Genomic and phylogenetic assessment of sea lamprey (*Petromyzon marinus*) *Hox* genes and analysis of *Hox* genes in association with myomeres across multiple lamprey genera

By

Darcy Childs

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

In Partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

Department of Biological Science

University of Manitoba

Winnipeg

Abstract

Lampreys are an important model for the study of early vertebrate development due to their unique evolutionary position as one of only two extant jawless vertebrates. In this study, 12 new putative *Hox* gene fragments were identified within the recently available *Petromyzon marinus* (sea lamprey) genome. These and the other previously-identified *Hox* genes were analyzed phylogenetically, which enabled the assignment of many of the new sequences to distinct paralogous gene clusters and showed distinctions between gnathostome and lamprey *Hox* sequences. An examination of *Hox* genes in other lamprey species was conducted using genomic PCR-based detection methods and identified 26 putative *Hox* gene homeobox fragments from multiple *Hox* genes across nine lamprey species. A study of *Hox10* coding sequences in different lamprey species failed to find any correlation with variable numbers of trunk myomeres in lampreys, which suggests that other sequences or factors regulate the number of myomeres in different species.

Acknowledgments

I would like to acknowledge a great number of people for their parts taken in the completion of my MSc thesis. First and foremost, I would like to thank my supervisors Dr. Steve Whyard and Dr. Margaret Docker for their support, guidance and encouragement during my stay in the Department of Biological Sciences.

I would also like to recognize the members of both the Whyard/Docker labs for their help and expertise, which helped me greatly. A special thanks to: George Heath for his collection of lamprey samples at the Hammond Bay Biological Station; Cass Erdelyan for his *in silico* knowledge and expertise; and Craig McFarlane and Erin Spice for their roles in RNA sequencing. I would also like to thank all of the other students and who helped me during field seasons and who helped me finish my work.

I thank my advisory committee: Dr. Mark Belmonte and Dr. Mojgan Rastegar for providing me the guidance and advice needed to help complete my thesis. I would also like to thank the Animal Holding staff for their expertise and assistance.

I greatly acknowledge the financial assistance provided to me through the Faculty of Science Graduate Studentship.

I would lastly like to thank my friends and family for their huge amount of support that they provided me. Especially, I thank Kris Fleet who was encouraging, supportive and always there to help me through.

Glossary and Abreviations

Monophyletic - A group that includes all organisms descended from a single common ancestor

Myomere - A block of segmented skeletal muscle fibres

Orthologous genes - Two genes separated by a speciation event

Paralogous genes - Two genes separated by gene duplication event

Paraphyletic - A group that includes the most recent common ancestor, but not all of its descendents

Synapomorphy - A derived trait that is shared between two or more taxa

MYA - Million years ago

RACE - Rapid amplification of cDNA ends

ML - Maximum likelihood

MP - Maximum parsimony

ORF - Open reading frame

PG - Paralogous group

Table of Contents

List of Tables

Table

List of Figures

Figure

1. Introduction

Lampreys are primitive fishes that are of interest to researchers from both basic and applied research perspectives. They are considered distant relatives to all other vertebrate species, and the study of lampreys helps provide insights into the modern vertebrate body plan (Shimeld and Donoghue 2012). Lampreys are also important for commercial reasons. The sea lamprey (*Petromyzon marinus*) for example, has also attracted considerable interest in North America, as it is an invasive pest species in the Great Lakes that has contributed to the collapse of the Great Lakes fisheries in the 1950s and remains a major problem for commercial fisheries (Smith and Tibbles 1980; Madenjian *et al.* 2008). Of additional research interest, particularly in recent years, is the relationship between lamprey "paired" species. In most genera of lampreys, there are species pairs, in which the larvae of the two species are morphologically similar but the adults can differ considerably. One adult type becomes parasitic, feeding on actinopterygian fishes, while the other adult type becomes nonparasitic and does not feed at all following metamorphosis, living on stored fats during that time (see Docker 2009).

1.1 Biology of Lampreys

1.1.1 Lamprey Taxonomy

Lampreys, along with hagfishes, are ancient jawless fishes known as cyclostomes or agnathans (Shimeld and Donoghue 2012). Lamprey fossil records date back 360 million years ago (mya) and it is estimated that the divergence time between lampreys and jawed vertebrates (gnathostomes) is approximately 500 mya (Gess *et al*. 2006; McEwen *et al*. 2009). Agnathans and gnathostomes belong to the phylum Chordata and subphylum Craniata. Craniates are further divided into three superclasses: Myxinomorphi

(hagfishes); Petromyzontomorphi (lampreys); and Gnathostomata (the jawed vertebrates; Nelson, 2006). Extant lampreys belong to the class Petromyzontida and the order Petromyzontiformes (also known as Hyperoartii; Nelson, 2006). Within the order of Petromyzontiformes, there are 43 recognized species organized into three families of lampreys (Renaud 2011; Mateus *et al.* 2013). These families are Geotriidae and Mordaciidae, which are restricted to the southern hemisphere excluding Africa and Antarctica, and Petromyzontidae, which are restricted to the northern hemisphere (Renaud 2011). The phylogenetic relationship of the three families has not been properly resolved using a cladistic approach, but it is speculated that Geotriidae and Mordaciidae are sister groups based on two synapomorphies during what has been termed as a type of second metamorphosis, in which they transition from immature to mature adults (Renaud 2011). The first **synapomorphy** (see Glossary, page 3, for definitions of newly-introduced, bolded words) is based on the number of radial plates (teeth arranged in an inner circle) in the posterior field into individual cusps (pointed cap of keratin) and the second synapomorphy is based on the number of transverse lingual laminae cusps (Renaud 2011). Molecular data, however, suggest that the three lamprey families are distinct from one another (Lang *et al*. 2009). The Petromyzontidae family comprises 39 of the 43 lamprey species, while the Geotriidae family only consists of one species and Mordaciidae is comprised of three (Renaud 2011).

The phylogenetic relationship among the craniates is still a point of debate (Heimberg *et al.* 2010). Morphological data suggest that lampreys and gnathostomes are **monophyletic**, while hagfishes are basal to this group, which suggests that cyclostomes are **paraphyletic** (Near 2009). Most molecular data, on the other hand, shows that

lampreys and hagfishes are monophyletic (e.g., Mallat and Sullivan 1998; Heimberg *et al.* 2010). Many early studies have examined phylogenetic relationships among lampreys using morphological data (e.g., Gill *et al.* 2003; Monette and Renaud 2005), but in more recent years, mitochondrial *cytochrome B* genetic data has been used to make phylogenetic comparisons (Docker *et al.* 1999; Lang *et al.* 2009). Further expansion of lamprey molecular data to support the phylogenetic relationships between lamprey species, as well as other vertebrates, is essential to better understanding lamprey and vertebrate evolution. Lampreys are excellent models in studying gnathostome evolution because they represent some of the most primitive extant vertebrates and are becoming important in many other areas of investigation (Osório and Retaux 2008), including studies focused on growth and developmental differences among lamprey species, as well as genomic studies with the recent sequencing of the *P. marinus* genome that has been available prior to 2009 (see Section 1.3.6; Smith *et al*. 2013).

1.1.2 Lamprey Life Cycle

In addition to their interesting evolutionary history, lampreys also exhibit a different life cycle compared to most of their distant fish relatives. They undergo a metamorphosis event within their lifetime, which effectively divides their lifespan into three main stages of development. The first stage of life is the larval or ammocoete stage, which begins when they hatch from eggs, taking approximately 10-15 days under ideal conditions, into a river and emerge as prolarvae (Piavis 1961; Langille and Hall 1988; Tahara 1988; Renaud 2011). The emerging prolarvae live off nutritional reserves from their yolk sacs for approximately 30 days and during this time, they burrow into sediment beds in rivers and streams (Piavis 1961). Once the gut develops, usually one month post-

hatch, lampreys become filter-feeding larvae known as ammocoetes (Hill and Potter 1970). Ammocoetes emerge only briefly, usually at night, to seek out more suitable environments, often downstream, when either food is limited or water quality diminishes (Potter 1980). After sufficient growth in length and mass, and once lipid stores within the ammocoete reach an appropriate level, the ammocoetes undergo metamorphosis into juvenile lampreys (Youson 1997). The second stage or metamorphosis of the lamprey from ammocoete to juvenile typically takes approximately 3-4 months, during which time the animal develops many different external and internal features (Richardson *et al*. 2010). Externally, the lamprey's transverse lip transforms into an oral disc, teeth and tongue develop, the rudimentary eye of the ammocoete changes into a functional eye, distinct dorsal fins form, and the seven triangular branchiopores change to an oval shape (Richardson *et al*. 2010). Internal changes include metamorphosis of the endostyle into thyroid follicles, the respiratory and digestive systems undergo further development, the larval kidney regresses while the adult kidney develops, and the branchial skeleton undergoes structural modifications (Youson 1997). The third and final stage is the adult stage, which consists of the sexually immature and sexually mature lamprey.

1.1.3 Paired Species

Lamprey species can exhibit two different distinct types of lifestyles. One is a non-parasitic lifestyle, where the ammocoetes metamorphose into a non-feeding, sexually-mature adult. The second type is a parasitic lifestyle, where the lamprey undergoes metamorphosis into an immature adult that feeds on fish for several months to several years, depending on the species and environmental conditions (Docker 2009). Lampreys feed by attaching to a prey fish using their oral disc. Once the lamprey is

attached, it will rasp away flesh using a tooth-bearing tongue and will feed on the blood, lymph, and other tissues (Farmer 1980). Following the feeding phase, the parasitic lamprey will undergo sexual maturation, and at this point, the lifestyles of the two different types of lamprey converge. Sexually mature lampreys migrate up streams from either downstream, lake or ocean, to find spawning sites (Docker 2009). At an appropriate spawning site, the lamprey will construct a nest of silt and small pebbles, moving large stones and pebbles to the boundary of the nest, using their oral discs to grasp and move the debris (Manion and Hanson 1980). At the completion of the nest building and the commencement of mating, the male will attach himself to the top of the female's head using his oral disc, and proceed to wrap around the female, squeezing her to extrude her eggs. During the process, the male releases his sperm to fertilize the eggs. The adults will fan the substrate with their posterior fins to cover the eggs, and being semelparous, they will die soon after spawning (Manion and Hanson 1980).

Interestingly, in seven of the ten genera of lampreys, there are pairs of very closely-related species that exhibit different lifecycles, with one species being parasitic and the other being non-parasitic (Docker 2009). Some of these so-called paired species are so closely related to each other that no genetic distinguising features have yet been found from one another (Docker 2009; Docker et al. 2012) despite the obvious phenotypic differences in the adults (described below). In some cases, a genus can have one parasitic species and several non-parasitic species, and these are called satellite species (Vladykov and Kott 1980). The paired non-parasitic lampreys typically have a longer larval development period in comparison to the parasitic lampreys to compensate

for the shorter, non-trophic post-larval life. Thus, it has been suggested that there is relatively no overall change to the life span between the paired species (Docker 2009).

The paired species relationship is thought to mark an evolutionarily recent speciation event, of which the non-parasitic species arose from the parasitic counterpart (Docker 2009). The observation that some pairs are genetically distinct where others are not is thought to reflect the amount of time since the occurrence of the speciation event (Docker 2009). For example, *Lampetra richardsoni* (western brook lamprey) and *Lampetra ayresii* (North American river lamprey) from British Columbia are genetically indistinguishable within the study range (Docker et al. 1999), although there may be diagnostic differences elsewhere in their range (Boguski *et al.* 2012). *Ichthyomyzon unicuspis* (silver lamprey) and *I. fossor* (northern brook lamprey) show no identifiable differences, despite examining 10,230bp of mitochondrial DNA sequence (often containing highly variable sequence), including DNA barcodes, and 14 microsatellite loci (Hubert *et al*. 2008; McFarlane 2009; Docker et al. 2012). *Lethenteron appendix* (American brook lamprey) and *Lethenteron camtschaticum* (Arctic lamprey) on the other hand shows a morphological and genetic divergence from one another (Docker *et al*. 1999); unlike the first two pairs, American brook and Arctic lampreys no longer co-occur in the same geographic area. Lamprey paired species often share or have overlapping habitats, which makes it challenging and yet important to be able to distinguish between species (Docker 2009). After metamorphosis, lampreys are more readily identifiable. Morphological differences include variation in the sizes of the adults, size of the eye, and oral disc, and differences in dentition at stage-dependent intervals. Other histological differences include a lack of a functional gut in non-parasitic metamorphosed lampreys,

whereas the gut is clearly functional during the feeding phase of the parasitic species. These differences allow for species identification after metamorphosis, though differences are not always clear (e.g., in small-bodied parasitic species or at spawning, when the gut has atrophied in both life history types; Docker 2009).

1.1.4 Lamprey Embryonic Development

For all species of lamprey, spawning occurs in the spring, during which environmental conditions such as temperature (10.0 - 26.1 °C), river velocity (0.5 - 1.5m/s), and water depth (13-170cm) are important (Carlander 1969; Manion and Hanson 1980). In *P. marinus*, water temperatures above 10°C induce sexual maturation, which occurs over several weeks. Once the river has reached the optimal temperature, typically around 16-17°C, the lampreys spawn and embryo development begins (Langille and Hall 1988). At optimal growth temperatures of approximately 18°C, lampreys undergo the typical stages of embryogenesis seen in vertebrates. The first two cleavages are meridional and occur between 2-8 and 8-10 hours, respectively, after fertilization. The third cleavage occurs around 11-12 hours, forming the 8-cell stage. This is followed by the fourth division at 13-15 hours, fifth at 16-19 hours, and sixth at 19-24 hours (Piavis 1961). Following the sixth cleavage or 64-cell stage at approximately 24-64 hours, the cell divisions become irregular and the blastocoel becomes visible. The next stage of embryogenesis, which occurs at 4-5 days, is the development of the gastrula with the formation of the blastopore. Following this, at 5-6 days, is the appearance of the neural groove and folds. At 6-8 days post-fertilization, the first somites appear and are accompanied by the lifting of the head from the yolk sac and appearance of the stomodeum. Continued development of the somites, embryo spiralling and spontaneous

movements can be seen during 8-12 days post-fertilization. Finally, at the time of hatching, 10-13 days post-fertilization, development of the olfactory pit occurs, 18-35 somites and the pericardial cavity can be seen externally, and the body of the now prolarva is pistol-shaped (Piavis 1961).

1.2 Myomeres

1.2.1 Myomeres in Lampreys

Myomeres are a mass of skeletal muscle fibres that are found in chordates (Chen 2011). In fish, they are separated by a thin layer of connective tissue and are seen as a segmentation of muscle fibres that run along the dorso-ventral axis (Wolpert 2007). They are formed from the dermomyotome of the somites during embryogenesis, which subsequently gives rise to the myotome. The myotome gives rise to the skeletal muscle fibres, which form the myomeres (Wolpert 2007). Myomeres develop from somites during embryogenesis, and in some paired species of lamprey, the number of trunk myomeres can help identify individual species (Vladykov and Kott 1980; Docker 2009). Trunk myomere counts typically include myomeres from the last branchial pore opening to the anterior tip of the cloacal slit (Hubbs and Trautman 1937). In general, the non-parasitic lamprey of a pair has a lower myomere count than its corresponding parasitic species, presumably corresponding with its smaller adult body size. For example, the paired species of western brook lamprey and river lamprey differ in their average trunk myomere counts of 60.7 and 66.8, respectively (Vladykov and Follett 1958; 1965). However, there is sufficient variation and overlap of myomere counts within both species to prevent accurate diagnostic identification of individuals using this morphometric character (Vladykov and Follett 1965; Docker 2009). Other species, like the European

river (*Lampetra fluviatilis*; 62.4 mean trunk myomeres) and brook (*Lampetra planeri*; 61.2 mean trunk myomeres) lampreys, show even less difference in myomere counts (Potter and Osborne 1975); there are statistical differences between populations but, again, insufficient variation to use myomere count as a diagnostic character. Some species pairs, like the three pairs recognized in the *Ichthyomyzon* genus, show no difference at all (Hubbs and Trautman 1937). Reduction in myomere number is pronounced in relict non-parasitic species; these are species that have had a long time to diverge from their presumed parasitic ancestor (Docker 2009). For example, with *La. richardsoni* and *La. pacifica* (Pacific brook lamprey) there is a noticeable difference in mean trunk myomere counts (Reid *et al.* 2011). This reduction presumably takes time to evolve, and therefore it is presumed that more recently-derived non-parasitic species show less difference in myomere counts relative to the parasitic ancestor (Docker 2009). Myomeres therefore, often show differences between parasitic and non-parasitic lamprey species. Understanding these difference can help lead to a better evoultionary understanding between the two distinct lifesytles.

1.2.2 Myomere Variation

Interestingly, myomere counts (and number of vertebrae) have been shown to vary between fish species and within a single species located over different geographical clines (Jordan 1891; McBride and Horodysky 2004). Jordan (1891) found that an inverse relationship exists between temperature and the number of myomeres or vertebrae, which has become known as Jordan's rule. A plausible explanation for this phenomenon is that individuals developing in cooler temperatures could experience slower development and therefore have a longer differentiation phase that results in the formation of more serial

elements such as vertebrae and myomeres (Bosley and Conner 1984; McPhee *et al*. 2012). McBride and Horodysky (2004) have shown that two distinct morphs of ladyfish occur in the western North Atlantic Ocean, with the northern morph (*Elops saurus)* possessing 79-87 myomeres and the southern morph (*Elops* sp.) possessing 73-78 myomeres. They note that meristic (segment) differences appear to arise from latitudinal clines due to differences in the temperature of the spawning areas (McBride and Horodysky 2004). Yamahira and Nishida (2009) also noted a latitudinal patterning of myomeres in medaka (*Oryzias latipes*) off the coast of Japan. Goodman *et al*. (2009) shows that this pattern is also true within the lamprey genus *Lampetra*, but appears to be the result of taxonomic uncertainty; ongoing studies are showing that most, if not all of the low-myomere southern populations are distinct species (Reid *et al*. 2011; Boguski *et al*. 2012). Thus, the role that latitude plays on myomere development is not yet clear; however, given the nature of the development of myomeres, the mechanisms involved may be mediated by embryological developmental genes.

1.3 *Hox* Genes

1.3.1 *Hox* Gene Overview

Myomeres develop during embryogenesis, and are therefore under the control of embryological developmental genes (Wolpert 2007). These genes, which play essential roles in the development of the anterior-posterior (AP) axis in all bilaterian organisms are known as *Hox* genes (Takio *et al*. 2007; Mallo *et al.* 2010). *Hox* genes encode a group of developmental homeobox transcription factors. The homeobox is a DNA sequence about 180 base pairs (bp) in length that codes for a homeodomain, which is a sequence of amino acids that can bind DNA and thereby regulate the expression of genes involved in

development. *Hox* genes have been organized into gene clusters in which some organisms may contain multiple sets of these clusters (Duboule 2007). In most vertebrates, the *Hox* genes have four **paralogous** gene clusters situated on different chromosomes. These paralogous group (PG) clusters are the result of ancestral genome duplications and are commonly known as the *Hox*A, *Hox*B, *Hox*C, and *Hox*D clusters (Takio *et al*. 2007). Lampreys have at least three PGs of *Hox* genes and it is yet to be determined whether or not they contain a fourth (Kuraku *et al*. 2008). Typically, there are 13 *Hox* genes in any given vertebrate *Hox* cluster (Takio *et al.* 2007; Barber and Rastegar 2010), but the full array of genes is not always found due to gene redundancy and subsequent loss of function of some paralogues. Redundancy of the *Hox* genes also allows for novel functions to arise through mutations and selection (Takio *et al*. 2007). In the amphioxus (subphylum Cephalochordata) and other non-lamprey primitive fishes (coelacanths and cartilaginous fishes), this process has presumably led to the generation of a fourteenth *Hox* gene (Kuraku *et al*. 2008). Lampreys, like the previous taxa, have also been shown to express **orthologous** *Hox14* genes (Kuraku *et al*. 2008). These genes have been subsequently lost in the Actinopterygii (bony ray-finned fish) lineage (Ferrier 2004).

1.3.2 *Hox* Expression

Hox genes are expressed in a temporal collinear pattern that runs 3′ - 5′ along the chromosome (Barber and Rastegar 2010; Pick and Heffer 2012). In other words, the genes closer to the 3′ end of the cluster are initially expressed at a more anterior location in the developing animal than genes that are closer to the 5′ end. The overlapping expression of subsets of *Hox* genes will determine how a particular tissue will develop along the AP axis. The temporal collinear expression of the *Hox* genes is therefore

associated with growth and elongation of the embryonic axis (Duboule 1994). Studies examining the expression of *Hox* genes suggest the presence of a mechanism called posterior prevalence, which accounts for the observations that despite the broad, overlapping expression of the *Hox* genes, phenotypes are the result of the most posteriorly-expressed genes (Duboule and Morata 1994). Strangely, lamprey *Hox* genes have been noted to not follow strict spatial collinearity (Cohn 2002; Takio *et al*. 2007).

1.3.3 *Hox* Structure

The structure of the lamprey *Hox* genes is usually represented by two exons, but triple exons can be found in *Hox14* genes and some of the *Hox13* genes (Kuraku *et al.* 2008), where exon3 is produced by the insertion of an intron into exon2, dividing it into two separate exons. Exon1 in *Hox* genes is often highly dissimilar to other *Hox* genes and typically ranges from 150-600 bp in length (Kuraku *et al*. 2008). Exon2 (and exon3) is more highly similar across different *Hox* genes. It contains a 180bp domain known as the homeodomain. The homeodomain is highly conserved across all bilaterians, which allows for the comparison between more highly divergent individuals across different taxa.

1.3.4 Anterior *Hox* Genes

Anterior *Hox* genes (1 through 8) within lampreys are those involved in anterior development, particularly those involved with the development of the head. In lampreys, *Hox2* - *Hox8* provide positional cues to the migrating neural crest-derived ectomesenchyme, which pattern the pharyngeal arches, contribute to the hyoid arch (second pharyngeal arch) and also help to contribute to mouth formation (Takio *et al*. 2007; Cerny 2010). Interestingly, however, differences in anterior *Hox* expression,

relative to other vertebrates, have been seen in the lamprey mandibular arch (or first pharyngeal arch) (Takio *et al*. 2004). Cohn (2002) reported expression of anteriorly-expressed *Hox* genes in *La. fluviatilis*, whereas Takio *et al*. (2004) reported no expression of anteriorly expressed *Hox* genes in *Le. camtschaticum* (formerly known as *Lethenteron japonicum* or *Lampetra japonica*). Studies on mice have shown that *Hoxa2* expression is exclusively responsible for hyoid arch formation and disruption of its expression results in homeotic transformation into a secondary mandibular arch (Barrow and Capecchi 1996), thus showing that lack of *Hox* expression in gnathostomes allows for the development of the mandibular arch (Kuratani 2004). Therefore, the lack of *Hox* expression can occur in both agnathans (jawless vertebrates) and gnathostomes (jawed vertebrates). It is the rise of the mandibular arch and subsequent evolution of dorsal-ventral genes that allowed for the dorsoventrally articulated jaw (Cerny 2010). Because lampreys have no hinged jaw, yet possess a mandibular arch as a result of lack of *Hox* gene function, they have been key model species in studies of how the dorsoventrally articulated jaw evolved (Kuratani 2005).

1.3.5 Posterior *Hox* Genes

Posteriorly expressed *Hox* genes (9 through 14) are also important, and are responsible for posterior development along with elongation of the individual (Takio *et* al. 2007). *Hox* genes functioning in posterior development have been identified in lamprey species *P. marinus*, *Le. camtschaticum* and *La. planeri*. First identified were *Hox9(A,B and C)* and *Hox10* genes (GenBank Accession numbers: AF044809 - AF044812) in *La. planeri* by Sharman and Holland (1998), followed by Irvine *et al*. (2002) with the identification of *Hox9*, *Hox10* and *Hox11* genes (AF410918 - AF410925)

in *P. marinus*. Following this, Takio *et al*. (2007) identified *LjHox9r*, *LjHoxW10a, LjHox10s* and *LjHox11* genes in *Le. camtschaticum*, and examined their developmental and tissue-specific expression. *LjHox9, LjHoxW10a* and *LjHox11* were expressed in both the putative tailbud and posterior neural tube. Curiously, *LjHox10s* expression is restricted from the neural tube and is only expressed in the posteriorly developing tail, of which expression was especially strong in the dorsal mesenchyme (Takio *et al.* 2007). More recently, *LjHox13-alpha*, *LjHox13-beta* and *LjHox14-alpha* were described in *Le. camtschaticum* (Kuraku *et al*. 2008). None of these additional genes are expressed in the mesenchyme of the dorsal fin fold, and in addition, *LjHox14-alpha* was found to be absent in the central nervous system, somites and fin buds/folds, where *Hox* expression is normally found. This alteration from normal *Hox* expression may be associated with its loss of function in tetrapod and teleost lineages (Kuraku *et al*. 2008). Fragments of the putative *Hox13* genes have been described by Fried *et al*. (2003) and Sharman and Holland (1998), but the short sequences obtained did not allow them to be properly categorized. Curiously, no putative *Hox12* gene has been identified in lampreys, but *Hox13* and *Hox14* genes have been identified (Kuraku *et al*. 2008).

1.3.6 *Petromyzon marinus* Genome Assembly

Petromyzon marinus is the first and currently only lamprey genome to have been sequenced and there are currently two different assemblies available. The first annotation of the *P. marinus* genome was produced in 2007 by the Washington University Genome Sequencing Center. An approximate 7x genome coverage was sequenced using a whole genome shotgun (WGS) approach using DNA obtained from *P. marinus* liver. It was later found that lamprey somatic tissue undergoes a loss of about 20% of its DNA (Smith *et al*.

2009). The somatic tissue shows consistent patterns of loss for some specific genes such as *Germ1*; however, some tissue-specific variation can also be observed among genome sizes estimated by flow-cytometry (Smith *et al*. 2009). The second annotation of the *P. marinus* genome, built using the Ensembl pipeline, was produced in 2011 with a total of 5x genome coverage and using WGS and BAC-end sequencing (Smith *et al.* 2013).

