-3824.

Zappavigna V, Falciola L, Helmer-Citterich M, Mavilio F, and Bianchi ME. 1996. HMG1 interacts with HOX proteins and enhances their DNA binding and transcriptional activation. EMBO J 15:4981-4991.

Zwickl DJ. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation,The University of Texas at Austin.

Curriculum Vitae

Jeremy Don Raincrow

Education:
2010 Ph.D. in Cell and Developmental Biology-Rutgers, The State University of New Jersey and University of Medicine and Dentistry of New Jersey
2004 B.S. in Biology-University of Central Oklahoma

Publications:

Zhang YQ, Guo N, Peng G, Han M, Raincrow J, Chiu CH, Coolen LM, Wenthold RJ, Zhao ZQ, Jing N and Yu L. 2009. Role of SIP30 in the development and maintenance of peripheral nerve injury-induced neuropathic pain. Pain. 146(1-2): 130-40.

Chambers KE, McDaniell R, Raincrow JD, Deshmukh M, Stadler P, and Chiu CH. 2009. Hox cluster duplication in a basal teleost fish Hiodon alosoides (Osteoglossomorpha). Theory Biosci. 128(2). 109-20.

Won YJ, Wang Y, Sivasundar A, Raincrow J, and Hey J. 2006. Nuclear gene variation and molecular dating of the cichlid species flock of Lake Malawi. MBE. 23(4). 828-37