1.4 Research Objectives

Prior to commencing this study, the full complement of *Hox* genes in lampreys had not been examined. Previous *Hox* genes have been described using many different methods such as targeted amplification or detection from genomic DNA, complementary DNAs (cDNAs), cosmid libraries, lambda phage libraries and P1 artificial chromosomes (Amores *et al.* 1998; Carr *et al.* 1998, Sharman and Holland 1998; Cohn 2002; Force *et al.* 2002; Irvine *et al.* 2002; Takio *et al.* 2004; Takio 2007; Kuraku *et al.* 2008). With the recent release of the partially annotated genome of *P. marinus* that has been available prior to 2009 (Smith *et al*. 2013), it was possible to now search more thoroughly for more than just the small subset of *Hox* genes previously identified in this species. A primary goal of this study was therefore to use this new genomic resource, along with molecular biology approaches, to identify more *Hox* genes in *P. marinus*, and to identify putative homologs in the three lamprey species found in Manitoba, *Ichthyomyzon castaneus, I. fossor*, *I. unicuspis*. By identifying and characterizing additional *Hox* genes in lampreys, it will be possible to gain a better understanding of both evolutionary and developmental aspects of lampreys as a group and their phylogenetic relationship to other vertebrate species. The following is an outline of the specific objectives of this thesis.

1.4.1 Analysis of Full Length Coding Region of Posteriorly Expressed *Hox* Genes

Most of the studies examining *Hox* genes within lamprey species have focused on the homeobox region of these genes, and hence, little is known of the sequences flanking this conserved region of this family of genes. The current study set out to identify full length coding sequences of posteriorly expressed *Hox* genes *Hox9*, *Hox10* and *Hox11* in *P. marinus* by using the Rapid Amplification of cDNA Ends (RACE) technique. By acquiring more sequence information of these genes, it would be possible to gain a better understanding of the origins of these genes and to make more informed comparisons of these genes with those found in other species.

1.4.2 *Hox G*ene Identification Within *P. marinus* Genome

The full complement of *Hox* genes within lamprey species had not yet been identified when this study started. Previous studies have only been able to guess at the number of *Hox* genes present within their *Hox* clusters. Furthermore, as a result of lack of information on complete *Hox* gene numbers in lampreys, it is still unknown as to how many *Hox* clusters lampreys have. This study aims to identify *Hox* gene sequences from within the *P. marinus* genome databases to further understand *Hox* composition within lampreys and to help better understand *Hox* gene evolution.

1.4.3 Identification of *Hox10s* Homeobox Orthologs and Relationship to Myomere Development

Lamprey myomere counts can vary greatly among different lamprey species (Docker 2009). The genetic mechanisms behind these varying myomere counts are still not understood in lampreys. Takio *et al*. (2007) showed that expression of *LjHox10s* in *Le. camtschaticum* was not fixed at any axial levels and followed the extension of the

body axis. This fact suggested that *Hox10s* could be a potential candidate for the genetic control of varying myomere numbers in different lamprey species. In this study, I examined *Hox10s* orthologs in nine species across five genera, with the overall trunk myomere counts ranging from 47 to 78, in an effort to determine whether any gene sequence changes were correlated with differences in myomere numbers.

1.4.4 Identification of Novel Lamprey *Hox* Genes in Lamprey Species

Hox genes are very important in studying evolutionary and developmental biology and yet have only been explored in a small number of lamprey species. In this study, I examined methods for identifying multiple *Hox* gene targets in a variety of lamprey species. Through PCR amplification of *Hox* genes from genomic DNA, I determined the efficiency to which new genes could be discovered in previously unstudied lamprey species.

2. Materials and Methods

2.1 Posterior *Hox* Gene Amplification

2.1.1 Isolation of Lamprey Genomic DNA

Transverse whole body tissue samples, approximately 2mm in length, were obtained from recently euthanized larval (1-3 year old) *Ichthyomyzon castaneus* (chestnut lamprey), and *I. fossor* (northern brook lamprey)/*I. unicuspis* (silver lamprey). Skin and muscle samples were obtained from 100% ethanol-preserved specimens from *Entosphenus tridentatus* (Pacific lamprey), *Eudontomyzon danfordi* (Carpathian lamprey), *Geotria australis* (pouched lamprey), *I. castaneus*, *I. fossor*, *I. gagei* (southern brook lamprey), *I. unicuspis*, *Lampetra pacifica* and *La. richardsoni* (western brook lamprey), collected from a selection of river systems across the world (Table 2.1). Tissue samples were approximately 6mm x 4mm x 2mm and were obtained from the right or left side of the individual posterior to the gill pores and anterior to the cloacal opening. DNA extractions were performed using either the CTAB gDNA purification method (for skin and muscle tissue) and Qiagen DNeasy Blood & Tissue Kit for all other tissues; DNA quantity was assessed using a GE Nanovue spectrophotometer.

DNA extracted using the Qiagen DNeasy Blood & Tissue Kit was performed according to the manufacturer's specifications, with some minor modifications to the initial tissue disruption process. Tissue samples were cut into small (3-5mm) pieces and placed into 180µL ATL buffer plus 20µL (500mg/ml) proteinase K and crushed using a melted pipette tip as a pestle. The tissue was then incubated overnight, for approximately 14-18 hours, at 56°C on a platform rotating at 300rpm. Thereafter, the DNA was extracted using the spin columns according to the manufacturer's specifications.

Table 2.1: Species and locations of lampreys used for samples. (En. = *Entosphenus*, Eu. = *Eudontomyzon*, G. = *Geotria*, I. = *Ichthyomyzon*, La. = *Lampetra*).

Species	Location	Water source
En. tridentatus	USA, CA	Cape Mendocino to Cape Conception
Eu. danfordi	Ukraine	Borzhava River (Black Sea basin, Danube R. system)
Eu. vladykovi	Slovenia	Krka River (Black Sea basin, Danube R. system)
G. australis	New Zealand	Mataura Falls
G. australis	New Zealand	Kaniwhaniwha Stream
I. castaneus	Can, MB	Rat River
I. castaneus	USA, AR	Current River
I. castaneus	USA, WI	Mississippi River
I. fossor	Can, MB	Rat River
I. fossor	Can, MB	Birch River
I. fossor	Can, ON	Lake Huron
I. gagei	USA, LA	Sandy Creek
I. unicuspis	Can, ON	Lake Huron
La. pacifica	USA, WA	Big Creek
La. richardsoni	Can, BC	Smith Creek

To extract DNA using the CTAB method, lamprey tissue was placed into a 1.5ml tube and frozen using liquid nitrogen and disrupted using a plastic disposable pestle. CTAB buffer (500µL; 2% hexadecyltrimethylammonium bromide, 100nM Tris-HCl, pH 8, 29mM EDTA, pH 8, 1.4M NaCl, 0.2% β-mercaptoethanol, 0.1mg/mL proteinase K) was added to the disrupted tissue and incubated for 3 hours at 65°C and shaking at 300rpm. The solution was centrifuged at 15,000xg for 10 minutes to pellet any undigested debris, and the supernatant was subjected to two repetitions of phenol:chloroform extractions (Taggart *et al*. 1992). The aqueous phase was then ethanol-precipitated using 0.3M sodium acetate and 60% ethanol. The DNA was dried in a vacuum centrifuge for 10 minutes. The dried pellet was then re-suspended in 50µl of Nanopure water (nH2O).

2.1.2 Primer Design

Degenerate primers were designed to amplify paralogous *Hox* gene groups *Hox9*, *Hox10* and *Hox11* using sequences that were previously identified in other lamprey species (Table 2.2). Multiple sequence alignments were performed using ClustalW2 (Appendix A), and primers were designed in conserved regions to target regions of similarity across the species, with preferential weight given to lamprey gene sequences, with an approximate fragment size of 150bp (Table 2.3).

Table 2.2: Genes used to design primers to amplify *Hox* genes *Hox9*, *Hox10* and *Hox11*. (*C. milii* = *Callorhinchus milii* (elephant shark), *Le. camtschaticum* = *Lethenteron camtschaticum* (Arctic lamprey), *L. menadoensis* = *Latimeria menadoensis* (Indonesian coelacanth), *P. marinus* = *Petromyzon marinus*)

2.2a. *Hox9* genes for *Hox9* primer design

Gene	Species	Accession Number
HoxV9	*P. marinus*	AF410919
HoxT9	*P. marinus*	AF410918
LjHox9r	*Le. camtschaticum*	AB125271
HoxA9	*C. milii*	FJ824598
HoxB9	*C. milii*	FJ824599
HoxC9	*C. milii*	FJ824600
HoxD9	*C. milii*	FJ824601
HoxA9	*L. menadoensis*	FJ497005
HoxB9	*L. menadoensis*	FJ497006
HoxC9	*L. menadoensis*	FJ497007
HoxD9	*L. menadoensis*	FJ497008

2.2b. *Hox10* genes for *Hox10* primer design

Gene	Species	Accession Number
HoxX10	*P. marinus*	AF410922
HoxW10b	*P. marinus*	AF410921
HoxW10a	*P. marinus*	AF410920
LjHox10s	*Le. camtschaticum*	AB286673
LjHoxW10a	*Le. camtschaticum*	AB286672
HoxA10	*C. milii*	FJ824598
HoxB10	*C. milii*	FJ824599
HoxC10	*C. milii*	FJ824600
HoxA10	*L. menadoensis*	FJ497005
HoxB10	*L. menadoensis*	FJ497006
HoxC10	*L. menadoensis*	FJ497007

2.2c. *Hox11* genes for *Hox11* primer design

Gene	Species	Accession Number
HoxZ11b	*P. marinus*	AF410925
HoxZ11a	*P. marinus*	AF410924
HoxY11	*P. marinus*	AF410923
LjHox11t	*Le. camtschaticum*	AB286674
HoxA11	*C. milii*	FJ824598
HoxC11	*C. milii*	FJ824600
HoxD11	*C. milii*	FJ824601
HoxA11	*L. menadoensis*	FJ497005
HoxC11	*L. menadoensis*	FJ497007
HoxD11	*L. menadoensis*	FJ497008

Table 2.3: PCR primers used to amplify target gene fragments from lamprey genomic DNA (gDNA). (F = forward primer, R = reverse primer, standard IUB base codes were used degenerate nucleotide sites - S = G+C, K = G+T, Y = C+T, R = A+G, N = A+C+T+G, D = G+A+T, H = A+T+C)

Target Gene	Primer Sequences	Amplicon
Hox9 homologs	F: 5' CCCTGGAGCTGGAGAAGGAG 3' R: 5' TTCATSCKYCKGTTCTGRAACCA 3'	167bp
Hox10 homologs	F: 5' AAGAARCGNTGCCCYTACAC 3' R: 5' TCADYTTCTTSAGYTTCAT 3'	158bp
Hox11 homologs	F: 5' AGAAGCGCTGCCCCTACACC 3' R: 5' CTTCTCSYTCATYCGHCGGTTC 3'	119bp

2.1.3 Polymerase Chain Reaction Amplification

PCR amplifications were performed using the standard protocol for recombinant Taq DNA Polymerase from Invitrogen with a standard $MgCl_2$ final concentration of 2mM and primer final concentration of 0.8µM. Standard PCR was performed using an initial 94°C denaturation of double stranded DNA for 180 seconds. This was followed by 25 rounds of DNA amplification (number of rounds of replication were increase up to 40 if the quantity of the product was low), which consisted of a 94°C denaturation step for 30 seconds, followed by a 48-65°C annealing step (based on specified primer melting temperature (Tm) unless otherwise noted) for 30 seconds, and finally an extension step of 72°C for 30-360 seconds. A final extension period of 68°C or 72°C for 180 seconds followed the DNA amplification cycles. PCR optimization, which was sometimes required to amplify some genes, involved adjusting $MgCl_2$ concentrations between 1mM and 6mM, reducing annealing temperatures and increasing the rounds of replication.

Amplification of *Hox9*, *Hox10* and *Hox11* genes was performed using primers from Table 2.3. Optimal annealing temperatures were determined based on the melting

temperatures of the primers, extension times of 30 seconds to amplify a maximum fragment of 500bp and 30 rounds of amplification.

2.1.4 Cloning and Verification

PCR amplification products were visualized on 1-2% agarose gels, stained with SYBR-Gold Nucleic Acid Gel Stain (Invitrogen) according to the manufacturer's instructions, and fragments of expected length were then excised and extracted using either QIAquick Gel Extraction Kit or BioRad Quantum Prep Freeze 'N Squeeze DNA Gel Extraction Spin Columns according to the manufacturers' specifications. Other DNA fragments, varying considerably in size from that expected, were not inspected as the expected DNA fragments were the brightest products and provided the expected sequences.

DNA fragments were cloned using one of the two blunt end cloning kits, pSTBlue-1 (EMD Millipore) or CloneJET (Fermentas) according to the manufacturers' specifications. Following ligation of the PCR products into the cloning plasmids, the plasmids were transformed into chemically competent *E. coli* bacteria and were grown and selected on LB agar ampicillin plates. Bacterial colonies were screened using PCR using a small amount of bacteria as template and a single plasmid specific primer (Table 2.4) and a single amplicon specific primer. Colonies generating PCR products of the expected size were selected and incubated in 2mL of LB-ampicillin broth for approximately 16 hours and plasmid DNA was then purified using Qiagen's QIAprep Spin Miniprep Kit according to the manufacturer's specifications. The DNA concentration of the purified plasmid was then determined using a GE Nanovue spectrophotometer. Sequencing was performed by London Regional Genomics Centre -

Robarts or The Centre for Applied Genomics - TCAG Facilities. Raw sequence data was analyzed using Geneious software and Basic Local Alignment Search Tool (BLAST) was used to identify sequences with similarities to the sequenced DNA fragments.

Table 2.4: PCR primers used to amplify target gene fragments from cloning plasmids.

Primer	Sequence
R20 (pstBlue)	5' ATGACCATGATTACGCCAAG 3'
U19 (pstBlue)	5' GTTTTCCCAGTCACGACGT 3'
pJET1.2 Forward	5' CGACTCACTATAGGGAGAGCGGC 3'
pJET1.2 Reverse	5' AAGAACATCGATTTTCCATGGCAG 3'

2.2 - Rapid Amplification of cDNA Ends

2.2.1 - Lamprey Handling

Hox gene expression occurs typically during embryonic development, and since obtaining wild lamprey embryos is very difficult, it was therefore necessary to obtain live spawning-phase adult lamprey. Three spawning-phase adult *I. castaneus* (two males and one female) were captured from the Rat River on June 15th, 2011, in St. Malo, Manitoba, at the base of a dam that prevented their migration further upstream. The animals were transported back to the University of Manitoba's Animal Holding Facility in water coolers containing river water. The lampreys were maintained in the Animal Holding Facility in an 80 gallon aquarium, with constant flow (0.25m/s) of 16°C dechlorinated water. They were held in the dark to prevent further stress on the animals (Langille and Hall 1988). Given that they were spawning-phase animals (i.e., nontrophic), no food was provided. The water temperature was raised to 18°C to prepare the animals for spawning. After 2 days, the lampreys were anesthetized in a solution of 0.05% tricaine methanesulfonate (MS-222) for approximately 30-60 seconds (Langille and Hall 1988) and attempts were made to collect gametes for artificial spawning using techniques

described by Piavis (1961). No eggs were produced from the single female and she was therefore revived in clean dechlorinated water before being returned to the tank for further development. Two successive attempts to extrude eggs were unsuccessful. The female was then euthanized using a 0.2% MS-222 overdose for 5 minutes and a surgical incision on the ventral surface posterior to the gill pores and anterior to the cloacal opening. Upon visual inspection it was discovered that the female had already spawned, and no eggs remained.

2.2.2 - Isolation of *Petromyzon marinus* (Sea Lamprey) Embryo RNA

Adult spawning-phase lamprey were obtained from the Trout River near Roger's City, MI, and transported to Hammond Bay Biological Station in Millersburg, MI, USA. Gametes were manually expressed as described above in Section 2.2.1 and eggs and sperm were mixed as previously described to produce embryos (Lakiza *et al*. 2011). The embryos were cultured in 500mL Pyrex bowls covered with screened lids in re-circulated Lake Huron water at 18°C. RNA was extracted from groups of 20 embryos at time points listed in Table 2.5 using an RNeasy Mini Kit (Qiagen) according to the manufacturer's specifications. The purified RNA was used for cDNA synthesis (described below). Samples were then shipped on ice via courier to the University of Manitoba.

Table 2.5: Embryo samples taken for RNA isolation (20 embryos per sample).

Cohort	Days post-fertilization
1	3, 6
2	3, 6, 9
3	3, 6, 9
4	9
5	Post hatch, stream prolarvae (approximately 15 days post-hatch, approximately 25 days post-fertilization)

2.2.3 - cDNA Synthesis and PCR Verification

cDNA was produced from *P. marinus* embryo RNA using QuantiTect Reverse Transcription Kit (Qiagen) performed according to the manufacturer's specifications. PCR was performed using *P. marinus* embryo cDNA and *Hox9*, *Hox10* and *Hox11* homolog primers (Table 2.3) and the PCR products were visualized on agarose gels as previously described (2.1.4) to confirm that the genes of interest were expressed in the embryos. PCR products of the expected sizes were gel-purified and sequenced to confirm their identity.

2.2.4 - Rapid Amplification of cDNA Ends (RACE)

Primers were designed based on the specification of Clontech's SMARTer RACE cDNA Amplification Kit (Table 2.6). Four universal primers were designed to collectively amplify paralogous groups *Hox9*, *Hox10* and *Hox11* genes in *P. marinus*. Degenerate primary and nested primers were designed to amplify anterior and posterior sections of the gene of interest. These primers were designed to the homeobox region of the genes. 5' and 3' RACE was performed using Clontech's SMARTer Race cDNA Amplification Kit and according to the manufacturer's specifications with some minor adjustments made to optimize the amplification of any potential products. A range of

$MgCl_2$ concentrations between 1mM and 6mM, annealing temperatures between 62 and 70°C and an increase to 40 rounds of replication were used to produce PCR products. PCR products were assessed using gel electrophoresis and extracted for cloning, cloned into the pSTBlue-1 plasmid, sequenced and assessed as described in Section 2.1.4.

Table 2.6: Universal *Hox9*, *Hox10* and *Hox11* RACE PCR primers.

Primer	Sequence
5HxRACE	R: 5' TTCATBCGYCKGTTCTGGAACCAGAT 3'
5HxRACE(n)	R: 5' TTGGWGTAGGGGCAGCGCTTCTT 3'
3HxRACE	F: 5' AAGAAGCGCTGCCCCTACWCCAA 3'
3HxRACE(n)	F: 5' ATCTGGTTCCAGAACMGRCGVATGAA 3'

2.3 - *Petromyzon marinus* Genome Search and Phylogenetic trees

2.3.1 - *Hox* Sequence Identification

Petromyzon marinus genomic contigs were obtained from the Genome Institute at Washington University (Genome1; http://genome.wustl.edu/genomes/detail/petromyzon-marinus/) and the Ensembl Genome Browser (Genome2; release 70 - January 2013; http://uswest.ensembl.org/Petromyzon_marinus/Info/Index). The contig databases were examined using the BLAST feature in the program Geneious to search for the presence of sequences that resembled *Hox* genes using previously identified *Hox* gene sequences derived from lampreys and the elephant shark *Callorhinchus milii*, Japanese killifish *Oryzias latipes*, and house mouse *Mus musculus* (Table 2.7). All contigs were then taken and re-analyzed using BLAST and ClustalW2 - Multiple Sequence Alignment to assess the extent of sequence identity of the various *P. marinus Hox* genes. Contigs from the two different annotations were given different designations: "Contig#####" for The Genome Institute at Washington University and "GL#####" for the Ensembl Genome Browser (##### indicates the contig number). Prefix number designations, ranging from

01 - 14, were then used to tentatively ascribe each gene to a family of *Hox* genes based on BLAST results.

Table 2.7: Previously identified lamprey *Hox* genes. (superscript number indicates the publication in which the *Hox* gene was first described, 1 indicates Irvine *et al*. 2002, 2 indicates Carr *et al*. 1998, 3 indicates Amores *et al*. 1998, 4 indicates Force *et al*. 2002, 5 indicates Cohn 2002, 6 indicates Takio *et al*. 2007, 7 indicates Takio *et al*. 2004, 8 indicates Sharman *et al*. 1998, 9 indicates Kuraku *et al*. 2008)

Petromyzon marinus (Sea lamprey)		*Lethenteron camtschaticum* (Arctic lamprey)		*Lampetra planeri* (European brook lamprey)		*Lampetra fluviatilis* (European river lamprey)	
Gene	Accession	Gene	Accession	Gene	Accession	Gene	Accession
Hox1W[4]	AF434665	*LjHox1W*[6]	AB286671	*LpHox1A*[8]	AF044797	*HoxL5*[3]	AY089981
HoxE2[1]	AF410908	*Hox2*[7]	AY497314	*LpHox1B*[8]	AF044798	*HoxL6*[3]	AY089982
Hox3[1]	AF410909	*LjHox3d*[7]	AB125270	*LpHox1C*[8]	AF044799		
HoxG4[1]	AF410911	*LjHox4w*[7]	AB125269	*LpHox2A*[8]	AF044800		
Hox4W[4]	AF434666	*LjHox4x*[7]	AB125278	*LpHox3A*[8]	AF044801		
Hox4X[4]	AY056469	*LjHox5i*[7]	AB125276	*LpHox4-7A*[8]	AF044802		
HoxF5[1]	AF410910	*LjHox5w*[7]	AB125277	*LpHox4-7B*[8]	AF044803		
HoxJ5[1]	AF410912	*LjHox6w*[7]	AB125275	*LpHox4-7C*[8]	AF044804		
HoxN5[1]	AF410915	*LjHox7m*[7]	AB125272	*LpHox4-7D*[8]	AF044805		
Hoxw5[3]	AF071234	*LjHox8p*[7]	AB125273	*LpHox4-7E*[8]	AF044806		
HoxL5/6[1]	AF410914	*LjHoxQ8*[7]	AB125274	*LpHox8B*[8]	AF044808		
Hoxw6[3]	AF071235	*LjHox9r*[7]	AB125271	*LpHox8A*[8]	AF044807		
HoxK6[1]	AF410913	*LjHoxW10a*[6]	AB286672	*LpHox9C*[8]	AF044811		
HoxN6[1]	AF410916	*LjHox10s*[6]	AB286673	*LpHox9B*[8]	AF044810		
HoxN7[1]	AF410917	*LjHox11t*[6]	AB286674	*LpHox9A*[8]	AF044809		
HoxQ8[2]	AH005896	*LjHox13-alpha*[9]	AB293597	*LpHox10A*[8]	AF044812		
HoxQ8a[2]	AF035589	*LjHox13-beta*[9]	AB293598				
HoxR8[2]	AF035588	*LjHox14-alpha*[9]	AB293599				
HoxT9[1]	AF410918						
HoxV9[1]	AF410919						
HoxX10[1]	AF410922						
HoxW10b[1]	AF410921						
HoxW10a[1]	AF410920						
HoxZ11b[1]	AF410925						
HoxZ11a[1]	AF410924						
HoxY11[1]	AF410923						

2.3.2 – Next Generation RNA Sequencing

An *I. fossor* lamprey transcriptome, prepared by fellow students Craig McFarlane and Erin Spice, was also used to identify *Hox* sequences within this species. These sequences were obtained from next generation direct RNA sequencing derived from an *I. fossor* ammocoete. RNA was extracted from the gonad of one female approximately 97mm in length and an approximate age of 2–4 years old, DNase treated and sent to Oklahoma Medical Research Foundation for direct sequencing using Illumina HiSeq 2000, which produced a database of 24,373 sequenced transcripts ranging from 40 to 17,352bp in size. BLAST analyses were performed as described above and all putative *Hox* sequences were compared to other known *Hox* genes. Number designations were provided to the identified *Hox* genes based on the BLAST similarities, where Hox##_I._fos (## represented *Hox* number) represented each gene identified.

2.3.3 - *Hox* Phylogenetic Trees

To determine the identity of the sequences obtained in Section 2.3.1, maximum likelihood (ML) and maximum parsimony (MP) trees were produced using MEGA (Molecular Evolutionary Genetics Analysis) software 5.05 on amino acid sequences using 1000 bootstrap replications (Tamura *et al.* 2011). ML phylogenetic analysis was based on the Jones-Taylor-Thornton (JTT) model and MP was obtained using the Close-Neighbor-Interchange algorithm (Jones *et al*. 1992; Nei and Kumar 2000). Previous studies have used these two types of analyses to analyze *Hox* and other similar relationships (Stadler *et al*. 2004; Kuraku *et al*. 2009; Kuraku 2011; Boguski *et al.* 2012). *Hox* phylogenetic trees were constructed using three other vertebrate species *M. musculus*, *C. milii*, and *L. menadoensis* (L. men) as a basis for similarity. For single *Hox*

paralogous group phylogenetic trees, *Branchiostoma lanceolatum* (amphioxus; B. lan) was used as a root. Sequences used for the phylogenetic trees were based on predicted amino acid alignments consisting of the first amino acid in the homeobox and continuing to the stop codon. All truncated sequences, possibly derived from either pseudogenes or incomplete PCR amplifications or failed sequencing, were not used in the alignments and subsequent phylogenetic trees.

2.4 - *Petromyzon marinus Hox* Gene Verification

Four predicted *P. marinus* exon1 coding sequences found within *Hox2*, *Hox7*, *Hox8* and *Hox13* genes were discovered in Section 2.3.1 with no observed exon2 and homeobox regions. Primers were designed to the 3′ regions of the newly discovered sequences (Table 2.8). Forward primers from Table 2.8 along with primers 5HxRACE1 and 5HxRACE1(n) from Table 2.6 were used in an attempt to amplify PCR products from *P. marinus* gDNA obtained in 2.1.1. Optimal PCR annealing temperatures were determined based on melting temperatures of the primers and extension times were set for 360 seconds to allow for a potential maximum product of 6000bp using LongAmp Taq DNA Polymerase (New England BioLabs). Products were cloned and sequenced as described above in Section 2.1.4.

Table 2.8: Primers designed for *Petromyzon marinus* exon1 *Hox* genes (F = Forward primer).

Target gene	Primer sequence
Hox2	F: 5' CCCGCTCAGCCTCCCGAGTA 3'
Hox7	F: 5' GGGGCTGCGCATTTACCCGT 3'
Hox8	F: 5' CTCGTCGGCGCAGCTCTTCC 3'
Hox13	F: 5' GCACCTGTGGAAGTCGCCCC 3'

2.5 - Identification of Lamprey *Hox* Genes in Multiple Lamprey Genera

2.5.1 - Amplification of Exon1 in Posterior *Hox* Genes

Degenerate nucleotide primers were designed to amplify the 3′ coding region of exon1 of *Hox* genes *HoxB9*, *HoxC9*, *HoxA10*, *HoxC10*, *HoxD10*, *HoxA11*, *HoxC11* and *HoxD11* using previously identified sequences in other species (Table 2.9). Multiple sequence alignments were performed using ClustalW2 (Appendix B), and primers were designed to target regions of greatest similarity across the species (Table 2.10). Forward primers from Table 2.10 along with reverse primers 5HxRACE and 5HxRACE(n) from Table 2.6 were used in an attempt to amplify PCR products from *I. fossor, I. castaneus, I. unicuspis* and *P. marinus* gDNA obtained in 2.1.1. The optimal PCR protocol from Section 2.1.4 was used, with annealing temperatures based on melting temperatures of the primers. Extension times were set for 360 seconds to allow for a potential maximum product of 6000bp and 30 and 40 rounds of replication were used for amplification. LongAmp Taq DNA Polymerase (New England BioLabs) was used, using variable concentrations of $MgCl_2$ (1mM - 6mM) and annealing temperatures between 3 and 8°C below the predicted Tm of the primers to optimize the PCR when no bands were observed under standard PCR conditions.

Table 2.9: Genes used to design primers to amplify the coding region of exon1 *Hox* genes *HoxB9*, *HoxC9*, *HoxA10*, *HoxC10*, *HoxD10*, *HoxA11*, *HoxC11* and *HoxD11*. (*Anguis fragilis* (slow worm; limbless reptile), *Aspidoscelis uniparens* (desert grassland whiptail lizard), *Boa constrictor* (boa constrictor*), Callorhinchus milii* (elephant shark), *Chalcides bedriagai* (Bedriaga's skink), *Chalcides ocellatus* (ocellated skink), *Danio rerio* (zebrafish), *Gallus gallus* (red junglefowl), *Gekko ulikovskii* (golden gecko), *Heterodontus francisci* (horn shark), *Homo sapiens* (human), *Ichthyophis kohtaoensis* (Koa tao island caecilian), *Oryzias latipes* (Japanese rice fish), *Saiphos equalis* (three-toed skink), *Salmo salar* (Atlantic salmon), *Takifugu rubripes* (pufferfish), *Thamnophis sirtalis* (common garter snake), *Trachemys scripta* (pond slider), *Varanus prasinus* (emerald tree monitor), *Xenopus laevis* (African clawed frog))

2.9a. *HoxB9* genes for *HoxB9* primer design

Gene	Species	Accession Number
HoxB9	*Callorhinchus milii*	FJ824599
HoxB9a	*Danio rerio*	AF071259
HoxB9	*Homo sapiens*	NM_024017
HoxB9ab	*Salmo salar*	NM_001141626
HoxB9a	*Takifugu rubripes*	DQ481665

2.9b. *HoxC9* genes for *HoxC9* primer design

Gene	Species	Accession Number
HoxC9	*Callorhinchus milii*	FJ824600
HoxC9a	*Danio rerio*	NM_131528
HoxC9	*Homo sapiens*	NM_006897
HoxC9a	*Oryzias latipes*	AB208008
HoxC9aa	*Salmo salar*	NM_001139526
HoxC9ba	*Salmo salar*	NM_001139540
HoxC9bb	*Salmo salar*	NM_001139546
HoxC9a	*Takifugu rubripes*	DQ481667

2.9c. *HoxA10* genes for *HoxA10* primer design

Gene	Species	Accession Number
HoxA10	*Anguis fragilis*	GU320330
HoxA10	*Aspidoscelis uniparens*	GU320331
HoxA10	*Boa constrictor*	GU320327
HoxA10	*Callorhinchus milii*	FJ824598
HoxA10	*Chalcides bedriagai*	GU320332
HoxA10	*Gekko ulikovskii*	GU320328
HoxA10	*Heterodontus francisci*	AF224262
HoxA10	*Thamnophis sirtalis*	GU320326
HoxA10	*Trachemys scripta*	GU320325
HoxA10	*Varanus prasinus*	GU320329

2.9d. *HoxC10* genes for *HoxC10* primer design

Gene	Species	Accession Number
HoxC10	*Callorhinchus milii*	FJ824600
HoxC10	*Homo sapiens*	NM_017409
HoxC10	*Ichthyophis kohtaoensis*	GQ176257

2.9e. *HoxD10* genes for *HoxD10* primer design

Gene	Species	Accession Number
HoxD10	*Boa constrictor*	GU320312
HoxD10	*Callorhinchus milii*	FJ824601
HoxD10	*Chalcides bedriagai*	GU320313
HoxD10	*Chalcides ocellatus*	GU320314
HoxD10	*Heterodontus francisci*	AF224263
HoxD10	*Saiphos equalis*	GU320315
HoxD10	*Xenopus laevis*	NM_001090166

2.9f. *HoxA11* genes for *HoxA11* primer design

Gene	Species	Accession Number
HoxA11a	*Danio rerio*	NM_131544
HoxA11b	*Danio rerio*	NM_131147
HoxA11a	*Oryzias latipes*	AB207983
HoxA11b	*Oryzias latipes*	AB207988
HoxA11aa	*Salmo salar*	NM_001139560
HoxA11ab	*Salmo salar*	NM_001139565
HoxA11b	*Salmo salar*	NM_001141672
HoxA11a	*Takifugu rubripes*	DQ481663
HoxA11b	*Takifugu rubripes*	DQ481664

2.9g. *HoxC11* genes for *HoxC11* primer design

Gene	Species	Accession Number
HoxC11a	*Danio rerio*	NM_131165
HoxC11	*Homo sapiens*	NM_014212
HoxC11a	*Oryzias latipes*	AB232922
HoxC11aa	*Salmo salar*	NM_001139525
HoxC11ab	*Salmo salar*	NM_001141665
HoxC11bb	*Salmo salar*	NM_001139545
HoxC11a	*Takifugu rubripes*	DQ481667

2.9h. *HoxD11* genes for *HoxD11* primer design

Gene	Species	Accession Number
HoxD11a	*Danio rerio*	NM_131167
HoxD11	*Gallus gallus*	NM_204620
HoxD11	*Heterodontus francisci*	AF224263
HoxD11	*Homo sapiens*	NM_021192
HoxD11a	*Oryzias latipes*	AB232923
HoxD11a	*Oryzias latipes*	AB208017
HoxD11aa	*Salmo salar*	NM_001139552
HoxD11a	*Takifugu rubripes*	DQ481668

Table 2.10: PCR primers used to amplify target paralogous group exon1 from lamprey gDNA (F = forward primer).

Target Gene	Primer Sequence
HoxB9	F: TATTATGTSGAYTCBATHATAAGTCA
HoxC9	F: ATGTCGRCNACGGGTCCYATAASTAA
HoxA10	F: ATGKCATGYTCSGASARCCCGGCTGCAAACTCKTTTTT
HoxC10	F: ATGTCATGYCCSAAMAATGTGACT
HoxD10	F: ATGTCCTKYCCCARCAGCTCTCC
HoxA11	F: ATGTATTTRCCCAGYTGCACYTAYTACGT
HoxC11	F: AACTCRGTBAATCTGGGMAACTTCTGCTC
HoxD11	F: TTTTTRCCVCARACTACKTCSTGTCA

2.5.2 - *Hox10* Gene Analysis Across Multiple Lamprey Species

Primers were designed to amplify the homeobox region of the gene *LjHox10s* (Accession number AB286673) in *Le. camtschaticum* (Table 2.11). Forward and reverse primers from Table 2.11 were used to amplify *Hox* gene fragments from *En. tridentatus*, *Eu. vladykovi*, *G. australis*, *I. castaneus*, *I. fossor*, *I. gagei*, *I. unicuspis*, *La. pacifica*, and *La. richardsoni* gDNA samples. The optimal PCR protocol from Section 2.1.4 was used, with annealing temperatures based on melting temperatures of the primers; extension times were set for 30 seconds to allow for a potential maximum product of 500bp and 40 rounds of replication. Products were cloned and sequenced as described Section 2.1.4.

Table 2.11: Primers used to amplify *Hox10* orthologs.

Primer name	Primer sequence	Amplicon
10-F1	F: 5' GCCGCGCGCGAGGCCC 3'	261bp
10-R1	R: 5' GGGGTACGGGGCCGTCATCT 3'	

Maximum likelihood and maximum parsimony trees were produced using methods as described above in Section 2.3.3 to assess phylogenetic relationships between

the species. The putative genes identified were also analyzed based on nucleotide and amino acid alignments. These sequences were used to look for indicative polymorphisms in comparison to high and low trunk myomere counts in the identified lamprey species. All lamprey myomere counts were obtained from Docker (2009) except for *G. australis, La. aepyptera* and *La. pacifica*, whose myomere counts were obtained from Neira (1984), Seversmith (1953) and Reid *et al. (*2011) respectively. All myomere counts were based on trunk myomere counts which include myomeres from the posterior margin of the last branchial pore opening to the cloacal slit (Table 2.12) (Hubbs and Trautman 1937).

Table 2.12: Lamprey trunk myomere counts. (*En. = Entosphenus*, *Eu. = Eudontomyzon*, *G. = Geotria*, *I. = Ichthyomyzon*, *La. = Lampetra*, *Le. = Lethenteron*)

Species	Trunk myomere counts
G. australis	66-78
En. tridentatus	61-77
Le. camtschaticum	60-74
La. richardsoni	58-67
Eu. vladykovi	58-68
La. pacifica	54-58
La. aepyptera	53-60
I. unicuspis	49-56
I. castaneus	47-56
I. fossor	47-55

2.5.3 - General *Hox* Gene Screening in Multiple Lamprey Genera

Primers designed in the studies by Pendleton *et al*. (1993) and Force *et al*. (2002) were used to screen *Hox* genes in lamprey species other than *P. marinus* (Table 2.12). The Pendleton *et al.* (1993) study probed cosmid DNA libraries with probes derived from PCR products amplified with primers designed to amplify small segments of the *Hox* homeobox region in acorn worm (*Saccoglossus kowalevskii*; hemichordate), amphioxus (*Branchiostoma floridae*; cephalochordate) and *P. marinus.* The Force *et al*. (2002) study

used degenerate primers to amplify *Hox* gene fragments from cDNA and cosmid libraries solely in *P. marinus*. Forward and reverse primers from Table 2.12 were used to amplify *Hox* gene fragments from *En. tridentatus*, *Eu. vladykovi*, *G. australis*, *I. castaneus*, *I. fossor*, *I. gagei*, *I. unicuspis*, *La. pacifica*, and *La. richardsoni* gDNA samples. Optimal PCR template from Section 2.1.4 was used, with annealing temperatures based on melting temperatures of the primers, extension times were set for 30 seconds to amplify an expect amplicon of 500bp with 30 rounds of replication were used. Products were cloned and sequenced as described above in Section 2.1.4.

Table 2.13: General primers for amplifying P. marinus Hox genes, intended to amplify the homeobox region from multiple Hox gene families. (F = forward, R = reverse).

2.13a. Primers designed in Pendleton et al. (1993)

Target genes	**Primer sequence**	**Primer name**
All Hox	F: 5' AAAGGATCCTGCAGARYTIGARAARGARTT 3'	HoxE
All Hox	R: 5' ACAAGCTTGAATTCATICKICKRTTYTGRAACCA 3'	HoxF

2.13b. Primers designed in Force *et al*. (2002)

Target genes	**Primer sequence**	**Primer name**
Hox 1-9	F: 5' GAATTCCACTTCAACMRSTACCT 3'	1-9Hx
All Hox	R: 5' CATCCTGCGGTTTTGGAACCAIAT 3'	HxReverse

2.13c. Forward and reverse primer pairs and expected amplicon sizes

Primer pairing	**Expected amplicon size**
HoxE - HoxF	150bp
1-9Hx - HoxF	125bp
1-9Hx - HxReverse	111bp
HoxE - HxReverse	136bp

3. Results

3.1 - Identification of Posterior *Hox* Genes in Three *Ichthyomyzon* species.

Using degenerate primers from Table 2.3, putative *Hox* gene fragments from the homeobox region were amplified from three local lamprey species, *Ichthyomyzon castaneus* (chestnut lamprey)*, I. fossor* (northern brook lamprey) and *I. unicuspis* (silver lamprey). Putative *Hox9*, *Hox10* and *Hox11* gene fragments (~120 - 106bp) were obtained for the *I. castaneus* and *I. fossor* and were successfully cloned and sequenced. *Ichthyomyzon unicuspis* gene fragments were also isolated, yet multiple attempts were not successful in cloning and sequencing these fragments. Gene fragments for *I. unicuspis* were faint compared to *I. castaneus* and *I. fossor* gene fragments. This suggested either poor quality of genomic DNA extraction or that the primers were not specific enough to amplify the fragments. Continued attempts to optimize the *I. unicuspis Hox* gene PCR with multiple genomic DNA samples were unsuccessful which further suggested that poor quality of genomic DNA was the problem. The *Hox* sequences from the *I. castaneus* and *I. fossor* samples showed high similarity (e.g., 96-98% in nucleotide sequences) to previously identified *P. marinus* and *Le. camtschaticum Hox8* - *Hox11* genes (Table 3.1).

Table 3.1: Sequence similarity of homeobox region amplified using *Hox* degenerate primers. (*P.m. = Petromyzon marinus*, *L.c. = Lethenteron camtschaticum*)

3.1a. Nucleotide alignments

Species	Primer	Highest similarity (accession)	% nucleotide similarity (total nucleotide length)
Ichthyomyzon castaneus	*Hox9*	*P.m. HoxV9* (AF410919)	98 (125)
	Hox10	*P.m. HoxW10a* (AF410920) and *L.c.LjHox10s* (AB286673)	98 (167)
	Hox11	*L.c.LjHox10s* (AB286673)	97 (157)
Ichthyomyzon fossor	*Hox9*	*P.m.HoxQ8a* (AF035589)	96 (125)
	Hox10	*P.m. HoxX10* (AF410922)	96 (165)
	Hox11	*P.m. HoxW10a* (AF410920) and *L.c.LjHox10s* (AB286673)	96 (157)

3.1a. Amino acid alignments

Species	Primer	Highest similarity (accession)	% amino acid similarity (total amino acid length)
Ichthyomyzon castaneus	*Hox9*	*P.m. HoxV9* (AF410919)	100 (41)
	Hox10	*P.m. HoxW10a* (AF410920) and *L.c.LjHox10s* (AB286673)	100 (55)
	Hox11	*P.m. HoxW10a* (AF410920), *HoxW10b* (AF410921) and *L.c.LjHox10s* (AB286673) *LjHoxW10a* (AB286672)	93 (52)
Ichthyomyzon fossor	*Hox9*	*P.m. HoxQ8 (*AH005896), and *L.c. LjHoxQ8* (AB125274), *LjHox8p* (AB125273)	100 (41)
	Hox10	*P.m. HoxX10* (AF410922)	96 (55)
	Hox11	*P.m. HoxW10a* (AF410920), *HoxW10b* (AF410921) and *L.c.LjHox10s* (AB286673)	96 (52)

Due to high sequence similarities within the homeobox fragments described above, it was not possible to distinguish without ambiguity the posterior *Hox* genes from one another within this 125-165bp region. For example, the *Hox11* primer set in both species appeared to amplify a region of the homeobox that appeared similar to a number of *Hox10* genes within both *P. marinus* and *Le. camtschaticum* (Table 3.1). To help discriminate each of these genes from one another, it was necessary to examine

sequences 5′ and 3′ of the homeobox region. As *Hox9*, *Hox10* and *Hox11* exon1 sequences had not yet been identified within any lamprey species, a 5′ and 3′ RACE approach was used to obtain additional gene sequences.

3.2 - Rapid Amplification of cDNA Ends

To determine the entire coding sequence of the posteriorly expressed *Hox* genes *Hox9*, *Hox10* and *Hox11,* rapid amplification of cDNA ends (RACE) was performed on RNA obtained from *P. marinus* embryos and stream prolarvae (Table 2.5). The cDNA samples derived from the RNA were first tested for the presence of *Hox* gene transcripts using degenerate primers designed to amplify the homeobox region of the posteriorly expressed *Hox* genes (Table 2.3). Confirmation of successful PCRs was determined by identifying the expected DNA fragment sizes, ~167bp for *Hox9* homologs and ~158bp for *Hox10* homologs. Confirmation of gene expression was observed in samples 9 days post-fertilization and 15 days post-hatch (Table 3.2). No observed expression was seen in any of the 3 and 6 day post-fertilization samples (see Section 4.2).

Table 3.2: Confirmation of posterior *Hox* gene expression in *P. marinus* embryo (3-9 days post fertilization) and stream prolarvae samples (15 days post-hatch).

Sample (Cohort/Days post-fertilization)	Confirmation of gene expression
1/3	No
1/6	No
2/3	No
2/6	No
2/9	Yes
3/3	No
3/6	No
3/9	Yes
4/9	Yes
5/15 days post-hatch	Yes

Obtaining RACE products was hindered due to poor quality RNA that was shipped from Michigan to Manitoba, as well as expired or otherwise ineffective RACE kits. As the RNA samples arrived partially thawed, it was suspected that much of the RNA was degraded due to poor spectrometer readings of RNA concentrations. The first two RACE kits that were used were deemed ineffective as positive mouse RNA control samples failed to amplify any products. By the time of use of the third RACE kit, there was very little RNA sample left to use. Only a single RACE PCR product was successfully cloned and sequenced. This RACE product showed a high identity (99.6% nucleotide identity over 254bp) to the *P. marinus* collagen gene *Col2a1a* (Accession DQ136024). Further analysis of the sequence revealed a moderate sequence similarity to the primer 5HxRACE from Table 2.7 (Figure 3.1). The 5HxRACE primer was 52.4% similar to the *P. marinus Col2a1a* gene at the site of binding. This shows a low complementarity to the gene, but it was noted that there were four regions of binding between the primer and the gene: a four nucleotide region in the middle of the primer, a three nucleotide region at the 3′ end and two other two nucleotide complementary sites, one close to the 5′ end and one in the middle. The lack of additional RNA samples prevented any further RACE analyses to be completed.

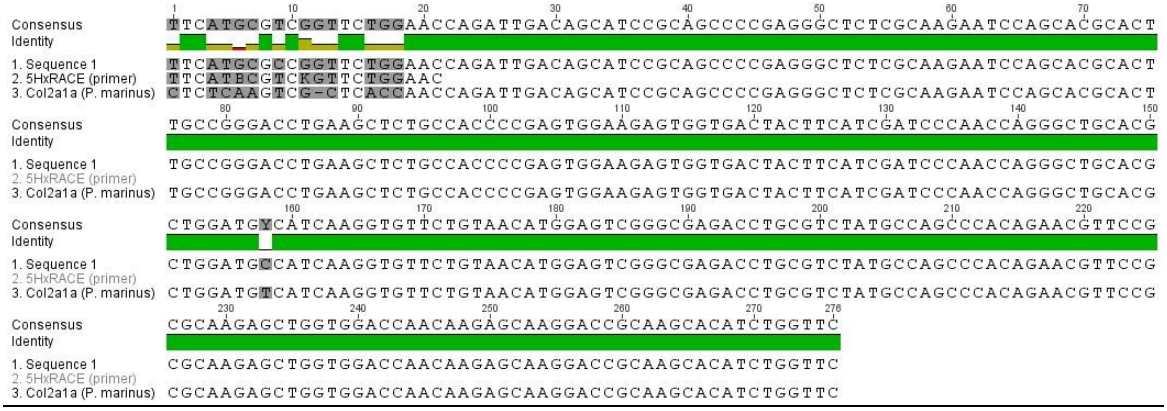


Figure 3.1: Nucleotide alignment of *P. marinus Col2a1a* (Accession DQ136024), RACE product and 5HxRACE primer. (Grey outlined nucleotides indicated less than 100% nucleotide identity across the alignment, white outlined nucleotides indicate 100% nucleotide identity, green identity bar areas indicate the level of identity with full green 100% identity and lack of green 0% identity)

3.3 - *Petromyzon marinus* Genome Search and Phylogenetic Trees

3.3.1 - *Hox* Sequence Identification

Forty-seven contigs were discovered to contain predicted coding sequence for *Hox* genes within the two *P. marinus* contig databases (Table 3. 3). A total of 30 unique putative *Hox* coding sequences were found within these contig sets. Within the set of unique coding sequences, six contigs contained only exon1 sequence, 21 contigs contained only exon2 sequences and three contigs contained both exons. In two instances with putative *Hox13* and *Hox14* genes, these sequences contained an intron (*Hox13* ~900bp and *Hox14* ~5340bp) within exon2's homeobox region. While *Hox14* genes are normally found to have three individual exons, *Hox13* genes normally only have two exons. Kuraku *et al*. (2008) observed in *Le. camtschaticum* that one of two *Hox13* genes discovered has an intron inserted in the middle of the homeodomain. Within the unique coding sequence set of 30 contigs, 12 of the coding sequences were exclusively found in genome set one, four coding sequences were found exclusively in genome two and 13 coding sequences were found in both. Of the 30 unique sequences discovered within the

two genome sets, 12 sequences are newly identified predicted coding sequences while 18 of the sequences matched to previously identified *P. marinus Hox* genes. Of the 18 matched sequences, new predicted coding sequence was identified for 16 of the sequences. Most of the additional identified sequences from the 16 genes were identified on the 3′ end of exon2, following the homeobox region and ending at the stop codon of the genes. New exon1 sequences were also identified for some genes. Exon1 was more easily identified on large contigs already found to contain homeobox sequences. One to ten open reading frames (ORFs) ranging in size from 250bp - 650bps were identified approximately 500bp - 6000bp upstream of the homeobox regions, and were matched to existing *Hox* sequences or identifying features. Anterior *Hox* genes with newly discovered exon1 ORFs were more easily identified using conserved amino acid motifs present at the 3′ region of the exon. Previously identified lamprey sequences containing exon1 were also helpful in identifying some sequences. The majority of exon1 sequences remained unidentified due to the genomes not being fully annotated and high variability between gene sequences. During the writing of this thesis, Smith *et al.* (2013) released a new *P. marinus* genome assembly in which new *Hox* gene information became available. They were able to identify new *Hox* genes by probing bacterial artificial chromosomes with known lamprey sequences for *Hox2*, *Hox4* and *Hox9* then subsequent sequencing and analysis.

Table 3.3: *Petromyzon marinus Hox* genes' exons mapped to genome contigs. (Grey indicates previously identified *Hox* genes not found in contig search, yellow indicates newly identified putative *Hox* genes in this study, numbers in brackets indicate nucleotide length in base pairs (bp) of newly discovered sequences, PG indicates paralogous group, * indicates exon2 divided by an intron producing two exons, ** indicates exon1 identified while exon2 reading frame is disrupted by nucleotide insertions, Contig## indicates a contig from the Genome Institute at Washington University genome database, GL## indicates contig from Ensembl genome database, new assembly of *P. marinus.* genome became available on Ensembl, blue indicates sequence was identified in this study as well as Smith *et al.* (2013).

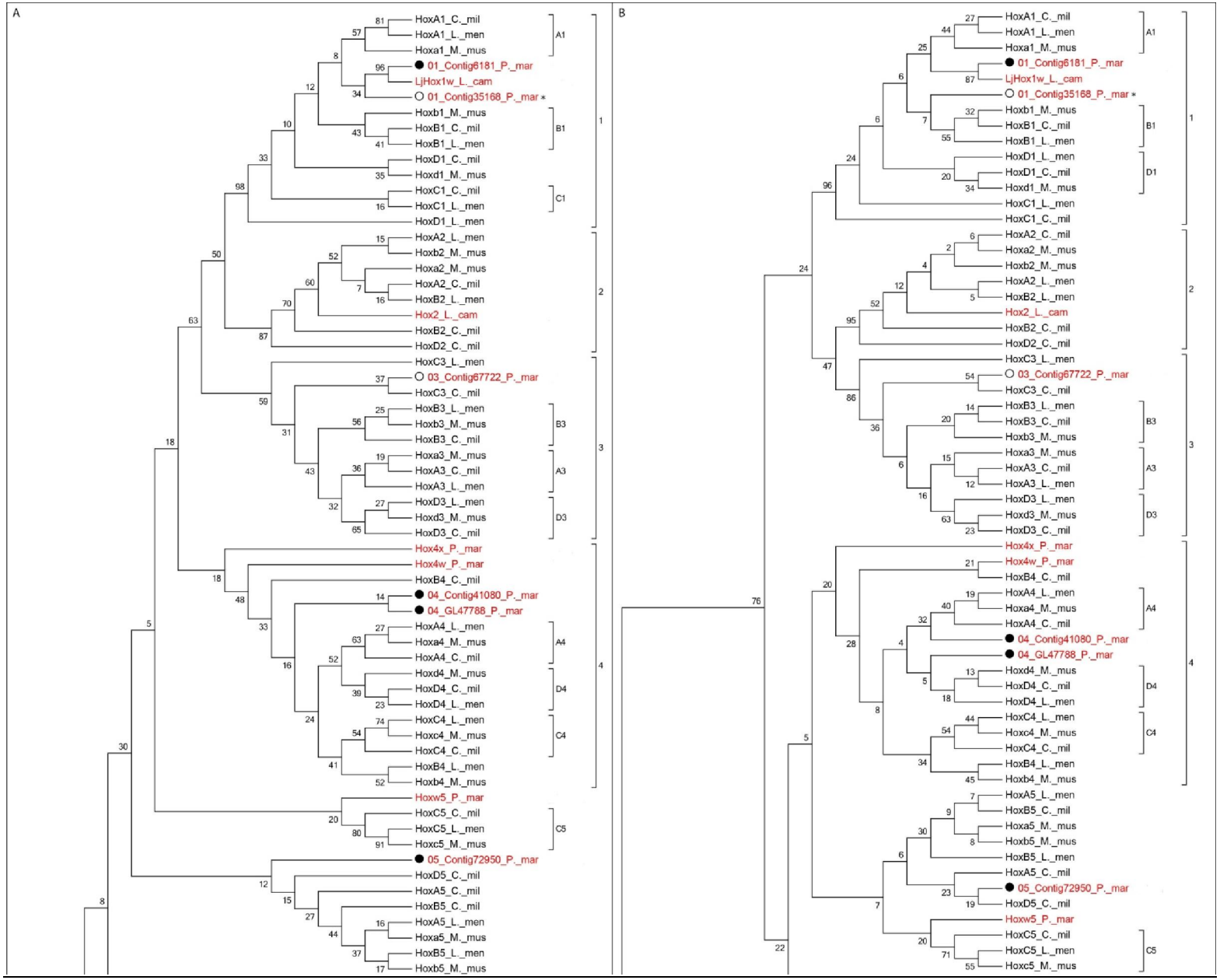
***Hox* PG**	**Gene name**	**Exon1**	**Exon2**	**Exon1 and Exon2**	**Identified in Smith *et al.* 2013**
1	*Hox1w*			Contig6181, GL477571 (36bp)	Yes
	?		Contig35168, GL486885 (471bp)		Yes
2	*HoxE2*				No
	?	Contig68831 (411bp)			No
3	3				Yes
	?		Contig67722 (649bp)		No
4	*HoxG4*		GL477881 (417bp)		Yes
	Hox4W	Contig64676 (0bp) **			Yes
	Hox4X				No
	Pm88-H		Contig41080 (182bp)		No
5	*HoxW5*				No
	HoxN5		Contig72950 (72bp)		Yes
	HoxJ5				No
	HoxF5		Contig8613, GL482944 (108bp)		Yes
5/6	*HoxL5/6*				Yes
6	*HoxK6*		Contig22678, GL477571 (87bp)		Yes
	HoxN6				No
	Hoxw6		Contig66553 (0bp)		Yes
7	*HoxN7*				No
	?	GL476758 (453bp)			Yes
8	*HoxQ8*	Contig82182, GL476758 (0bp)			Yes
	HoxQ8a		Contig49403 (105bp)		Yes
	HoxR8	Contig34634 **		GL483321 (480bp)	Yes
	?	GL477571 (510bp)			No

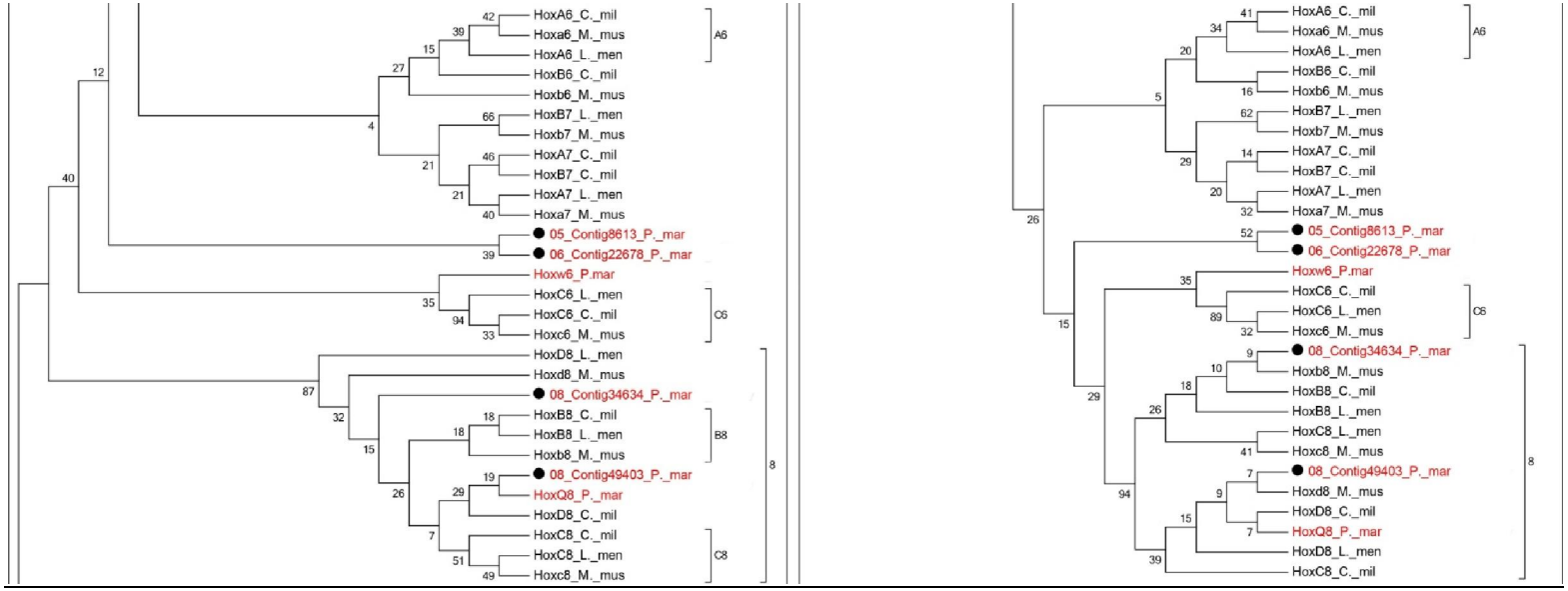
9	*HoxT9*		Contig62581 (42bp)		Yes
	HoxV9		Contig75333, GL476758 (39bp)		Yes
	Pm98-s		Contig38193, Contig49546, GL482503 (134bp)		Yes
10	*HoxW10a*		Contig52464, GL477881 (60bp)		Yes
	HoxW10b				No
	HoxX10		GL493006 (39bp)		Yes
11	*HoxY11*		Contig82219 (42bp)		Yes
	HoxZ11a				No
	HoxZ11b		Contig89919 (42bp)		Yes
	?		Contig38436, GL477014 (222bp)		No
12					
13	?			Contig25758, Contig21977, GL479015 (1255bp) *	Yes
	?		Contig23980, GL484353 (213bp)		No
	?		Contig60485 (261bp)		No
	?		Contig64214, Contig67493 (349bp)		No
	?	Contig30770 (552bp)			No
14	?	GL486262 (111bp)	Contig28189, GL494617 (330bp) *		No

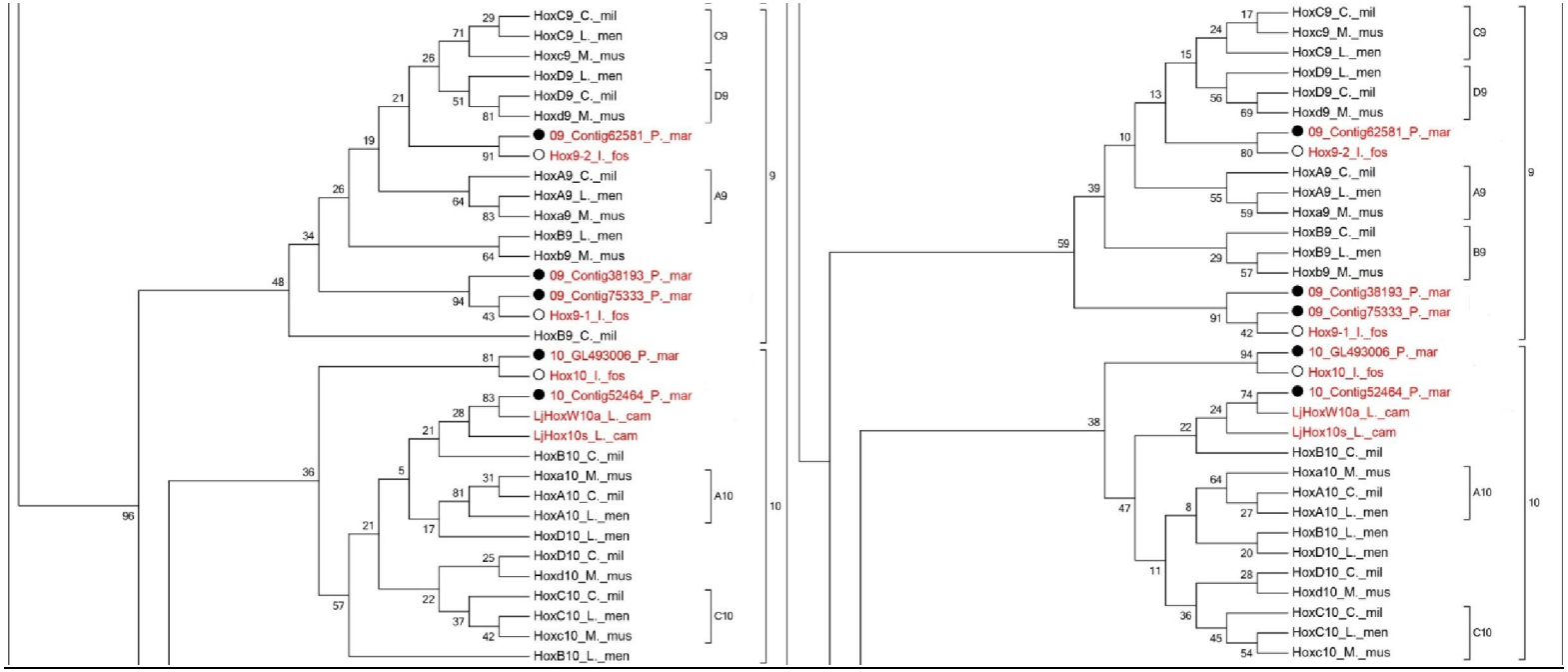
3.3.2 - *Hox* Phylogenetic Trees

Initial phylogenetic trees, maximum likelihood and maximum parsimony, containing all *Hox* genes 1-14 were constructed (Section 2.3.3), to determine if the phylogenetic trees placement of the *Hox* sequences into individual *Hox* clades were consistent with BLAST results to help identify groupings of paralogous genes in which smaller trees could subsequently be constructed (Figure 3.2). Three vertebrate species, whose complete *Hox* gene complements have been analyzed, were used in producing the phylogenetic trees. These *Hox* genes were used to create a comparison of paralogous genes. The three species used were *Mus musculus* (house mouse), *Latimeria menadoensis* (Indonesian coelacanth), and *Callorhinchus milii* (elephant shark). Initial analysis showed

that anteriorly expressed *Hox* genes *Hox1-Hox8* (Figure 3.3) and posteriorly expressed *Hox* genes *Hox9-Hox14* (Figure 3.4) each formed a clade. The *Hox9* and *Hox10* genes were then analysed separately to produce *Hox9-Hox10* (Figure 3.5) to help further resolve the relationships of the genes in these clades. Anteriorly and posteriorly expressed genes were then re-analyzed as separate groups and new phylogenetic trees were constructed containing these subsets of genes. Analysis of the anterior set of *Hox* genes (Figures 3.6 - 3.9) showed significant differences between *Hox1* (Figure 3.6), *Hox2*, *Hox3* (Figure 3.7) and *Hox8* (Figure 3.9) paralogous groups and were therefore isolated into separate clades and re-analyzed as independent sets (covered later in Sections 3.3.2.1, 3.3.2.2, 3.3.2.3 and 3.3.2.5). The *Hox* gene clade *Hox4-Hox7* (Figure 3.8) was then separated and re-analyzed in a grouped subset (Section 3.3.2.4). The analysis of this clade was not well supported by bootstrap analyses (42 for ML and 40 for MP) and no further resolutions could be made within this grouping of *Hox* genes. Independent analysis of the posterior *Hox* gene phylogenetic trees (Figures 3.10 - 3.14) showed separation of *Hox* genes *Hox11* (Figure 3.12), *Hox12*, *Hox13* (Figure 3.13) and *Hox14* (Figure 3.14) into separate clades (Sections 3.3.2.8, 3.3.2.9, 3.3.2.10 and 3.3.2.11). The *Hox* genes in *Hox9-Hox10* (Figure 3.5) paralogous groups were separated from the rest of the posterior *Hox* genes and re-analyzed to resolve gene phylogenetic to allow them to be separated into individual phylogenetic trees. *Hox9* (Figure 3.10) and *Hox10* (Figure 3.11) gene phylogenies were analyzed independently (Sections 3.3.2.6 and 3.3.2.7).


A
HoxA1_C._mil
HoxA1_L._men
Hoxa1_M._mus
01_Contig6181_P._mar
LjHox1w_L._cam
01_Contig35168_P._mar *
Hoxb1_M._mus
HoxB1_C._mil
HoxB1_L._men
HoxD1_C._mil
Hoxd1_M._mus
HoxC1_C._mil
HoxC1_L._men
HoxD1_L._men
HoxA2_L._men
Hoxb2_M._mus
Hoxa2_M._mus
HoxA2_C._mil
HoxB2_L._men
Hox2_L._cam
HoxB2_C._mil
HoxD2_C._mil
HoxC3_L._men
03_Contig67722_P._mar
HoxC3_C._mil
HoxB3_L._men
Hoxb3_M._mus
HoxB3_C._mil
Hoxa3_M._mus
HoxA3_C._mil
HoxA3_L._men
HoxD3_L._men
Hoxd3_M._mus
HoxD3_C._mil
Hox4x_P._mar
Hox4w_P._mar
HoxB4_C._mil
04_Contig41080_P._mar
04_GL47788_P._mar
HoxA4_L._men
Hoxa4_M._mus
HoxA4_C._mil
Hoxd4_M._mus
HoxD4_C._mil
HoxD4_L._men
HoxC4_L._men
Hoxc4_M._mus
HoxC4_C._mil
HoxB4_L._men
Hoxb4_M._mus
Hoxw5_P._mar
HoxC5_C._mil
HoxC5_L._men
Hoxc5_M._mus
05_Contig72950_P._mar
HoxD5_C._mil
HoxA5_C._mil
HoxB5_C._mil
HoxA5_L._men
Hoxa5_M._mus
HoxB5_L._men
Hoxb5_M._mus
A1
B1
C1
B3
A3
D3
A4
D4
C4
C5
1
2
3
4
B
HoxA1_C._mil
HoxA1_L._men
Hoxa1_M._mus
01_Contig6181_P._mar
LjHox1w_L._cam
01_Contig35168_P._mar *
Hoxb1_M._mus
HoxB1_C._mil
HoxB1_L._men
HoxD1_L._men
HoxD1_C._mil
Hoxd1_M._mus
HoxC1_L._men
HoxC1_C._mil
HoxA2_C._mil
Hoxa2_M._mus
Hoxb2_M._mus
HoxA2_L._men
HoxB2_L._men
Hox2_L._cam
HoxB2_C._mil
HoxD2_C._mil
HoxC3_L._men
03_Contig67722_P._mar
HoxC3_C._mil
HoxB3_L._men
HoxB3_C._mil
Hoxb3_M._mus
Hoxa3_M._mus
HoxA3_C._mil
HoxA3_L._men
HoxD3_L._men
Hoxd3_M._mus
HoxD3_C._mil
Hox4x_P._mar
Hox4w_P._mar
HoxB4_C._mil
HoxA4_L._men
Hoxa4_M._mus
HoxA4_C._mil
04_Contig41080_P._mar
04_GL47788_P._mar
Hoxd4_M._mus
HoxD4_C._mil
HoxD4_L._men
HoxC4_L._men
Hoxc4_M._mus
HoxC4_C._mil
HoxB4_L._men
Hoxb4_M._mus
HoxA5_L._men
HoxB5_C._mil
Hoxa5_M._mus
Hoxb5_M._mus
HoxB5_L._men
HoxA5_C._mil
05_Contig72950_P._mar
HoxD5_C._mil
Hoxw5_P._mar
HoxC5_C._mil
HoxC5_L._men
Hoxc5_M._mus
A1
B1
D1
B3
A3
D3
A4
D4
C4
C5
1
2
3
4


HoxA6_C._mil
Hoxa6_M._mus
HoxA6_L._men
HoxB6_C._mil
Hoxb6_M._mus
HoxB7_L._men
Hoxb7_M._mus
HoxA7_C._mil
HoxB7_C._mil
HoxA7_L._men
Hoxa7_M._mus
05_Contig8613_P._mar
06_Contig22678_P._mar
Hoxw6_P.mar
HoxC6_L._men
HoxC6_C._mil
Hoxc6_M._mus
HoxD8_L._men
Hoxd8_M._mus
08_Contig34634_P._mar
HoxB8_C._mil
HoxB8_L._men
Hoxb8_M._mus
08_Contig49403_P._mar
HoxQ8_P._mar
HoxD8_C._mil
HoxC8_C._mil
HoxC8_L._men
Hoxc8_M._mus
A6
C6
B8
8
C8


HoxC9_C._mil
HoxC9_L._men
Hoxc9_M._mus
HoxD9_L._men
HoxD9_C._mil
Hoxd9_M._mus
09_Contig62581_P._mar
Hox9-2_I._fos
HoxA9_C._mil
HoxA9_L._men
Hoxa9_M._mus
HoxB9_L._men
Hoxb9_M._mus
09_Contig38193_P._mar
09_Contig75333_P._mar
Hox9-1_I._fos
HoxB9_C._mil
10_GL493006_P._mar
Hox10_I._fos
10_Contig52464_P._mar
LjHoxW10a_L._cam
LjHox10s_L._cam
HoxB10_C._mil
Hoxa10_M._mus
HoxA10_C._mil
HoxA10_L._men
HoxD10_L._men
HoxD10_C._mil
Hoxd10_M._mus
HoxC10_C._mil
HoxC10_L._men
Hoxc10_M._mus
HoxB10_L._men
C9
D9
A9
B9
C10
9
10

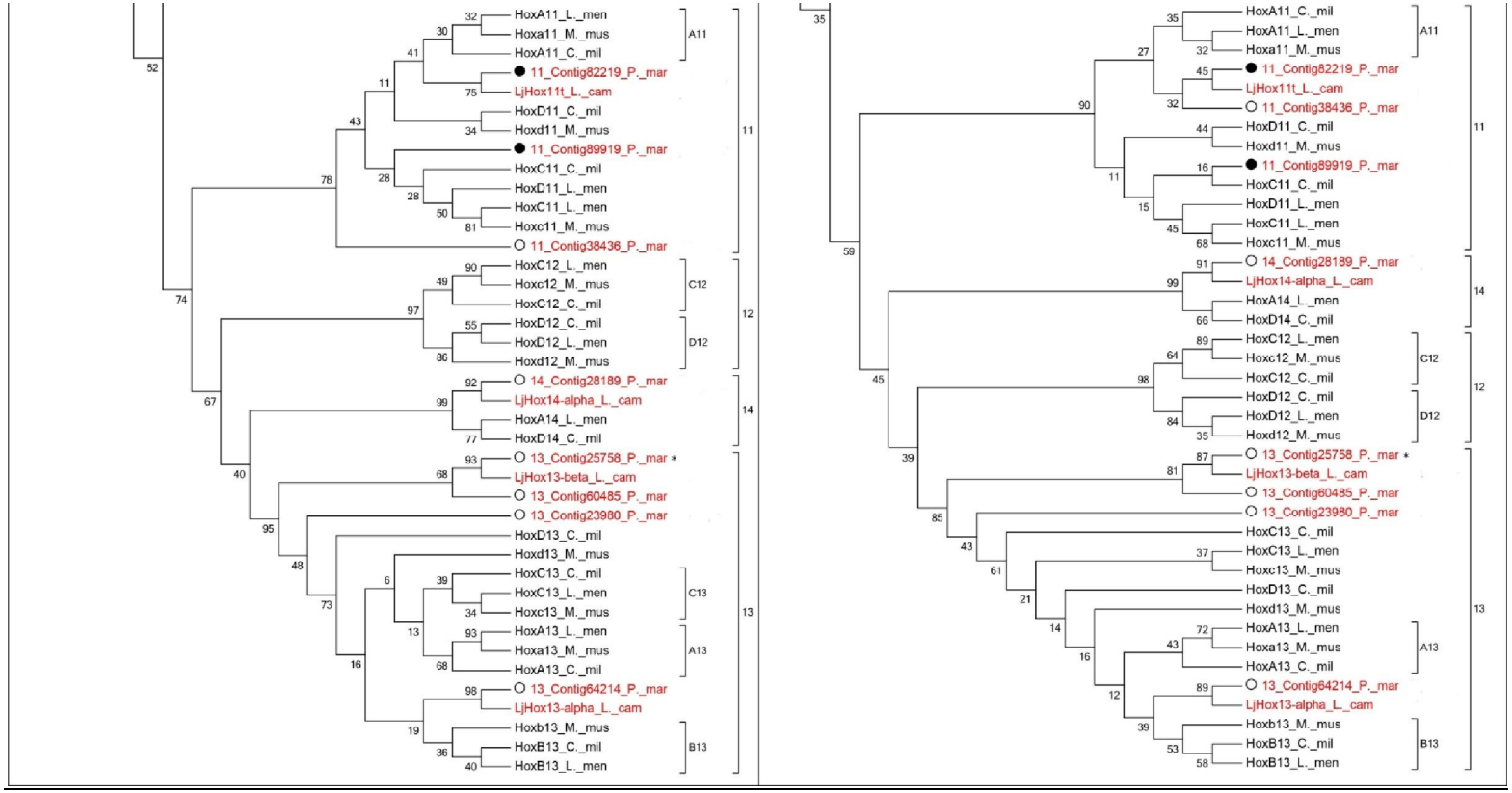


Figure 3.2: *Hox1-Hox14* phylogenetic trees. (A indicates maximum likelihood, B indicates maximum parsimony; Open circle indicates a sequence newly identified (in this study) as a *Hox* gene, closed circle indicates extended sequence on a previously described gene, red indicates a lamprey species; Paralogous group clades indicated where properly separated; Notations along right side of phylogenetic trees indicate paralogous clades; Number values on respective clade indicates bootstrap values, * indicates homeodomain identified in Smith *et al.* 2013, C._mil = *Callorhinchus milii*, M._mus = *Mus musculus*, L._men = *Latimeria menadoensis*, P._mar = *Petromyzon marinus*, L._cam = *Lethenteron camtschaticum*)

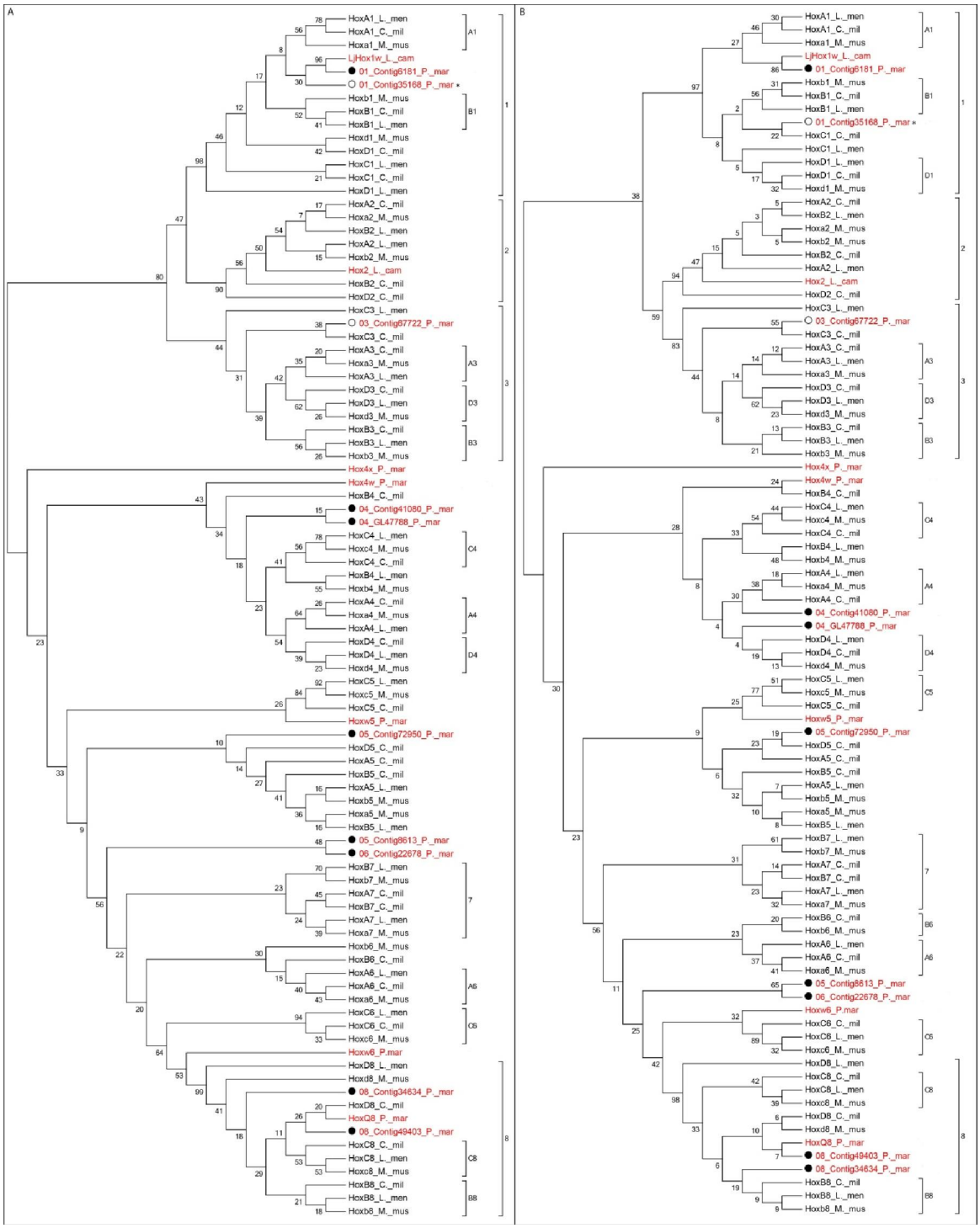


Figure 3.3: *Hox1-Hox8* phylogenetic trees. (A indicates maximum likelihood, B indicates maximum parsimony; Open circle indicates a sequence newly identified (in this study) as a *Hox* gene, closed circle indicates extended sequence on a previously described gene, red indicates a lamprey species; Paralogous group clades indicated where properly separated; Notations along right side of phylogenetic trees indicate paralogous clades; Number values on respective clade indicates bootstrap values, * indicates homeodomain identified in Smith *et al.* 2013, C._mil = *Callorhinchus milii*, M._mus = *Mus musculus*, L._men = *Latimeria menadoensis*, P._mar = *Petromyzon marinus*)

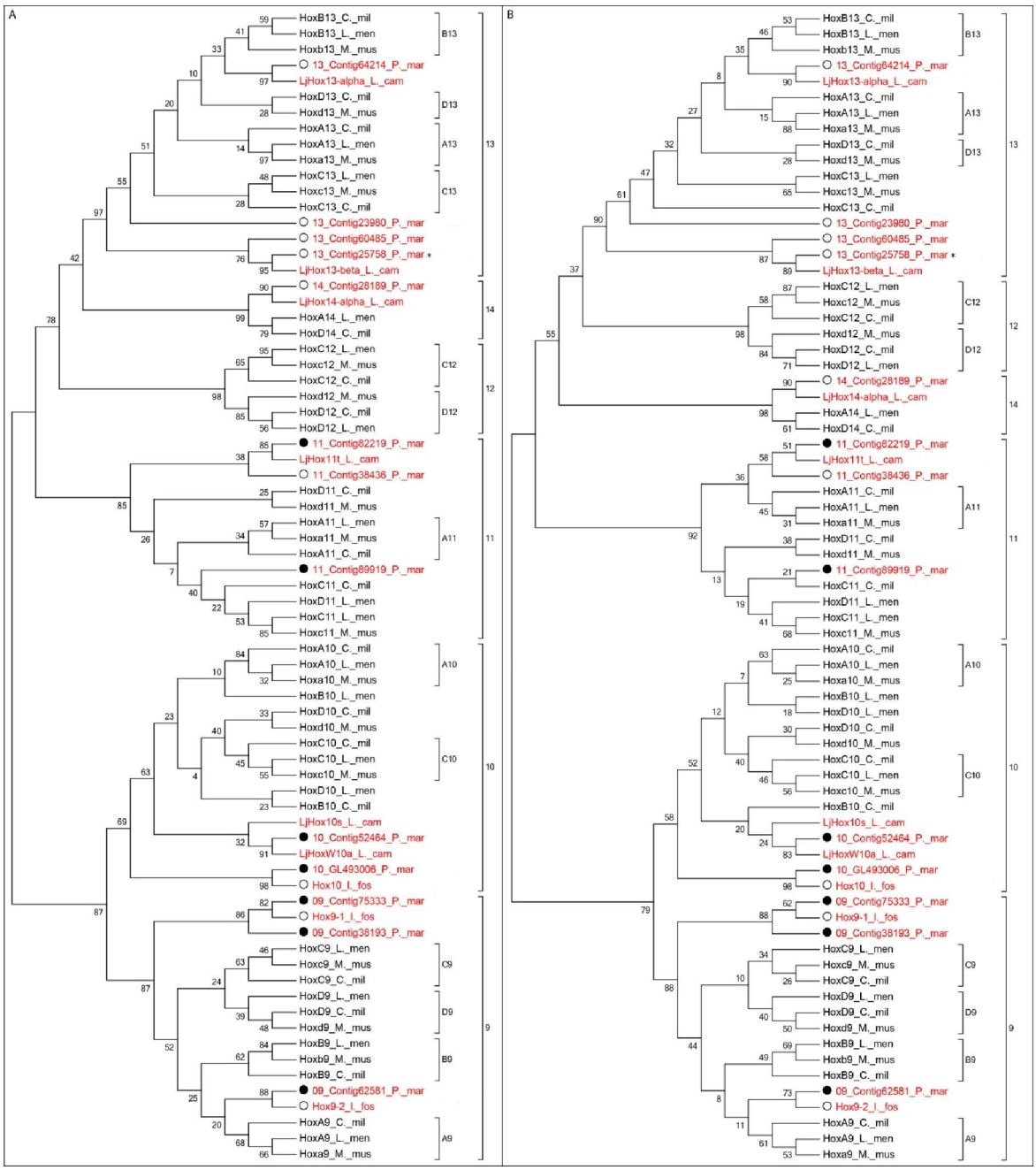


Figure 3.4: *Hox9-Hox14* phylogenetic trees. (A indicates maximum likelihood, B indicates maximum parsimony; Open circle indicates a sequence newly identified (in this study) as a *Hox* gene, closed circle indicates extended sequence on a previously described gene, red indicates a lamprey species; Paralogous group clades indicated where properly separated; Notations along right side of phylogenetic trees indicate paralogous clades; Number values on respective clade indicates bootstrap values, * indicates homeodomain identified in Smith *et al.* 2013, C._mil = *Callorhinchus milii*, I._fos = *Ichtyomyzon fossor*, M._mus = *Mus musculus*, L._men = *Latimeria menadoensis*, P._mar = *Petromyzon marinus*, L._cam = *Lethenteron camtschaticum*)

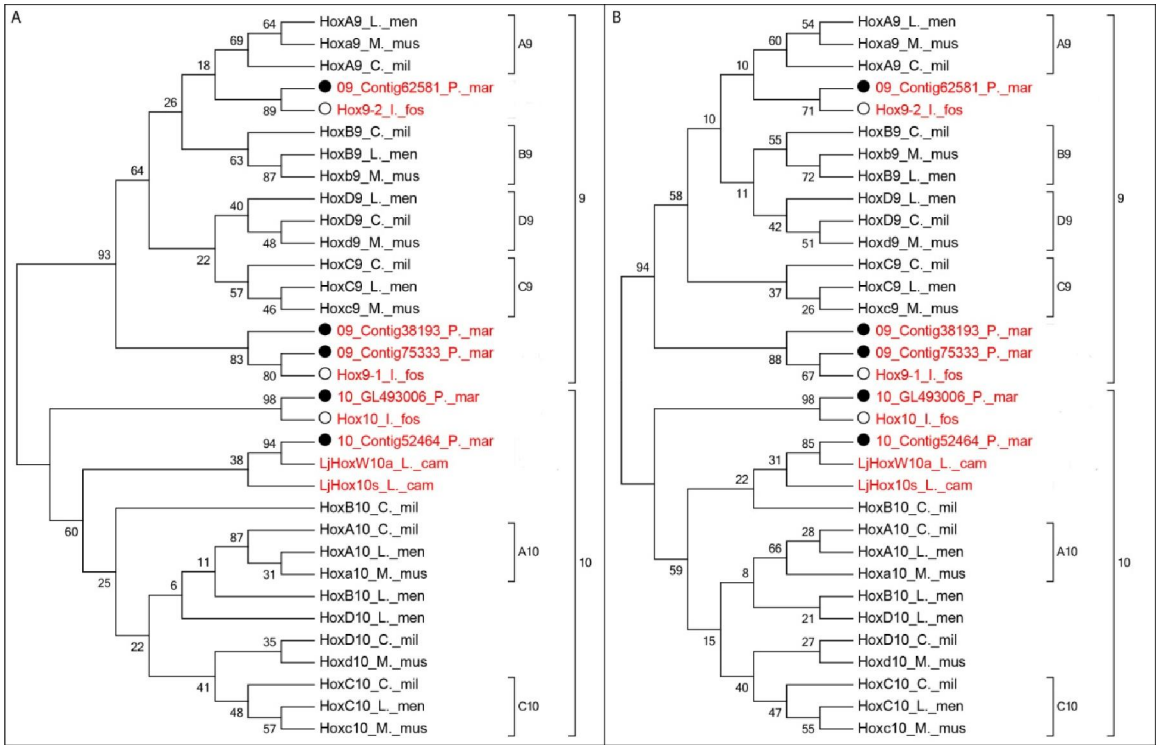


Figure 3.5: *Hox9-Hox10* phylogenetic trees. (A indicates maximum likelihood, B indicates maximum parsimony; Open circle indicates a sequence newly identified (in this study) as a *Hox* gene, closed circle indicates extended sequence on a previously described gene, red indicates a lamprey species; Paralogous group clades indicated where properly separated; Notations along right side of phylogenetic trees indicate paralogous clades; Number values on respective clade indicates bootstrap values, C._mil = *Callorhinchus milii*, I._fos = *Ichtyomyzon fossor*, M._mus = *Mus musculus*, L._men = *Latimeria menadoensis*, P._mar = *Petromyzon marinus*, L._cam = *Lethenteron camtschaticum*)

3.3.2.1 - *Hox1* Gene Clade

The *Hox1* clade was separated from all of the *Hox* genes with a very high degree of confidence (bootstrap values of 95 or greater) for both ML and MP phylogenetic trees in both sets of *Hox1-Hox14* and *Hox1-Hox8* (Figure 3.3). Within the individual *Hox1* phylogenetic trees (Figure 3.6), all paralogous groups formed individual clades, although with low bootstrap values (Table 3.4). Two new sequences found within the *P. marinus* genome were placed within the *Hox1* clade. These two genes were grouped together within a single clade with the pre-existing *Le. camtschaticum LjHox1w* (Accession AB286671) sequence. The *Hox* sequence on Contig6181, which was previously

identified as *P. marinus Hox1w* (Accession AF434665) was placed in the same clade as the *Le. camtschaticum Hox* gene *LjHox1w,* of which in both ML and MP were shown to have bootstrap values greater than 95. The *Hox* sequence on Contig35168 was subsequently placed in a clade with the lamprey *Hox1w* genes yet bootstrap values were less than 35 in both ML and MP phylogenetic trees. This putative *Hox* gene sequence should be considered a newly described *Hox1* gene within *P. marinus*. Curiously, the *Hox* sequence on Contig35168 shows a similarity to the previously identified *HoxQ8* gene (Accession AF410907), which is not to be confused with the other labelled *HoxQ8* gene (Accession AH005896). With a nucleotide similarity of 94.2% within the homeobox region and three non-synonymous mutations, these two genes should be considered paralogous.

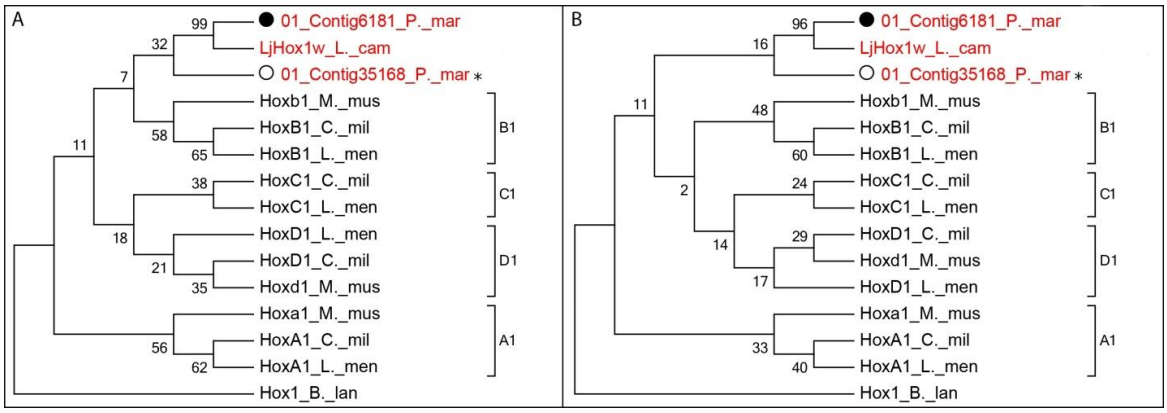


Figure 3.6: *Hox1* phylogenetic trees. (A indicates maximum likelihood, B indicates maximum parsimony; Open circle indicates a sequence newly identified (in this study) as a *Hox* gene, closed circle indicates extended sequence on a previously described gene, red indicates a lamprey species; Paralogous group clades indicated where properly separated; Notations along right side of phylogenetic trees indicate paralogous clades; Number values on respective clade indicates bootstrap values, * indicates homeodomain identified in Smith *et al.* 2013, B._lan = *Branchiostoma lanceolatum*, C._mil = *Callorhinchus milii*, M._mus = *Mus musculus*, L._men = *Latimeria menadoensis*, P._mar = *Petromyzon marinus*, L._cam = *Lethenteron camtschaticum*)

Table 3.4: Bootstrap ratings used to analyze phylogenetic trees.

Bootstrap values	Rating
95-100	very high
90-94	high
80-89	moderate
60-79	low
0-59	very low

3.3.2.2 - *Hox2* Gene Clade

The *Hox2* clade was separated from all of the other *Hox* genes with a moderate to high degree of confidence, with bootstrap values >85 for both ML and MP *Hox1-Hox14* (Figure 3.2) phylogenetic trees and bootstrap values >90 for both ML and MP *Hox1-Hox8* (Figure 3.3) phylogenetic trees. No *Hox* sequences that were found within the *P. marinus* genome were placed within the *Hox2* clade; therefore, no *Hox* phylogenetic trees were produced for this clade.

3.3.2.3 - *Hox3* Gene Clade

The *Hox3* clade was separated from all of the other *Hox* genes with a very low to moderate degree of confidence; a bootstrap value greater than 80 was observed in only the MP phylogenetic trees *Hox1-Hox14* (Figure 3.2) and *Hox1-Hox8* (Figure3.3). The gnathostome *Hox* paralogous groups all formed separate clades, as expected, and showed a moderate to high bootstrap value (Figure 3.7). One sequence found within the *P. marinus* genome was placed within the *Hox3* clade. This putative *Hox* gene on Contig67722 showed 80.6% similarity to that of the previously identified *P. marinus Hox3* gene (Accession AF410909), over 180bp within the homeobox region. This, combined with three non-synonymous mutations, suggests that this gene is a newly described *Hox3* gene in *P. marinus*. Contig67722 was separated within both ML and MP

phylogenetic trees and placed outside of the gnathostome *Hox3* clade, showing a greater divergence from the gnathostome *Hox3* genes.

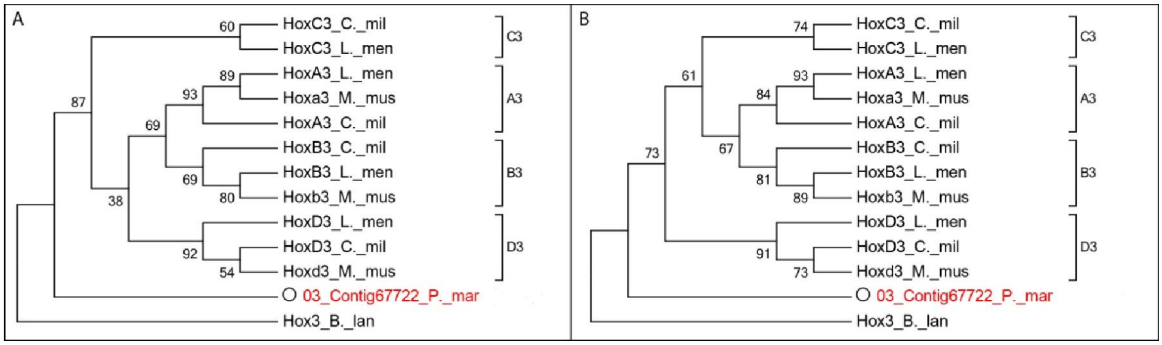


Figure 3.7: *Hox3* phylogenetic trees. (A indicates maximum likelihood, B indicates maximum parsimony; Open circle indicates a sequence newly identified (in this study) as a *Hox* gene, red indicates a lamprey species; Paralogous group clades indicated where properly separated; Notations along right side of phylogenetic trees indicate paralogous clades; Number values on respective clade indicates bootstrap values, B._lan = *Branchiostoma lanceolatum*, C._mil = *Callorhinchus milii*, M._mus = *Mus musculus*, L._men = *Latimeria menadoensis*, P._mar = *Petromyzon marinus*)

3.3.2.4 - *Hox4-Hox7* Gene Clade

Efforts to categorize *Hox* genes *Hox4-Hox7* into individual clades were unsuccessful using phylogenetic analysis with the available amino acid sequence data. The identity of the paralogous groups was too similar to allow for segregation into individual clades for further analysis. Bootstrap values were low, typically under 60, for the *Hox4-Hox7* ML and MP trees (Figure 3.8). Some paralogous groups were able to be separated into individual clades, most notably the *HoxC* paralogous groups, which showed the highest bootstrap values greater than 53. The *HoxA* and *HoxB* paralogous groups were either placed into clades where they were not expected (*Hox5* and *Hox7*) or separated with very low bootstrap values of less than 65 (*Hox4* and *Hox6*) (Figure 3.8). Seven *Hox* sequences found within the *P. marinus* genome were placed within the *Hox4-Hox7* clade. Two of these sequences were direct matches to previously identified *P. marinus* sequence and no new coding sequence could be extrapolated. These two genes

were *Hox4w* (Accession AF434666) and *Hoxw6* (Accession AF071235). Five of these genes were matched with previously identified *Hox* genes from NCBI in which new putative coding sequence could be identified. Contig41080 was matched with *Hox* clone Pm88-h (Accession FSAHOXP88H) from Pendleton *et al*. (1993), GL47788 with *HoxG4* (Accession AF410911), Contig8613 with *HoxF5* (Accession AF410910), Contig22678 with *HoxK6* (Accession AF410913), and Contig72950 with *HoxN5* (Accession AF410915). Sequences Contig41080 and GL47788 are both grouped into the clade consisting of *Hox4* genes. These two *Hox* genes, combined with *Hox4w* and *Hox4x* (Accession AY056469), show a potential of four paralogous *Hox4* genes. These four *Hox* genes show between 82-92% nucleotide identity and 81-95% amino acid identity within the homeobox region, but begin to differ more significantly outside of this region. The few differences observed in these sequences resulted in most groupings within the ML and MP phylogenetic trees to be only supported by weak bootstrap values between 18-58 (Figure3.8). Contig72950 was consistently grouped into the clade consisting of *Hox5* genes in both ML and MP phylogenetic trees, although its bootstrap values were very low (less than 25) and associations within the trees were variable (Figure 3.8). Contig8613 and Contig22678 were consistently grouped together in both ML and MP phylogenetic trees. Both of these genes in the MP phylogenetic tree appear to be grouped with the clade consisting of *Hox6* genes; however, in the ML tree, they appear to be grouped outside of the *Hox5*, *Hox6* or *Hox7* clades. These ambiguous results do not allow for the proper classification of these genes (see Section 4.3.2).

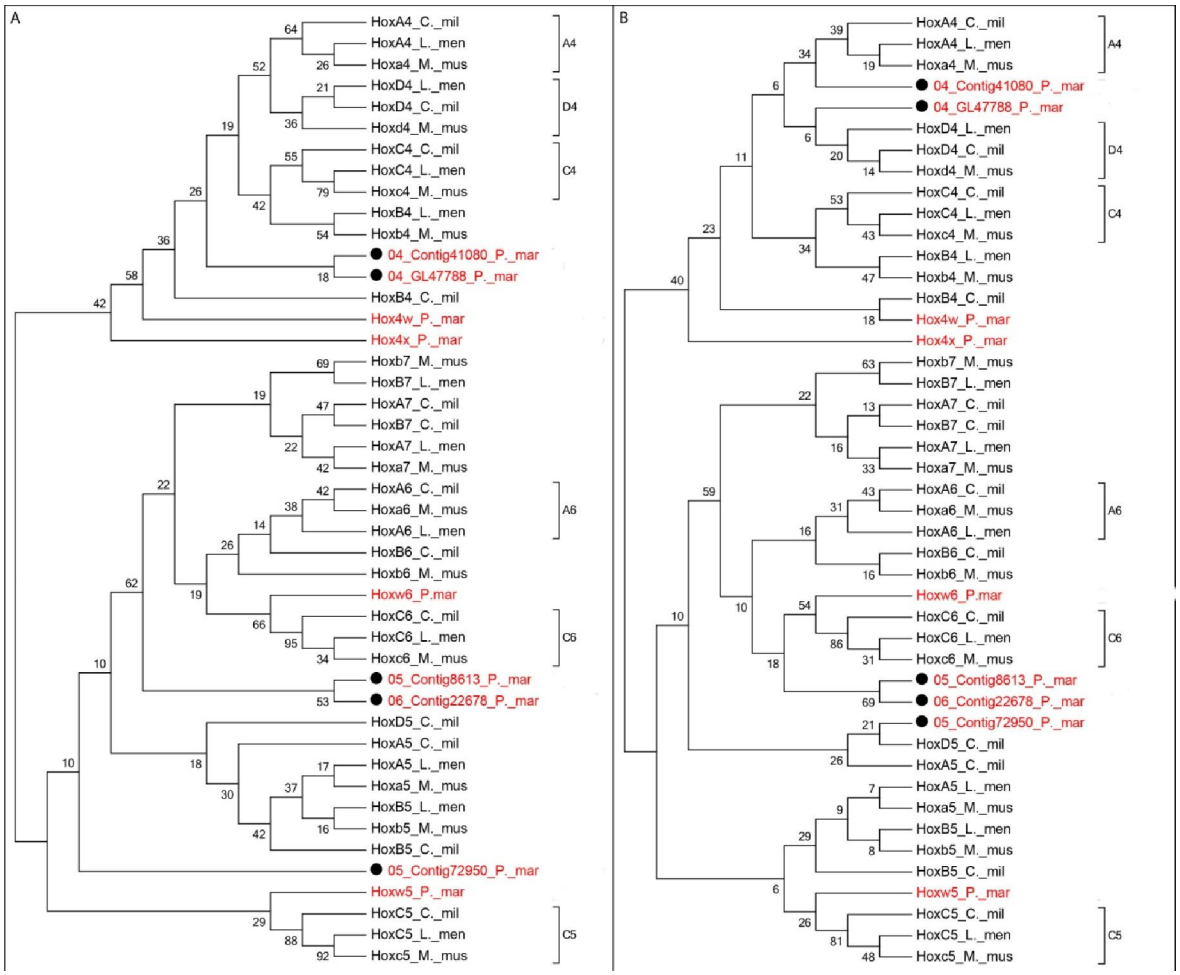


Figure 3.8: *Hox4-Hox7* phylogenetic trees. (A indicates maximum likelihood, B indicates maximum parsimony; Closed circle indicates extended sequence on a previously described gene, red indicates a lamprey species; Paralogous group clades indicated where properly separated; Notations along right side of phylogenetic trees indicate paralogous clades; Number values on respective clade indicates bootstrap values, C._mil = *Callorhinchus milii*, M._mus = *Mus musculus*, L._men = *Latimeria menadoensis*, P._mar = *Petromyzon marinus*)

3.3.2.5 - *Hox8* Gene Clade

The *Hox8* clade was separated from all of the other *Hox* genes with a moderate to very high degree of confidence, showing bootstrap values greater than 85 for ML *Hox1-Hox14* phylogenetic tree and greater than 95 for MP *Hox1-Hox14* and both ML and MP *Hox1-Hox8* phylogenetic trees (Figure 3.2; Figure 3.3). Gnathostome *Hox* paralogous groups *HoxC* and *HoxD* were grouped into individual clades with bootstrap values of 77 and 83 in ML and 57 and 34 in MP phylogenetic trees, whereas the *HoxB* paralogous

group clade in the ML tree only showed a bootstrap value of 27 and in the MP tree it was not properly grouped at all. Three *Hox* sequences found within the *P. marinus* genome were placed within the *Hox8* clade. Contig82182 was matched to previously identified *HoxQ8* (Accession AH005896) and no new putative coding sequence was identified. Contig34634 and Contig49403 were matched with previously identified *HoxR8* (Accession AF035588) and *HoxQ8a* (Accession AF035589) respectively, in which putative new coding sequence was identified. Contig34634 and *HoxQ8* sequences were grouped together in both ML and MP phylogenetic trees, although in both cases bootstrap values were very low with values from 24-37.

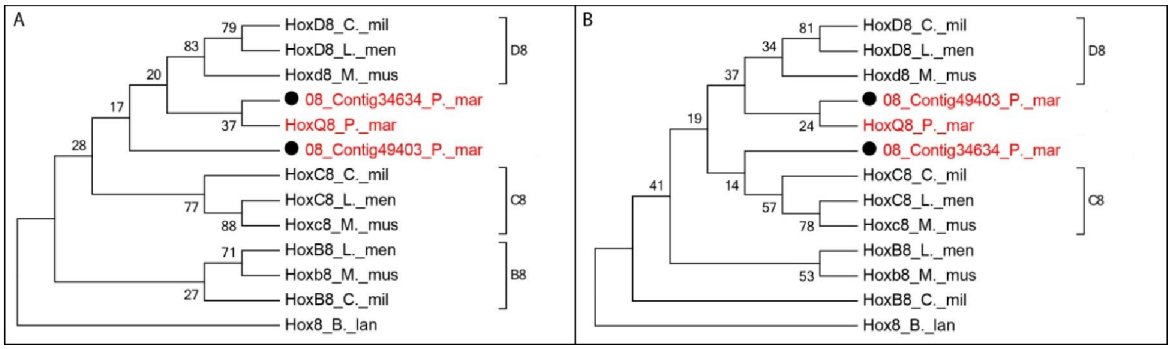


Figure 3.9: *Hox8* phylogenetic trees. (A indicates maximum likelihood, B indicates maximum parsimony; Closed circle indicates extended sequence on a previously described gene, red indicates a lamprey species; Paralogous group clades indicated where properly separated; Notations along right side of phylogenetic trees indicate paralogous clades; Number values on respective clade indicates bootstrap values, * indicates homeodomain identified in Smith *et al.* 2013, B._lan = *Branchiostoma lanceolatum*, C._mil = *Callorhinchus milii*, M._mus = *Mus musculus*, L._men = *Latimeria menadoensis*, P._mar = *Petromyzon marinus*)

3.3.2.6 - *Hox9* Gene Clade

The *Hox9* clade was separated from all of the other *Hox* genes with very low bootstrap values in the ML and MP *Hox1-Hox14* phylogenetic trees, low bootstrap values in ML and MP *Hox9-Hox14* phylogenetic trees and high bootstrap values in ML and MP *Hox9-Hox10* phylogenetic trees (Figure 3.2; Figure 3.4; Figure 3.5). Within the individual *Hox9* phylogenetic trees, all paralogous groups formed separate clades,

although with relatively low bootstrap values ranging from 32-76. Two sequences from the *I. fossor* gonadal transcriptome (see Section 2.3.2) and three sequences found within the *P. marinus* genome were placed within the *Hox9* clade (Figure 3.10). The three lamprey sequences were all matched to previously identified *Hox* genes. Contig75333 was matched to *HoxV9* (Accession AF410919), Contig38193 was matched to *Hox* clone Pm98-s (Accession FSAHOXP98S), and Contig62581 was matched to *HoxT9* (Accession AF410918). The two *I. fossor* sequences showed orthology to the genes identified in *P. marinus*. The first *Hox9* sequence Hox9-1_I.fos was matched with Contig75333 showing bootstrap values of 90 and 64 for ML and MP phylogenetic trees respectively. The Hox9-2_I._fos sequence was matched with Contig62581 showing bootstrap values of 84 and 50 for ML and MP phylogenetic trees.

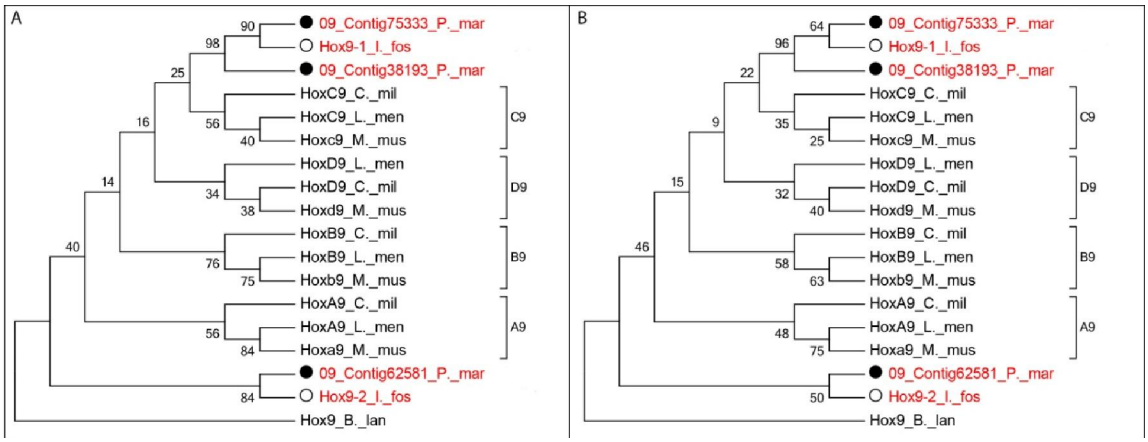


Figure 3.10: *Hox9* phylogenetic trees. (A indicates maximum likelihood, B indicates maximum parsimony; Open circle indicates a sequence newly identified (in this study) as a *Hox* gene, closed circle indicates extended sequence on a previously described gene, red indicates a lamprey species; Paralogous group clades indicated where properly separated; Notations along right side of phylogenetic trees indicate paralogous clades; Number values on respective clade indicates bootstrap values, B._lan = *Branchiostoma lanceolatum*, C._mil = *Callorhinchus milii*, I._fos = *Ichtyomyzon fossor*, M._mus = *Mus musculus*, L._men = *Latimeria menadoensis*, P._mar = *Petromyzon marinus*)

3.3.2.7 - *Hox10* Gene Clade

The *Hox10* clade was separated from all of the other *Hox* genes with very low bootstrap values in the ML and MP *Hox1-Hox14* phylogenetic trees, low bootstrap values in ML and MP *Hox9-Hox14* phylogenetic trees and low bootstrap values in ML and MP *Hox9* and *Hox10* phylogenetic trees (Figure 3.11). Within the individual *Hox10* ML and MP phylogenetic trees only *HoxA* and *HoxC* paralogous groups were separated into clades, although with low bootstrap values. *HoxB* and *HoxD* paralogous groups were not separated into clades within the phylogenetic trees. One sequence from the *I. fossor* transcriptome and two sequences found within the *P. marinus* genome were placed within the *Hox10* clade. Two previously described *Le. camtschaticum* sequences, *LjHoxW10a* (Accession AB286672) and *LjHox10s* (Accession AB286673), were also added to the *Hox10* phylogenetic trees to help identify sequence similarities. The two *P. marinus* sequences identified in the *P. marinus* genome search were both matched with previously identified genes. Contig52464 was matched with *HoxW10a* (Accession AF410920) and GL493006 was matched with *HoxX10* (Accession AF410922). The *I. fossor* sequence Hox10_I._fos is newly identified. GL493006 and Hox10_I._fos in both ML and MP phylogenetic tree are grouped together and show a very high similarity to each other (with bootstrap values of 99 and 100, respectively). These two genes are also loosely placed into a clade with *HoxA* paralogous group genes. Contig52464 is grouped with *LjHoxW10a* in the same clade and show bootstrap values of 98 and 93 for ML and MP respectively, suggesting these are possibly orthologous genes. *LjHox10s* was not grouped with any *P. marinus* gene sequence suggesting three different *Hox10* paralogous groups are present within these phylogenetic trees.

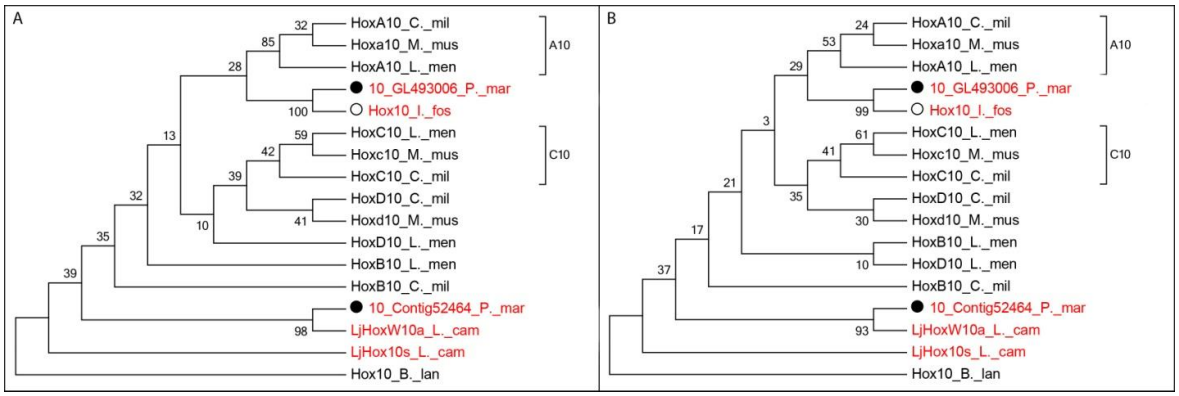


Figure 3.11: *Hox10* phylogenetic trees. (A indicates maximum likelihood, B indicates maximum parsimony; Open circle indicates a sequence newly identified (in this study) as a *Hox* gene, closed circle indicates extended sequence on a previously described gene, red indicates a lamprey species; Paralogous group clades indicated where properly separated; Notations along right side of phylogenetic trees indicate paralogous clades; Number values on respective clade indicates bootstrap values, B._lan = *Branchiostoma lanceolatum*, C._mil = *Callorhinchus milii*, I._fos = *Ichtyomyzon fossor*, M._mus = *Mus musculus*, L._men = *Latimeria menadoensis*, P._mar = *Petromyzon marinus*, L._cam = *Lethenteron camtschaticum*)

3.3.2.8 - *Hox11* Gene Clade

The *Hox11* clade was separated from all of the other *Hox* genes with low to moderate bootstrap values in the ML and MP *Hox1-Hox14* phylogenetic trees and moderate to high bootstrap values in ML and MP *Hox9-Hox14* phylogenetic trees (Figure 3.2; Figure 3.4). Within the individual *Hox11* phylogenetic trees all paralogous groups formed separate clades, although with low bootstrap values < 55 (Figure 3.12). Three sequences found within the *P. marinus* genome were placed within the *Hox11* clade. Two of these were matched to previously identified *Hox11* genes and one gene, Contig38436, appeared to be a putative new *Hox11* gene. Contig89919 was a match to *HoxZ11b* (Accession AF410925) and Contig82219 was a match to *HoxY11* (Accession AF410923). One previously identified *Le. camtschaticum Hox* gene *LjHox11t* (Accession AB286674) was also added. All four lamprey sequences were grouped together in both ML and MP phylogenetic trees. Contig89919 appeared most divergent from the other sequences and was only connected with a very low bootstrap value < 48, while the other three showed a

moderate bootstrap support of 75-88. Contig82219 was most closely related to *LjHox11t* within both ML and MP phylogenetic trees and shared a moderate/low bootstrap of 90 and 62 respectively.

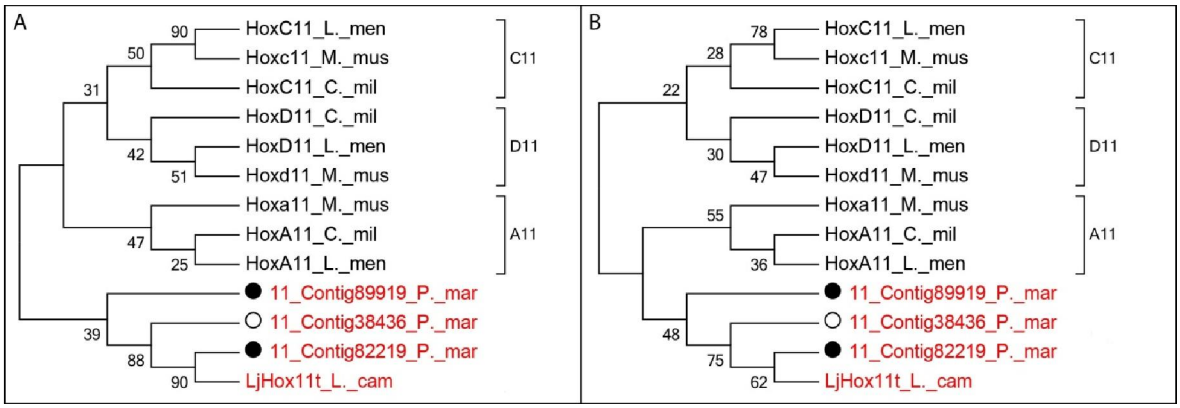


Figure 3.12: *Hox11* phylogenetic trees. (A indicates maximum likelihood, B indicates maximum parsimony; Open circle indicates a sequence newly identified (in this study) as a *Hox* gene, closed circle indicates extended sequence on a previously described gene, red indicates a lamprey species; Paralogous group clades indicated where properly separated; Notations along right side of phylogenetic trees indicate paralogous clades; Number values on respective clade indicates bootstrap values, C._mil = *Callorhinchus milii*, M._mus = *Mus musculus*, L._men = *Latimeria menadoensis*, P._mar = *Petromyzon marinus*, L._cam = *Lethenteron camtschaticum*)

3.3.2.9 - *Hox12* Gene Clade

The Hox12 clade was separated from all of the other *Hox* genes with very high degree of confidence, showing greater than a 95 bootstrap value for both ML and MP phylogenetic trees for both *Hox1-Hox14* and *Hox9-Hox14*. No *Hox* sequences that were found within the *P. marinus* genome were placed within the *Hox12* clade; therefore, no *Hox* phylogenetic trees were produced for this clade.

3.3.2.10 - *Hox13* Gene Clade

The *Hox13* clade was separated from all of the other *Hox* genes with very high/moderate bootstrap values in the ML and MP *Hox1-Hox14* phylogenetic trees and very high/high bootstrap values in ML and MP *Hox9-Hox14* phylogenetic trees (Figure 3.2; Figure 3.4). Within the individual *Hox13* ML phylogenetic tree all paralogous groups

formed separate clades, although with low bootstrap values < 50 (Figure 3.13). In the *Hox13* MP phylogenetic tree, only *HoxA* and *HoxB* paralogous groups separated as expected into individual clades. Four sequences found within the *P. marinus* genome were placed within the *Hox11* clade. All four of these putative sequences (Contig64214, Contig23980, Contig60485 and Contig25758) are previously undescribed as *Hox* genes. Two previously identified *Le. camtschaticum Hox* genes *LjHox13-alpha* (Accession AB293597) and *LjHox13-beta* (Accession AB293598) were also included in the *Hox11* clade. In both ML and MP phylogenetic trees Contig64214 was grouped into the *HoxB13* paralogous group along with a high bootstrap association to *LjHox13-alpha*, suggesting this new sequence could be orthologous to this *Le. camtschaticum Hox* gene. Contig60485 and Contig25758 were grouped together with *LjHox13-beta* in both ML and MP phylogenetic trees with bootstrap values of 68 and 87 for ML and MP phylogenetic trees respectively. Contig25758 was subsequently grouped with *LjHox13-beta* and are most closely related to each other in both ML and MP trees with bootstrap values of 94 and 89. The placement of both Contig25758 and *LjHox13-beta* together in both phylogenetic trees, as well as the presence of an intron within the homeobox regions of both, suggests orthology between these two genes. Contig60485, while related to Contig25758 and *LjHox13-beta*, does not contain a secondary intron, which suggests this putative *Hox* sequence is not orthologous with Contig25758. Contig23980 was placed individually within both ML and MP phylogenetic trees suggesting this putative *Hox13* gene is also a unique sequence. Altogether four newly described putative *Hox13* are shown within the *Hox13* gene clade.

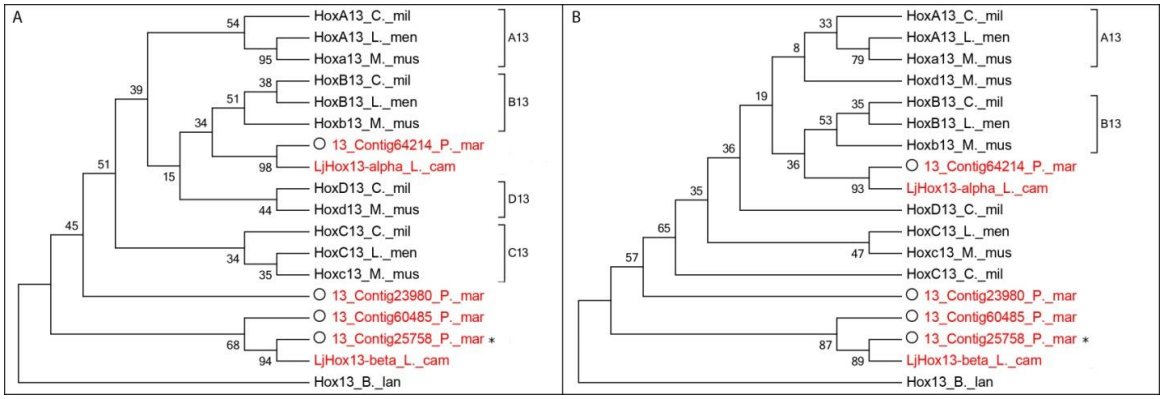


Figure 3.13: *Hox13* phylogenetic trees. (A indicates maximum likelihood, B indicates maximum parsimony; Open circle indicates a sequence newly identified (in this study) as a *Hox* gene, , red indicates a lamprey species; Paralogous group clades indicated where properly separated; Notations along right side of phylogenetic trees indicate paralogous clades; Number values on respective clade indicates bootstrap values, * indicates homeodomain identified in Smith *et al.* 2013, B._lan = *Branchiostoma lanceolatum*, C._mil = *Callorhinchus milii*, M._mus = *Mus musculus*, L._men = *Latimeria menadoensis*, P._mar = *Petromyzon marinus*, L._cam = *Lethenteron camtschaticum*)

3.3.2.11 - *Hox14* Gene Clade

The *Hox14* clade was separated from all of the *Hox* genes with very high bootstrap values in the ML and MP *Hox1-Hox14* and *Hox9-Hox14* phylogenetic trees (Figure 3.2; Figure 3.4; Figure 3.14). One sequence (Contig28189) found within the *P. marinus* genome was placed within the *Hox14* clade and was newly identified as a *Hox* gene. Contig28189 was grouped with *LjHox14-alpha* (Accession AB293599) with high similarity in both ML and MP phylogenetic trees, suggesting possible orthology between the two genes.

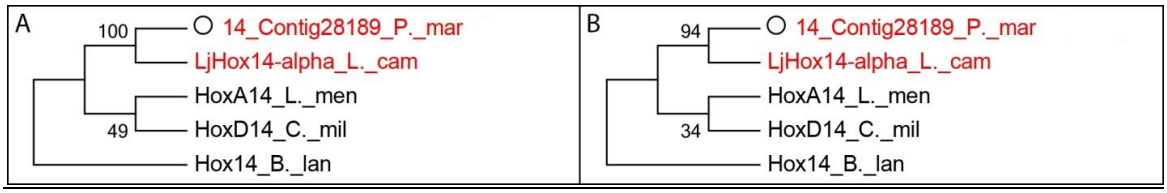


Figure 3.14: Hox14 phylogenetic trees. (A indicates maximum likelihood, B indicates maximum parsimony; Open circle indicates a sequence newly identified (in this study) as a *Hox* gene, red indicates a lamprey species; Number values on respective clade indicates bootstrap values, B._lan = *Branchiostoma lanceolatum*, C._mil = *Callorhinchus milii*, L._men = *Latimeria menadoensis*, P._mar = *Petromyzon marinus*, L._cam = *Lethenteron camtschaticum*)

3.4 - *Petromyzon marinus Hox* Gene Verification

Eight solitary exon1 sequences were identified within the *P. marinus* genome analysis. Sequence identification was achieved in three different ways. Exon1 sequences on Contig68831 (*Hox2*), Contig30770 (*Hox13*) and GL486262 (*Hox14*) were identified based on sequences previously identified in *Le. camtschaticum*. In contrast, exon1 sequences on GL476758 (*Hox7*), Contig34634 (*HoxR8*) and GL477571 (*Hox8*) were identified based on the search for 200-700bp ORFs combined with a similar sequence motif within the 3′ region of exon1. Searches for these latter sequences were performed on contigs with existing *Hox* sequence from other *Hox* genes. Finally Contig64676 (*Hox4w*) and Contig82182 (*HoxQ8*) were identified based on the previously identified *P. marinus* sequences. Exon2 was identified in both Contig64676 (*Hox4w*) and Contig34634 (*HoxR8)*; however, both reading frames were disrupted by nucleotide insertions and appeared to be non-functional. In the case of *HoxQ8*, there was two different contigs, one from genome1 (Genome Institute at Washington University) and the other from genome2 (Ensembl Genome Browser). Both lacked exon2, because in both cases, the sequence of exon1 was found at the 3′ end of the sequence read of the contig. Unfortunately, exon2 to *HoxQ8* was not identified in any other contigs.

Primers designed in Section 2.4 targeted four of the five remaining contigs: Contig68831 (*Hox2*); GL476758 (*Hox7*); GL477571 (*Hox8*); and Contig30770 (*Hox13*). No primers were designed for GL486262 (*Hox14*) because it was discovered after the *Hox* gene verification experiments were complete. This exon1 sequence along with Contig28189 (putative exon2 and exon3 of *P. marinus Hox14*) show orthology to *Le. camtschaticum LjHox14-alpha* gene, suggesting GL486262 and Contig28189 contain

exons from the same gene. As a result, confirmation of the *Hox14* gene was not undertaken. Designed primers targeted a 3′ region of the putative exon1 sequence and all of these primers were used in conjunction with the reverse RACE primers from Section 2.2.4. PCR amplification using the *Hox* forward and RACE reverse primers with *P. marinus* genomic DNA revealed a strong positive band for *Hox2* sequence while *Hox7*, *Hox8*, and *Hox13* showed faint bands. DNA fragments from each of the four PCR reactions were extracted and then cloned (Section 2.1.4). No bacterial clones were obtained for *Hox8* or *Hox13* cloning reactions, and the *E. coli* colonies for the *Hox7* cloning reactions lacked any *Hox* gene sequences. DNA fragments for the *Hox7*, *Hox8* and *Hox13* samples were fainter when observed on an agarose gel than the *Hox2* DNA fragments, suggesting that the fragments were not the intended targets or that the primers were not specific enough to amplify the exon-intron-exon DNA fragments. *Hox2* did show two potential positive colonies from PCR colony screens, which were subsequently sequenced. One of the two colonies showed positive sequence for that *Hox2* paralogous group. Exon1 of the sequence was matched to exon1 nucleotide sequence from Contig68831, which showed a 100% nucleotide identity to that of the sequence sample. Exon2 was aligned against *Hox2* (Accession AY497314) from *Le. camtschaticum* and showed a 98.3% nucleotide identity across 177bp and a 100% amino acid identity. The intron size was estimated at approximately 1600bp in length. The remaining three *Hox* sequences (*Hox7*, *Hox8* and *Hox13*) were not isolated from genomic samples. PCR optimization, using various $MgCl_2$ concentrations and annealing temperatures, produced multiple PCR products, but none of these contained *Hox* sequences. A single primer PCR

control showed that the observed bands thought to be potential positives for exon-intron-exon amplicons were a consequence of unspecific primer binding.

3.5 - Identification of Lamprey *Hox* Genes in Multiple Lamprey Genera

3.5.1 - Amplification of Exon1 in Posterior *Hox* Genes

Degenerate primers aimed at the 3′ coding region of exon1, designed from alternative species orthologous *Hox9, Hox10* and *Hox11* genes, failed to produce any amplicon (Tables 2.6; 2.10). Multiple attempts to optimize PCR conditions also failed to produce any amplicons. After exhausting PCR optimization techniques, no further experiments were conducted using these sets of primers.

3.5.2 - *Hox10* Gene Analysis Across Multiple Lamprey Genera

LjHox10s orthologs were isolated using primers designed from *Le. camtschaticum* (Table 2.11) from nine different lamprey species: *En. tridentatus*, *Eu. vladykovi*, *G. australis*, *I. castaneus*, *I. fossor*, *I. unicuspis*, *La. aepyptera*, *La. pacifica*, and *La. richardsoni*. A total of 226bp of putative nucleotide sequence was sequenced and identified from each individual lamprey consisting of the 180bp homeodomain and 27bp 5′ to the homeodomain and 19bp 3′ of the homeodomain. Sequence alignments showed a >90% nucleotide similarity (Figure 3.15) and >94% amino acid similarity (Figure 3.16) to the *LjHox10s* sequence. Nucleotide ML and MP phylogenetic trees were produced using the nine putative newly identified lamprey *LjHox10s* orthologs along with *LjHox10s* sequence (Figure 3.17). Both ML and MP phylogenetic trees produced similar results with *G. australis* rooted as the most distantly related lamprey in both phylogenetic trees. All three *Ichthyomyzon* species were grouped into a clade with bootstrap values >90 in both ML and MP trees. The *Lampetra* species were grouped into their own clade

with low bootstrap values <30. Each of the *Eudontomyzon* and *Lethenteron* species were also loosely grouped together with a low bootstrap values <40 and in turn were grouped with the *Lampetra* species with bootstrap values <50. *Entosphenus tridentatus* was placed to the exterior of the northern hemisphere lamprey clade, with bootstrap values of 69 and 87 for ML and MP respectively.

Orthologous *LjHox10s* sequences were compared to lamprey trunk myomere counts obtained from previous studies (Seversmith 1953; Strahan 1960; Neira 1984; Docker 2009). Differences in nucleotide and amino acid sequence within the coding region were examined in all lamprey sequences obtained above. The lamprey nucleotide and amino acid alignments were arranged from high - low trunk myomere counts and comparisons between high/high, high/low and low/low myomere counts were conducted (Figure 3.15; Figure 3.16). No observable sequence variation was consistent across any of the comparable myomere counts. Nucleotide comparisons of the orthologous *LjHox10s* sequences to *Le. camtschaticum* differed no more than 8.8%, which was seen in *G. australis* (Table 3.5). Amino acid comparisons of the orthologous *LjHox10s* sequences to *Le. camtschaticum* differed no more than 5.3%, which was seen in *En. tridentatus* (Table 3.5). The highest variation in nucleotide and amino acid was observed between lamprey species with the highest trunk myomere counts, whereas, the least amount of variation was seen when comparing low trunk myomere count species to *Le. camtschaticum LjHox10s* sequence. The lack of any definitive sequence similarities or identifiable synapomorphies between the high and low trunk myomere species within the *Hox10* sequences examined suggests that the 226bp region examined does not have a correlation

with the number of trunk myomeres present in lamprey or that any functional differences are overwhelmed by phylogenetic differences.

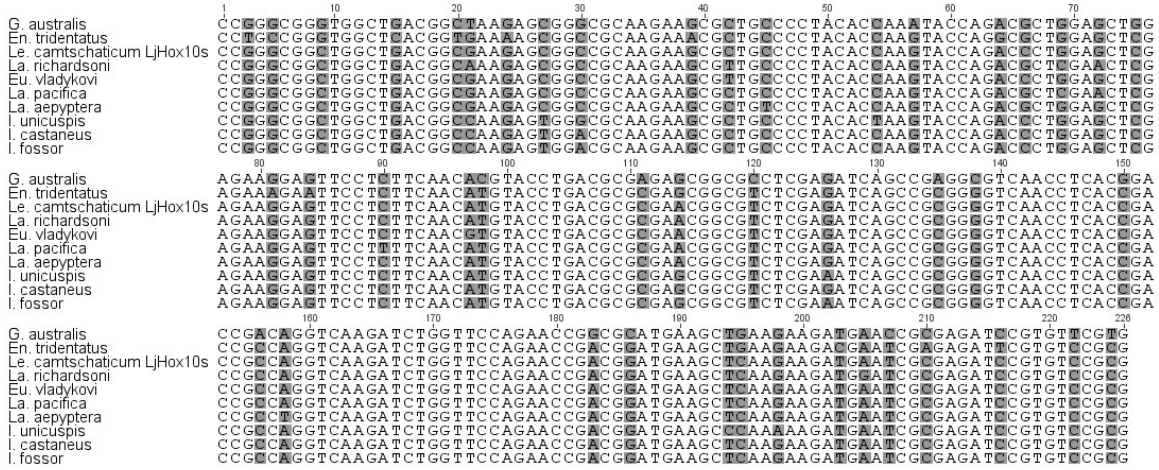


Figure 3.15: *Hox10s* ortholog nucleotide alignment (En = *Entosphenus*, Eu = *Eudontomyzon*, G = *Geotria*, I = *Ichthyomyzon,* La = *Lampetra*, Le = *Lethenteron*, Grey indicates regions of polymorphisms, sequences arranged from high - low myomere counts)

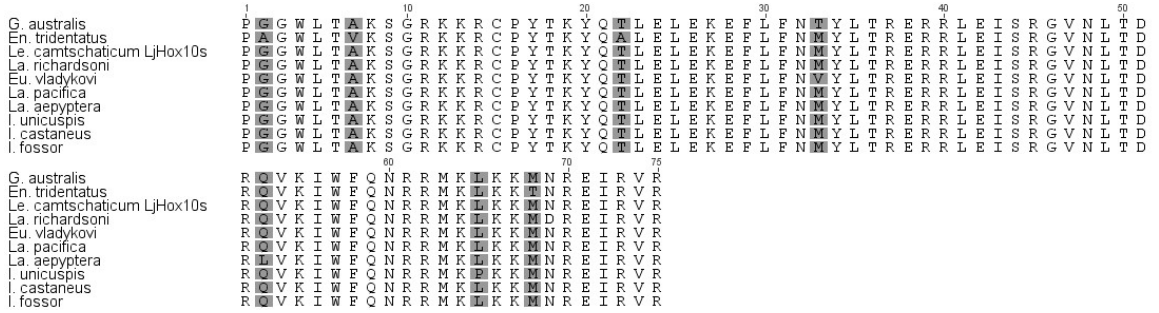


Figure 3.16: *Hox10s* ortholog amino acid alignment (En = *Entosphenus*, Eu = *Eudontomyzon*, G = *Geotria*, I = *Ichthyomyzon,* La = *Lampetra*, Le = *Lethenteron*, Grey indicates regions of polymorphisms)

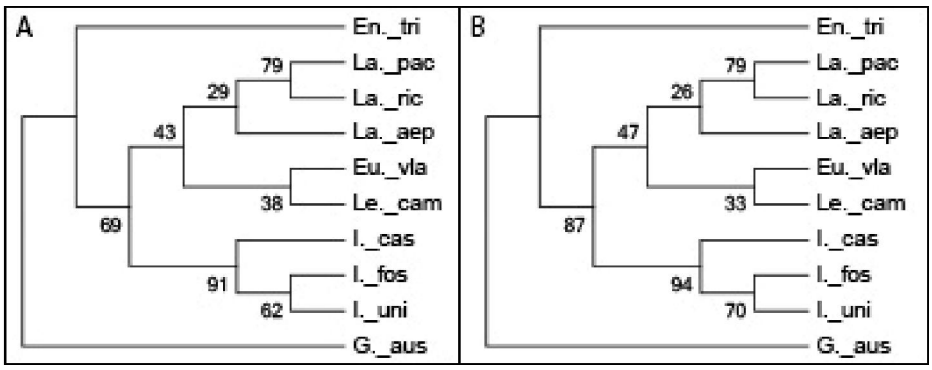


Figure 3.17: *Hox10s* ortholog phylogenetic trees. A = maximum likelihood, B = maximum parsimony (En = *Entosphenus*, Eu = *Eudontomyzon*, G = *Geotria*, I = *Ichthyomyzon,* La = *Lampetra*, Le = *Lethenteron*)

Table 3.5: Lamprey myomere counts and nucleotide/amino acid similarity of *LjHox10s* orthologs. (En. = *Entosphenus*, Eu. = *Eudontomyzon*, G. = *Geotria*, I. = *Ichthyomyzon,* La. = *Lampetra*, Le. = *Lethenteron*)

Species	Nucleotide similarity (%)	Amino acid similarity (%)	Trunk myomere counts	High or low myomere count
G. australis	91.2	98.7	66-78	High
En. tridentatus	97.3	94.7	61-77	High
Le. camtschaticum	100	100	60-74	High
La. richardsoni	92.9	98.7	58-67	High
Eu. vladykovi	99.1	98.7	58-68	High
La. pacifica	98.2	100	54-58	Low
La. aepyptera	98.7	98.7	53-60	Low
I. unicuspis	96.5	98.7	49-56	Low
I. castaneus	98.2	100	47-56	Low
I. fossor	97.3	100	47-55	Low

3.5.3 - General *Hox* Gene Screening in Lampreys

Following analysis of the *P. marinus* genome for the presence of *Hox* genes, an effort was made to determine efficiency of identifying *Hox* gene complements within other lamprey species. Primers were designed based on the studies by Pendleton *et al*. (1993) and Force *et al*. (2002), which identified *Hox* gene families in a variety of chordates including some lamprey sequences and *P. marinus* respectively (Table 2.12). Combinations of the primers were initially tested on *I. fossor* to determine efficacy of the primer sets (Table 2.12c). All four primer sets yielded products of expected size when

samples were resolved on agarose gels. The HoxE - HoxF primer set was selected for further experiments as it produced the largest amplicon. Following this, putative gene targets were isolated from nine species of lampreys: *En. tridentatus, Eu. vladykovi, G. australis, I. castaneus, I. fossor, I. gagei, I. unicuspis, La. pacifica and La. richardsoni*. At least two samples were sequenced from each species, except in *I. fossor* where eight samples were sequenced. The final sequencing results produced 30 positive sequence reads of 81bp within the homeodomain of the *Hox* genes across the different species (Table 3.6). All of these sequences are newly identified putative *Hox* gene fragments in these species. Sequences were provided with a preliminary *Hox* match base on sequence identity when analyzed using BLAST. Although assigning the putatively identified *Hox* gene fragments was difficult due to the small length of nucleotides, the HoxE - HoxF primer set appeared to be capable of amplifying the majority of the paralogous groups. With a large effort of amplifying, screening and sequencing it appears that a genomic analysis of *Hox* genes within these lamprey species may be possible.

Table 3.6: *Hox* gene identification in multiple lamprey species using degenerate primers HoxE - HoxF. (*Hox* match based on sequence similarity using BLAST)

Individual #	Sample name	Species	*Hox* match
1	34-4	*Entosphenus tridentatus*	*Hox3*
2	40-8	*Entosphenus tridentatus*	*Hox8*
1	B10-3	*Eudontomyzon vladykovi*	*Hox4/5*
2	D10-1	*Eudontomyzon vladykovi*	*Hox4/5/6*
1	2-1	*Geotria australis*	*Hox6/7*
2	3-1	*Geotria australis*	*Hox6/14*
3	4-3	*Geotria australis*	*Hox2*
3	4-7	*Geotria australis*	*Hox2*
1	CL1	*Ichthyomyzon castaneus*	*Hox4/5/6/7*
1	CL3	*Ichthyomyzon castaneus*	*Hox4/5/6/7*
1	NBL1	*Ichthyomyzon fossor*	*Hox1/8*
1	NBL2	*Ichthyomyzon fossor*	*Hox8*
1	NBL3	*Ichthyomyzon fossor*	*Hox1/8*
1	NBL4	*Ichthyomyzon fossor*	*Hox5/6/7*
1	NBL5	*Ichthyomyzon fossor*	*Hox1/8*
1	NBL6	*Ichthyomyzon fossor*	*Hox1/8*
1	NBL7	*Ichthyomyzon fossor*	*Hox1/8*
1	NBL8	*Ichthyomyzon fossor*	*Hox1/8*
1	G10	*Ichthyomyzon gagei*	*Hox5/6*
1	G12	*Ichthyomyzon gagei*	*Hox5/6*
1	G13	*Ichthyomyzon gagei*	*Hox5/6*
1	G14	*Ichthyomyzon gagei*	*Hox5/6*
1	SL3	*Ichthyomyzon unicuspis*	*Hox4/5*
1	SL6	*Ichthyomyzon unicuspis*	*Hox4/5*
1	P3-3	*Lampetra pacifica*	*Hox4/5/6/7*
1	P3-6	*Lampetra pacifica*	*Hox10*
1	R1-4	*Lampetra richardsoni*	*Hox3*
1	R1-5	*Lampetra richardsoni*	*Hox11*
1	R1-6	*Lampetra richardsoni*	*Hox11*
2	R2-6	*Lampetra richardsoni*	*Hox11*

4. Discussion

As agnathans, lampreys serve as very important models in understanding the evolution of vertebrates. Agnathans are one of only two extant vertebrate groups that lack a hinged jaw and represent an important evolutionary step in vertebrate evolution (Osorio and Retaux 2008). An understanding of the diversity and functions of *Hox* genes, which are essential to the anterior-posterior development during embryogenesis, will be a key factor in helping to understand these ancestral vertebrates. In this study, I examined the *P. marinus* genome databases (Genome Institute at Washington University (Genome1; http://genome.wustl.edu/genomes/detail/petromyzon-marinus/) and the Ensembl Genome Browser (Genome2; release 70 - January 2013; http://uswest.ensembl.org/Petromyzon_marinus/Info/Index) for the presence of *Hox* genes. In this search, I discovered multiple new putative *Hox* genes including four putative *Hox* genes within the 13th paralogous group. I examined the diversity of *Hox* genes within *P. marinus* and compared them to other orthologous *Hox* genes within other lamprey species (Amores *et al.* 1998; Carr *et al.* 1998, Sharman and Holland 1998; Cohn 2002; Force *et al.* 2002; Irvine *et al.* 2002; Takio *et al.* 2004; Takio 2007; Kuraku *et al.* 2008). I also examined posteriorly-expressed *Hox10* genes and their potential role in regulating differences in trunk myomere numbers among different lamprey species. Finally, I looked at the potential identification of some putative representative *Hox* genes in other lamprey species.

4.1 - Posterior *Hox* Gene Amplification

Few studies have examined *Hox* genes in lampreys and aside from the homeobox region, relatively little is even known about their sequences within the genomes of most

lamprey species (Amores *et al.* 1998; Carr *et al.* 1998, Sharman and Holland 1998; Cohn 2002; Force *et al.* 2002; Irvine *et al.* 2002; Takio *et al.* 2004; Takio 2007; Kuraku *et al.* 2008). The homeobox is a relatively well-conserved sequence throughout many vertebrate taxa (Holland and Garcia-Fernàndez 1996), and has therefore facilitated its identification in a growing number of species (see Mallo *et al*. 2010; Durston *et al.* 2011; Pick and Heffer 2012). Gene sequences flanking the homeobox region, however, can diverge considerably, making it more difficult to describe full length *Hox* gene sequences. To identify posterior *Hox* genes in the local lamprey species *Ichthyomyzon castaneus* (chestnut lamprey)*, I. fossor* (northern brook lamprey) and *I. unicuspis* (silver lamprey), primers were designed to the homeobox region to allow for the highest complementarity. In this study, I aimed to identify *Hox9*, *Hox10* and *Hox11* homologous sequences in these species. These sequences would then be used to design new primers to use in RACE experiments using cDNA produced from *Ichthyomyzon* species embryos to identify full coding sequence reads of the *Hox* genes.

Degenerate primers designed to amplify *Hox9*, *Hox10* and *Hox11* homologs in these *Ichthyomyzon* species were successful in amplifying *Hox* genes. The PCR specificity in some cases was not always accurate, as the primers sometimes amplified *Hox* genes from different paralogous groups. Due to the high sequence similarity of some *Hox* genes, accuracy could not be guaranteed with these primers. Despite these technical challenges, *Hox* homologs were successfully amplified, cloned and sequenced in both *I. castaneus* and *I. fossor*. In the case of *I. unicuspis*, faintly detectable PCR products were initially obtained, yet could not be cloned and sequenced. Further attempts to isolate *I. unicuspis* sequences was not attempted as a result of recent findings by Docker *et al.*

(2012) that showed nucleotide sequence similarities between *I. fossor* and *I. unicuspis* are not substantially different. The *Hox10* homolog primers were successful in amplifying putative *Hox10* genes in *I. castaneus* and *I. fossor*, although further cloning and sequencing would be required to determine whether these primers were specific to the *Hox10* paralogous group. *Hox9* and *Hox11* homolog primers were less specific than the *Hox10* primers in that the *Hox9* homolog primers were only partially successful and the *Hox11* homolog primers were unsuccessful in amplifying their desired paralogous group genes. These non-specific gene amplifications are not unexpected given that these families of *Hox* genes are highly similar in other vertebrates, especially over the homeobox region. Amino acid similarities between the *B. florida* (Florida lancet) and the *Mus musculus* (mouse) homeodomain were shown to have 55/60 (91.7%) identities and in the extreme C terminus end, 10/14 (71.4%) identities (Holland *et al.* 1992). Similarities such as those seen in *B. florida* and *M. musculus* can be seen between distantly related chordates (Carr *et al.* 1998; Irvine *et al.* 2002; Takio *et al.* 2004; Takio *et al.* 2007).

The exact gene identity of some of the *Hox* genes identified in the local species could not always be confirmed given the small amplified region within the homeobox gene, though in these cases, the paralogous groups to which they belonged were identified. Longer sequences were not obtained because there was insufficient information from orthologous lamprey sequences to design additional primers outside the homeobox regions. As longer *Hox* gene sequences would help identify the proper paralogous groups to which the *Hox* genes belonged, I considered different techniques. One method of obtaining flanking sequence to the identified gene regions would be to use

genome or primer walking methods using genomic DNA (Leoni *et al.* 2011). Genomic DNA is easily acquired, yet this method necessitates some knowledge of the target sequence in order to identify it among flanking non-coding DNA. It also requires multiple cloning and sequencing steps to be able to build contigs that would span introns. This could prove difficult if the introns were large. Genome walking is also unable to determine whether any sequence obtained is coding or non-coding sequence. Another method considered was rapid amplification of cDNA ends (RACE). This method requires RNA from the correct stage of development of the lamprey, but confines the search to a smaller set of short PCR products that are actively transcribed products (Frohman *et al.* 1988).

4.2 - Rapid Amplification of cDNA Ends

For the aforementioned reasons, I favoured the use of RACE to acquire additional, non-homeobox sequences within the posterior *Hox* genes. Lamprey embryos were needed to obtain RNA at stage-specific times. No local lamprey embryos were produced, despite the concerted effort to find local spawning phase lamprey. I therefore opted to use *P. marinus* embryos from Hammond Bay Biological Station. Due to difficulties in shipping RNA to Manitoba (see Section 3.2), efforts to use RACE were limited.

RT-PCR analysis of the RNA derived from 3- and 6-day old *P. marinus* embryos suggested that there was no expression of *Hox9* and *Hox10*. Expression of the putative *Hox9* and *Hox10* genes in the *P. marinus* embryos was detected from embryos 9 days post-fertilization and the stream prolarvae (approximately 15 days post hatch) samples. This suggests that expression of these *Hox* genes begins sometime between 6 and 9 days

post fertilization, which equates to the developmental time points between the formation of the neural plate and protrusion of the head (Piavis 1961; Tahara 1988). These findings are in agreement with a study by Takio *et al.* (2007), who demonstrated expression of *Hox9*, *Hox10* and *Hox11* genes in *Le. camtschaticum* at the time of hatching, which takes place approximately 9-11 days post-fertilization. In all vertebrates, *Hox* gene expression begins during early gastrulation (Wolpert *et al.* 2007). As development continues there is a temporal activation of subsequent *Hox* genes (Durston *et al.* 2011). Posteriorly expressed genes such as *HoxD11* have been shown in mice to activate during late gastrulation (Gérard *et al*. 1993), yet expression of posteriorly expressed genes such as mouse *Hoxa10*, *Hoxc10* and *Hoxd10* genes have been shown to continue expression on in to mid-to-late embryogenesis during neurulation (Choe *et al.* 2006). This suggests that lampreys follow the same basic developmental expression of gnathostomes.

Following this confirmation that at least putative homologs of *Hox9* and *Hox10* were expressed in day 9 embryos, RACE degenerate primers were then designed to amplify *Hox9*, *Hox10* and *Hox11* targets in *P. marinus* (Table 2.6). Degenerate primers were designed to capture all potential *Hox* candidates within this group of posteriorly expressed genes, as a means of identifying as many *Hox* genes as possible using a limited resource of RNA.

A 300bp RACE product was isolated and subsequently identified as the *P. marinus Col2a1a* collagen gene. The *Hox*-specific primer used in the RACE experiment was only 52.4% identical to the *Col2a1a* gene, but nevertheless, it was evidently sufficient to anneal to the target cDNA and promote PCR amplification, using the lowered annealing temperature and increased $MgCl_2$ concentration that was used to help

promote amplification using the degenerate primers. Also, during lamprey development, collagen is expressed at high levels (Dale and Topczewski 2011), increasing the chances that the gene would be amplified. Small nucleotide polymorphisms within the primer binding site may have also contributed to the successful binding of the primer the *Col2a1a* gene; however, this is speculation and further experimentation would need to be conducted to determine whether or not this is true. RNA samples were entirely consumed in the multiple reiterations and optimizations of the RACE experiment, with no *Hox* sequences identified. Amplification of remaining RNA may have been useful to contiune with RACE; however, the absence of *Hox* products in the final RACE experiments led to the conclusion that any remaining *Hox* RNA had already been degraded. Thus, the absence of lamprey embryo RNA and results in the RACE experiments caused a shift in the focus of the project priorities. However, given the discovery of *Hox* gene expression in 2-4 year old *I. fossor* gonadal tissue in the final stages of this study (see Section 2.3.2), an alternative approach to identifying full length *Hox* gene coding sequences could use RNA from larval stage lamprey. This could be pursued in the future.

4.3 - *Petromyzon marinus* Genome Search and Phylogenetic Trees

An alternative approach to identifying lamprey *Hox* genes involved studying the *P. marinus* sequenced genome for the presence of putative new *Hox* sequence or to extend the knowledge of *Hox* genes that have previously been described (previously identified lamprey *Hox* genes Table 2.7). Another approach to this new focus involved the design and use of additional *Hox* primers to amplify and study *Hox* genes in lamprey species other than *P. marinus*.

4.3.1 - Hox Sequence Identification

Most *Hox* genes contain only two exons (Liang *et al* 2011). In the case of the anterior *Hox* genes (1-8), a conserved region known as the YPWM motif, contained within the hexapeptide region, can be observed at the 3′ end of exon1 (Shanmugan *et al*. 1997; LaRonde-LeBlanc and Wolberger 2003). A variable degree of similarity in amino acid sequence can be seen in the 5′ region of exon1 between orthologous genes (Holland *et al.* 1992). Using this similarity often proves difficult, as nucleotide variation is more divergent. Another identifying feature of exon1 is the length of the sequence, which is often between 250bp and 650bp (Liang *et al*. 2011). Exon2, in contrast to exon1, has a highly similar region across all paralogous groups. This region and the main identifying feature of exon2 is the homeodomain (Hueber *et al*. 2010). This 180bp region, located towards the 5′ end of exon2, encodes a DNA binding domain, which is essential to the function of the Hox proteins and is therefore highly conserved (Murtha *et al.* 1991). Sequence similarity among the paralogous groups is greatly diminished outside of the homeobox region. Coding sequence 5′ and 3′ of the homeobox region is often highly variable between paralogous groups (Liang *et al*. 2011). The 3′ region can range from a couple of base pairs to several hundred base pairs before the coding region is terminated by a stop codon (Ravi *et al.* 2009; Amemiya *et al.* 2010).

Identifying the paralogous group for any particular *Hox* gene in lampreys can sometimes prove to be difficult. This is because the easiest portion of the *Hox* gene to identify, the homeobox, is the most highly conserved region within the gene. A number of analytical approaches have been used to place an unknown *Hox* gene within its appropriate paralogous group. Primarily, there is sequence similarity. This method is

often insufficient to identify to which group the *Hox* gene belongs but can usually place the genes within a smaller subset of paralogous genes (Koh *et al.* 2003; Hueber *et al* 2010). Secondarily, highly conserved amino acid residues present in exon2 can sometimes help narrow the identification of the *Hox* genes into their proper paralogous groups. The studies by Takio *et al*. (2004) and Takio *et al*. (2007) have compiled a list of these amino acid residues for *Hox* paralogous groups 1-11; however, the conserved residues do not always help to distinguish between the different paralogous groups, as they are not always conserved through every individual *Hox* gene. In the case of Contig75333, it was not obvious to which paralogous group the putative *Hox* sequences belong, because it showed similarity to both *Hox9* and *Hox10* nucleotide and amino acid sequences. With nucleotide alignment and amino acid conserved residue matching not being enough to identify the correct paralogous groups, a more strenuous test was needed. The list of putative *Hox* genes from the *P. marinus* genome annotations was therefore compiled and used in a phylogenetic analysis. This process takes orthologous genes from other reference species and, through construction of multiple phylogenetic trees, infers the most likely relationships through nucleotide or amino acid substitutions. The clades within each tree that are inferred more often (i.e., more consistently) receive a higher bootstrap value indicating the higher degree of confidence in those clades (Harrison and Langdale 2006).

A total of 47 contigs were discovered to contain predicted coding sequence for *Hox* genes within the two *P. marinus* contig databases, with a total of 30 unique sequences consisting of either exon1, exon2 or both exon1 and exon2 sequences. Previous studies suggesting the number of *Hox* genes in lamprey have varied. For *P.*

marinus, looking at genomic DNA, Pendleton *et al*. (1993) found 19 *Hox* genes within the *Hox1-Hox10* paralogous groups, Irvine *et al.* (2002) found 21 genes within the *Hox1-Hox11* paralogous groups, and Force *et al*. (2002) found 19 *Hox* genes when looking at all *Hox* genes. In *Lethenteron camtschaticum*, studies have identified and observed the expression of 19 *Hox* genes (Takio *et al*. 2004; Takio *et al*. 2007; Karaku *et al*. 2008). Finally, in *La. planeri*, a total of 18 *Hox* genes were estimated through PCR-amplified products using genomic DNA and identifying different *Hox* genes using 82bp of sequenced nucleotides (Sharman and Holland 1998). The short lengths of these amplified products made it difficult in some instances to distinguish the proper paralogous group to which the *Hox* gene belonged, yet the novel sequences were helpful in identifying a subset of *Hox* genes. These previous studies identified far fewer *Hox* genes than the current genome search suggests. Sharman and Holland (1998) suggest that "Lamprey are deduced to have approximately 21 (or very few more) *Hox* genes from PG1-PG10." This statement is in agreement with my findings, as 20 putative *Hox* genes appear to fall within PG1-PG10 and eight putative genes appear to fall within PG11-PG14. What is not known, however, is whether the identified putative *Hox* genes are functional genes or are pseudogenes. These pseudogenes would consist of either a non-fuctional promoter or contain nucleotide sequences with insertions or deletions that cause a disruption in the reading frame leading to a premature stop codon. As many as eight *Hox* pseudogenes have been shown in *Salmo salar* (Atlantic salmon) whose genome contains 13 *Hox* clusters (Mungpakdee *et al.* 2008). Transcribed *Hox* pseudogenes, like that of psi *hoxa2b* found in Japanese medaka (*Oryzias latipes)* can further complicate detection of these genes (Davis *et al.* 2008). *Hox* pseudogenes have been found within many *Hox*

containing organisms such as zebrafish, mouse and amphioxus (Meyer 1998). The presence of pseudogenes within the *Hox* clusters further complicates the identification of *Hox* genes when identifying these genes within genomic DNA. When this study commenced, only a partially annotated genome was available for the sea lamprey. However, by the time this thesis was being written, Smith *et al*. (2013) published a much more thorough genome analysis, including transcriptomic analyses of different life stages and tissues of the sea lamprey, although there is still only a partially annotated genome. In this analysis, they found 23 homeodomain-containing regions of exon2. Unfortunately, these genes have yet to be ascribed to developmental stages as the transcriptomic data has yet to be fully assembled and annotated. Once complete, these data will be able to provide much more insight into the expression of *Hox* genes within *P. marinus*.

4.3.2 - Hox Phylogenetic Trees

The phylogenetic analysis of *Hox* genes was performed using amino acid sequences derived from the database from the first amino acid of the homeobox to the stop codon at the end of the gene. These sequences were used for the phylogenetic analysis because of the relative ease and certainty of their identification, and they provided the longest available sequences that have been used in a phylogenetic analysis for lampreys prior to the writing of this thesis. Identification of coding sequence upstream of the homeodomain was difficult for two reasons. One is that these upstream sequences are highly variable, and the second is that the position of the intron is also variable within paralogous groups.

Maximum likelihood and maximum parsimony phylogenetic analyses were therefore based on the homeobox to stop codon amino acid sequences of the identified

Hox genes. Resolution of all the *Hox* paralogous groups except for *Hox4-Hox7* were confirmed within the ML and MP phylogenetic trees. Separation of the *Hox* clusters was observed in *Hox* clades *Hox1*, *Hox3*, *Hox9*, *Hox11* and *Hox14*, but not *Hox4-Hox8*, *Hox10* and *Hox13*. Longer sequences could offer a more accurate identification of these genes, provided that that the rate of divergence of these genes reflected phylogenetic differences. Previous studies have shown that *Hox* genes can prove to be reliable candidates for assessing phylogenetic differences. Henkel *et al.* (2012) used phylogenetic assessment to show the relationship between *Anguilla anguilla* (European eel) and other fish species *Hox9* paralogous. Phylogenetic analysis of groups or clusters of *Hox* genes also helps to provide greater resolution of interspecies relationships (Hoegg *et al.* 2007*;* Henkel *et al.* 2012). As the *P. marinus* genome annotation becomes more complete, *Hox* gene clusters may help to resolve evolutionary relationships that still remain ambiguous.

For paralogous groups *Hox4-Hox7*, low bootstrap values were observed due to their similarity to each other. Their placement within could therefore not be guaranteed. Previous studies also report various difficulties in identifying and properly separating these paralogous groups (Pendleton *et al*. 1993; Sharman and Holland 1998; Irvine *et al*. 2002; Takio *et al*. 2004). In order to improve the resolution of the paralogous groups *Hox4-Hox7* phylogenetic trees, longer amino acid sequences or nucleotide sequences would have to be studied. Although a confident assessment of the *Hox4-Hox7* clade could not be guaranteed, separation of the individual paralogous group clades is observed. Curiously, grouping of two different paralogous groups can be seen in the case of Contig8613 and Contig22678, previously identified as *HoxF5* and *HoxK6* respectively by Irvine *et al.* (2002). These two contigs group together into the tentative *Hox6* paralogous

group clade and show a low bootstrap values (53 ML, 69 MP). This suggests that either the genes are improperly identified or that the two different paralogous group genes are more closely related to each other than they are with the other *Hox* genes. Regardless, this shows that there is a large divergence between the lamprey and gnathostomes species used in the phylogenetic analysis.

Within this studies phylogenetic trees that compared lamprey genes with those of other vertebrates, the lamprey genes typically appeared as an outgroup to the gnathostome species. Most of the lamprey *Hox* genes were seen as outgroups to either paralogous group clades or even an outgroup to the *Hox* clusters. Only in one instance was a lamprey gene found within a gnathostome *Hox* clade. In the MP *Hox4-Hox7* phylogenetic tree Contig72950 was found within a *C. milli Hox5* clade, but the relationship was only weakly supported by a bootstrap value of 21. The same contig was placed outside of the *Hox5* clade in the ML phylogenetic, suggesting that Contig72950's placement within the phylogenetic trees is still ambiguous. The overwhelming placement of the *Hox* genes as outgroups to the other vertebrates' *Hox* genes helps to define the divergence between lampreys from other vertebrates. Showing even greater divergence from the vertebrate *Hox* genes was the marine invertebrate *B. lanceolatum*. The consistent placement of its *Hox* genes as an outgroup to the vertebrate *Hox* genes within all the phylogenetic trees suggests that the invertebrates are basal still to that of lamprey. These findings are consistent with current evolutionary theories that argue that vertebrates evolved from invertebrate ancestors and that gnathostomes evolved from agnathan ancestors (Heimberg *et al.* 2010).

The origin of the *Hox* genes in lampreys and their relationship to other species continues to be debated. There is still no consensus as to whether lampreys have three or four cluster groups of *Hox* genes. Some studies agree that lampreys have at least three clusters (Force *et al*. 2002), while other studies, such as Irvine *et al*. (2002), believe that in at least *P. marinus*, they contain four *Hox* clusters. When examining the number of *Hox* clusters, it is worth considering genome duplication events. Although there is a lack of consensus as to when the duplication events occurred, most studies agree that at least one genome duplication event happened before the divergence of agnathans and gnathostomes. Stadler *et al*. (2004) suggests a doubling event occurred before the agnathan/gnathostome divergence, followed by a loss of almost all of one set of paralogous genes, and then two separate duplication events after the divergence. Force *et al.* (2002) and Fried *et al*. (2003) suggest a duplication event before and after the agnathan/gnathostome divergence. Kuraku *et al*. (2009) suggest that two duplications occurred before the agnathan/gnathostome divergence. Finally, with the assembly of the lamprey genome, Smith *et al. (*2013) suggest that two whole-genome duplication likely occurred before the divergence of ancestral lamprey and gnathostome lineages. These inconsistencies can lead to confusion in accurately assessing the number of *Hox* clusters in lampreys. Force *et al.* (2002) identified four possible putative *Hox* genes for *Hox1*, *Hox4* and *Hox9*, by combining data from multiple studies. This suggests the presence of four paralogous groups in lampreys. The current analysis of the *P. marinus* genome in this thesis has uncovered four putative *Hox13* genes and is the first instance of identifying four genes within the same paralogous group from one source. The presence of four *Hox13* paralogous genes suggests two possible gene histories: one possible scenario

could involve two *Hox* cluster duplications, followed by subsequent loss of one full complement of all other *Hox* paralogous sets other than *Hox13* genes; while the other scenario involves a full genome duplication, followed by a partial duplication to create three *Hox* clusters, and that the fourth *Hox13* paralogous gene was the result of a single gene duplication. Curiously, the genome search was unable to identify four separate putative gene targets for the *Hox1, Hox4* and *Hox9* paralogous groups. Discovery of a fourth paralogous gene for any of these or other paralogous groups would help to explain the possible origins of the multiple *P. marinus Hox* clusters. Unfortunately, the *P. marinus* genome database is still not fully annotated and further insights into *Hox* cluster numbers could not be made, yet as further annotations of the *P. marinus* genome become available more information becomes available to help elucidate the nature of lamprey whole-genome duplications as well as more information on the nature of lamprey *Hox* clusters. Furthermore, functional studies of the putative *Hox13* genes discovered in the *Hox* genome search would be able to determine whether *P. marinus* contain four functional *Hox13* genes.

Of the putatively identified *Hox13* genes, one was found to have three exons instead of the normal two exons seen in other *Hox13* genes. The Contig25758 matched one of the two *Hox13* genes, *LjHox13-beta*, discovered in *Le. camtschaticum* (Kuraku *et al*. 2008). The additional intron within the *Hox13* gene present on Contig25758 was also found in *LjHox13-beta* from *Le. camtschaticum,* showing gene orthology. Phylogenetic data would also suggest that the presence of this extra intron in *Hox13* would at least be present in all northern hemisphere lampreys based on the last common ancestor of these two species (Lang *et al.* 2009). The additional intron site within these *Hox13* genes is also

found in the same location within the homeobox region in the putative *Hox14* genes. Kuraku *et al*. (2009) made the same observations, stating the secondary intron was not a sole *Hox14* hallmark. They suggest, based on phylogenetic analysis, that the *LjHox14-alpha* gene was not an ortholog to any *Hox* gene found in amphioxus, but was a tandem gene duplication of the *Hox13* genes (Kuraku *et al*. 2009). Further, this example of a unique *Hox13* gene found in distantly related species may support the idea that four *Hox13* genes are not representative of four paralogous groups, but rather there was a single gene duplication event.

No putative *Hox* genes were discovered in the *P. marinus* genome search for the *Hox12* paralogous group. In fact, no studies thus far have been able to identify any putative *Hox12* genes in any lamprey species (Pendleton *et al.* 1993; Sharman and Holland 1998; Force *et al*. 2002; Irvine *et al*. 2002; Fried *et al*. 2003; Takio *et al*. 2004; Takio *et al*. 2007; Kuraku *et al*. 2008; Smith *et al.* 2013). It is curious that lamprey would completely lack the *Hox12* paralogous group from their genomes given that it has been shown in other organisms to have an important role in neuronal and tail development (Ikuta *et al*. 2010). The study Smith *et al*. (2013) has very recently identified at least two clusters of *Hox* genes in *P. marinus*, as well as additional *Hox* genes that were not placed within a known cluster. Ambiguity of *Hox* gene arrangement within lampreys makes it difficult to identify whether or not any *Hox12* genes exist, yet all data to date on lamprey *Hox* gene clusters seem to suggest the absence of any *Hox12* genes. Once more complete annotations of the *P. marinus* genome become available, studies analyzing the regions between *Hox11* and *Hox13* PGs could perhaps shed more light on the absence of *Hox12* genes in lampreys, assuming all of the *Hox* clusters are collinearly arranged, as they are

in other species (Hueber *et al.* 2010; Mallo *et al.* 2010). More complete assemblies of RNA sequencing may also help to provide resolution of the absence of *Hox12* genes in *P. marinus*.

Three *I. fossor Hox* genes were also used in this studies phylogenetic analysis of lamprey *Hox* genes. These genes were identified from gonadal tissue from an individual female ammocoete approximately 97mm in length. The three sequences were placed into the *Hox9* and *Hox10* PGs, two as *Hox9* sequences and one as *Hox10* sequence. This ammocoete lamprey was able to maintain the RNA sequences to these *Hox9* and *Hox10* PG genes well after embryonic development was complete. The question stands as to whether or not these specific genes have any function in postembryonic development, whether the RNA was kept intact until the time of extraction or if there was genomic DNA contamination. Genomic DNA contamination is a concern when dealing with any RNA samples; however, many effective methods exist for removing genomic DNA (Añez-Lingerfelt *et al*. 2009). Analysis of known exon-intron boundaries post RNA sequencing also provides an effective way of determining if the samples contained contaminating DNA. No observable DNA contamination was observed when looking at *Hox* gene sequences within the *I. fossor* RNA seq contigs. Takio *et al*. (2007) has noted that unlike other vertebrates, there appeared to be no temporal collinear correlation of the *Le. camtschaticum Hox* genes with the PG to which the genes were assigned. Lacking a more sophisticated regulation of these developmental genes could have also led to the presence and detection of these genes post-hatch. Alternatively, *Hox* gene expression could continue past embryonic development. Expression of *Hox* genes have been detected in adult humans and mice, suggesting that some *Hox* genes play more than just a

role in embryonic development (Chen and Capecchi 1999; Golpon *et al*. 2001). While the focus of the functions of *Hox* genes has concentrated on embryonic pattern development, there is a growing body of evidence for expression of these genes in later stages of development. Sifuentes-Romero *et al*. (2010) have shown that *HoxD11* and *HoxA13* are expressed in the developing reproductive tract in the sea turtle *Lepidochelys olivacea* at female-inducing temperatures. Curiously, three *Hox* genes were found within the *I. fossor* RNA sequencing database, from a developing female gonad (see Section 2.3.2). Further functional studies could identify the nature of these genes and determine if there is a post-hatch function or if it is just a relic of embryonic development. If in fact these *Hox* genes do have a role in post-embryonic development, further stage specific lamprey RNA sequencing databases could lead to the identification of more potential *Hox* targets and uncover new directions of research.

To date, this study has been the most systematic search for *Hox* within lamprey species. Recently, however, during the writing of this thesis, new genome assembly information has become available for *P. marinus* (Smith *et al.* 2013). A total of 23 unique *Hox* genes were identified of which 23 homeodomain containing regions on exon2 along with 12 hexapeptide regions on exon1 were described (Smith *et al.* 2013). Twenty-two of these identified *Hox* genes were also identified in the *P. marinus* genome search performed in this thesis. Two sequences, *Hox1* (Contig6181) and *Hox13* (Contig25758), out of the eight newly identified putative *Hox* gene exon2 sequences identified in this thesis were matched to sequences identified in Smith *et al.* (2013), while one sequence, *Hox7* (GL476758), out of the five newly identified putative *Hox* gene exon1 sequences identified in this thesis were matched to sequences identified in Smith *et al.* (2013). The

fact that more putative *Hox* sequences were identified within this thesis project suggests a disparity in identification methods. Smith *et al.* (2013) used bacterial artificial chromosomes combined with *Hox2*, *Hox4* and *Hox9* probe hybridization and subsequent sequencing and alignment analysis. Perhaps some of the genes identified within this thesis are non-coding pseudogenes in which the absence of detection by Smith *et al.* (2013) from the BAC probing suggests a false positive confirmation. No indications of premature stop codons were found within the ambiguous genes that would indicate the presence of a pseudogene. Further testing with cDNA as well as functional gene testing would be required to resolve this matter further. Search algorithms used to identify *Hox* genes may have been too specific to detect the remaining genes and failed to identify all of the *Hox* genes that do not conform to other known *Hox* genes either in sequence or structure.

4.4 - *Petromyzon marinus Hox* Gene Verification

The *P. marinus* genome search for *Hox* genes revealed eight exon1 sequences with no identifiable exon2 sequences. In order to help further resolve if these exon1 sequences were in fact putative *Hox* sequences, efforts to identify their corresponding secondary exons were made by using low stringency PCR. This would help to determine whether or not the *P. marinus* genome databases lack full genetic coverage, and if so, help to identify further *Hox* sequences. Multiple PCR amplicons due to low stringency PCR conditions as well as unknown intron sizes made it difficult to identify positive DNA fragments. Out of the four genes under examination, no successful PCR products were isolated for *Hox7*, *Hox8*, and *Hox13* samples. Multiple PCR fragments within these samples were determined to be non-specific primer amplicons while any potential

positive amplicons lacked sufficient quantity of product to be cloned and sequenced, and could not be further amplified to produce sufficient product. Lack of success in isolating positive DNA fragments for the *Hox7*, *Hox8*, and *Hox13* targets suggests that either the primers were not specific enough to their target sequences or the exon1 sequences identified are pseudogenes. Unfortunately, time constraints did not allow for the testing of more primer sets. Curiously, none of these sequences could be identified with any corresponding exon2 sequences discovered in the genome search. It is possible that some of the four exon1 sequences investigated here are part of exon2 gene sequences also found in this study. This is most likely the case especially for the *Hox13* sequence where there are already four exon2 sequences that have been discovered. The relatively short length of some of these contigs and/or the location of exon2 putative gene sequences to the periphery of the contigs easily explains why the putative exon1 and exon2 sequences could not be found together on the same contigs. Further assemblies of the *P. marinus* genomes will allow for longer sequence reads which will likely help to identify if these putative identified exon1 sequences have a corresponding exon2 sequences, whether the identified exon1 and exon2 sequences belong to the same gene and finally help to determine the size of the intron.

Positive sequence was identified for the *Hox2* sample, in which the intron was estimated to be approximately 1600bp in length. The exon2 sequence identified was shown to have high sequence similarity to the previously identified *Le. camtschaticum Hox2* sequence. This presence of the second exon and further match to a known lamprey *Hox* gene suggests that this putatively identified sequence is likely a functional gene. The identification of the exon2 sequence through PCR raises the question as to why it was not

discovered in the *P. marinus* genome databases. The fact that this sequence is still missing from the databases suggests an incomplete coverage of the entire genome. This further suggests that the current search for *Hox* genes may not have uncovered the full *Hox* complement for *P. marinus*.

4.5 - Possible Use of *Hox* Genes for Resolving Lamprey Phylogenies

4.5.1 - *Hox10* Sequences in Different Lamprey Species

Lethenteron camtschaticum LjHox10s gene expression was shown to migrate posteriorly with the developing tail bud with respect to the myotomal numbers as the body axis extends posteriorly (Takio *et al*. 2007). This study aimed to identify possible roles that *LjHox10s* or an orthologous lamprey sequence may play a role in varying the number of myomeres within different lamprey species. To look at this, orthologs were cloned and sequenced from nine lamprey species. These species consisted of high and low myomere count individuals, as well as members of a pair that have different myomere numbers (*La. richardsoni* and *La. pacifica*) and members of a pair that have no difference in myomere numbers (*I. unicuspis* and *I. fossor*). No sequence similarities or nucleotide or amino acid synapomorphies could be associated with high or low myomere counts in these lampreys. This suggests that the 226bp region around the homeobox region of these genes had no bearing on the number of myomeres that develop within an individual. If the *LjHox10s* gene and other lamprey orthologs do play a role in myomere determination, the answer likely lies within the regulatory region of these genes. The ability to up- or down-regulate these genes would allow for varying lengths and possibly myotomal numbers within the individuals.

Considering the broad coverage of the orthologous genes analyzed above, a phylogenetic analysis was performed. Comparisons to morphological data provided by Gill *et al*. (2003) and genetic data looking at mitochondrial *cytB* provided by Lang *et al*. (2009) revealed some differences, some small and others larger. Most notable was that *En. tridentatus* was placed most basal in the southern hemisphere lamprey clade. Existing morphological and genetic data place *En. tridentatus* in a clade with *Eudontomyzon*, *Lampetra* and *Lethenteron* species (Gill *et al.* 2003; Lang *et al.* 2009). The grouping of the *Ichthyomyzon* species separate from other lamprey genera on the other hand did agree with previous studies. No other direct comparisons could be made when viewing individual species. Looking at genera placements within the trees, *Eudontomyzon* was placed into a clade with *Lethenteron* as opposed to being placed in a clade with *Lampetra* according to the previous morphological and genetic phylogenetic data (Gill *et al.* 2003; Lang *et al*. 2009). The bootstrap values for the *Ichthyomyzon* species clades in both ML and MP phylogenetic trees were the only values greater than 90, showing that these sequences were the only sequences that could be assigned with a high degree of confidence. In the remaining species clades the bootstrap values were less than 90, showing that the data was insufficient to properly analyze the gene relationships. With longer sequence reads and establishing a more complete set of *Hox* genes could help to better understand the evolutionary relationships among lamprey species. Relationships like the still ambiguous relationship between the three families of lampreys as well as discrepancies between European and North American *Lampetra* species in addition to inconsistancies between *Eudontomyzon* species and *Lethenteron* species may be further resolved with new analysis techniques (Lang *et al.* 2009).

4.5.2 - General *Hox* Gene Screening in Lampreys

Several studies have previously investigated *Hox* gene complements in lamprey species. More specifically, three lamprey species have been studied to a greater extent than others. *Petromyzon marinus*, the sea lamprey, is probably the best studied lamprey species with six studies contributing to the understanding of their *Hox* clusters (Pendleton *et al*. 1993; Carr *et al*. 1998; Amores *et al.* 1998; Irvine *et al*. 2002; Force *et al.* 2002; Smith *et al.* 2013). Next is *Le camtschaticum*, which has had 18 of its genes identified and is the only species studied using gene expression experiments (Takio *et al*. 2004; Takio *et al.* 2007; Kuraku *et al*. 2008). Finally, Sharman and Holland (1998) identified 18 *Hox* gene fragments in *La. planeri*. Following the work of Pendleton *et al*. (1993) and Force *et al.* (2002), general *Hox* primers were used to analyze *Hox* genes in other lamprey species. The aim of the experiment was to determine the efficiency of amplification *Hox* genes in alternative lamprey species. Thirty *Hox* genes were cloned and sequenced from nine different lamprey species. A minimum of two genes were sequenced for each of the species tested. Duplicate gene repeats were noticed immediately, within two to three gene sequences. Putative *Hox* genes were identified from each of the paralogous groups *Hox1-Hox14* with exception to *Hox9*, *Hox12* and *Hox13*, across the 30 sequences analyzed. The 30 sequenced samples were likely insufficient to determine if a full coverage of all representative lamprey genes. Never the less, sequence results showed a wide range of putative *Hox* targets across the lamprey species analyzed. This method proved effective in identifying novel putative *Hox* gene targets in lamprey species with no previously identified *Hox* genes.

4.6 – Conclusions and Future Directions

At the time this study started, *Hox* genes in lampreys were not well described. Since lampreys represent a key evolutionarily transitional vertebrate group, it is important to gain a more thorough understanding of both the composition and organization of this important group of developmental genes. This thesis was the most systematic search of *Hox* genes conducted to date using the *P. marinus* genome, an *I. fossor* transcriptome, and a targeted PCR approach across multiple lamprey species. This research helps to improve the understanding of *Hox* genes in vertebrates by further elucidating the number of *Hox* genes in this ancestral vertebrate group and potential roles that *Hox* genes may have on myomere development and number.

Investigation of the two *P. marinus* genome assemblies discovered 15 extended sequences of previously identified *Hox* genes as well as 12 new putative *Hox* gene sequences. These genes were categorized and assessed based on sequence similarities to, and phylogenetic analysis with, full complements of previously identified *Hox* genes from *B. lanceolatum, C. milii*, *L. menadoensis* and *M. musculus*. This analysis showed candidates from all paralogous groups except for the still absent *Hox12* paralogous group. Considering the exhaustive search conducted for all *Hox* genes within the two *P. marinus* assemblies, these findings along with all other previous work with lamprey *Hox* genes would suggest that *P. marinus* lacks the *Hox12* PG. Of course, the absence of the *Hox12* PG could not be definitively demonstrated; future investigation into the genomic DNA regions between *Hox11* and *Hox13* or full RNA sequencing databases will be able to help resolve this question fully. Also uncovered in this thesis, for the first time in a single study, were four putative *Hox13* genes, which supports the hypothesis that there are four

PGs in lampreys (see Section 4.3.2). Further work, however, is needed to confirm this (e.g., to identify first, if these genes are true functional *Hox* genes and second, if these four genes truly belong to four distinct *Hox* clusters).

The thesis also observed expression of three posteriorly expressed *Hox* genes within the RNA sequencing transcriptome produced from the ovary of a 97mm *I. fossor* female (approximately 2-4 years old). The expression of these three genes was well outside the timing of embryogenesis and as such, it was curious to see expression of these genes. *Hox* gene expression has been previously seen beyond embryogenesis in mice and human, but this was the first instance in lamprey. How many *Hox* genes are expressed post-embryogenesis, where and to what extent they are expressed, and what their function is has yet to be investigated and could provide new areas of research for the future.

This thesis also investigated the myomere differences across nine species in five lamprey genera and the potential role the *Hox10* genes homeobox region has on their number. Although no discernible differences in the homeobox region could be identified in the species assayed, the sequences identified may help to uncover differences in other coding regions of these genes or in regulatory elements. Furthermore, phylogenetic analysis of this gene region was one of the first to use nuclear (rather than mitochondrial) gene data to assess phylogenetic relationships in lampreys. Finally, this thesis identified one to four putative *Hox* homeobox gene fragments from nine lamprey species that were previously unidentified, to help better understand the capabilities of identifying *Hox* genes in lamprey species with no previous *Hox* genes identifications.

This study focused on lamprey *Hox* genes, which are important genes in development and generally not yet well characterized. A better understanding of *Hox*

genes in lampreys will permit a better understanding of lamprey development and, as one of only two groups of extant agnathans, the information will help to improve understanding of vertebrate evolution.

5. References

Amemiya, C. T., Powers, T. P., Prohaska, S. J., Grimwood, J., Schmutz, J., Dickson, M., Miyake, T., Schoenborn, M. A., Myers, R. M., Ruddle, F. H., et al. (2010). Complete HOX cluster characterization of the coelacanth provides further evidence for slow evolution of its genome. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 3622–7.

Amores, A., Force, A., Yan, Y., Joly, L., Amemiya, C. T., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y., Westerfield, M., Ekker, M. and Postlehwait, J. H. (1998). Zebrafish *Hox* clusters and vertebrate genome evolution. *Science* **282**, 1711–14.

Añez-Lingerfelt, M., Fox, G. E. and Willson, R. C. (2009). Reduction of DNA contamination in RNA samples for reverse transcription-polymerase chain reaction using selective precipitation by compaction agents. *Analytical Biochemistry* **384**, 79–85.

Barber, B. A. and Rastegar, M. (2010). Epigenetic control of *Hox* genes during neurogenesis, development, and disease. *Annals of anatomy* **192**, 261–74.

Barrow, J.R. and Capecchi M.R. (1996). Targeted disruption of the *Hoxb-2* locus in mice interferes with expression of *Hoxb-1* and *Hoxb-4*. *Development* **122**, 3817–28.

Boguski, D. A., Reid, S. B., Goodman, D. H. and Docker, M. F. (2012). Genetic diversity, endemism and phylogenetic of lampreys within the genus *Lampetra sensu stricto* (Petromyzontiformes: Petromyzontidae) in western North America. *Journal of Fish Biology* **81**, 1891–14.

Bosley, T.R. and Conner J. V. (1984). Geographic and temporal variation in numbers of myomeres in fish larvae from the lower Mississippi River. *Transactions of the American Fisheries Society*. **113**, 238-42.

Carlander, K.D. (1969). Handbook of freshwater fishery biology. Volume 1. Life history data on freshwater fishes of the United States and Canada, exclusive of the Perciformes. *The Iowa State University Press, Ames*. v-vi, 3-752.

Carr, J. L., Shashikant, C. S., Bailey, W. J. and Ruddle, F. H. (1998). Molecular evolution of *Hox* gene regulation: cloning and transgenic analysis of the lamprey *HoxQ8* gene. *The Journal of Experimental Zoology* **280**, 73–85.

Cerny, R., Cattell, M., Sauka-Spengler, T., Bronner-Fraser, M., Yu, F. and Medeiros, D.M. (2010). Evidence for the prepattern/cooption model of vertebrate jaw evolution. *Proceedings of the National Academy of Sciences of the United States of America* **107,** 17262–7.

Chen, F. and Capecchi, M. R. (1999). Paralogous mouse *Hox* genes, *Hoxa9*, *Hoxb9*, and *Hoxd9*, function together to control development of the mammary gland in response to pregnancy. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 541–6.

Chen, J.-Y. (2011). The origins and key innovations of vertebrates and arthropods. *Palaeoworld* **20**, 257–278.

Choe, A., Phun, H. Q., Tieu, D. D., Hu, Y. H. and Carpenter, E. M. (2006). Expression patterns of *Hox10* paralogous genes during lumbar spinal cord development. *Gene Expression Patterns* **6**, 730–7.

Cochran, P. A., Lyons, J. and Gehl, M. R. (2003). Parasitic attachments by overwintering silver lampreys , *Ichthyomyzon unicuspis* , and chestnut lampreys , *Ichthyomyzon castaneus*. *Environmental Biology of Fishes* **68**, 65–71.

Cohn, M. J. (2002). Brief communications: visualization of coronary arteries in rats by 3-dimensional real-time contrast echocardiography. *Nature* **416**, 386–7.

COSEWIC (2010). COSEWIC assessment and status report on the Chestnut Lamprey *Ichthyomyzon castaneus* (Great Lakes - Upper St. Lawrence populations and Saskatchewan - Nelson River populations) in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. xii + 35 pp. (www.sararegistry.gc.ca/status/status_e.cfm).

Dale, R. M. and Topczewski, J. (2011). Identification of an evolutionarily conserved regulatory element of the zebrafish *col2a1a* gene. *Developmental Biology* **357**, 518–31.

Davis, A., Scemama, J. L. and Stellwag, E. J. (2008). Japanese medaka *Hox* paralog group 2: insights into the evolution of *Hox* PG2 gene composition and expression in the Osteichthyes. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution* **310B**, 623-41.

Docker, M. F., Youson, J. H., Beamish, R. J. and Devlin, R. H. (1999). Phylogeny of the lamprey genus *Lampetra* inferred from mitochondrial cytochrome *b* and ND3 gene sequences. *Canadian Journal of Fisheries and Aquatic Sciences* **56**, 2340–2349.

Docker, M. F. (2006). Bill Beamish 's Contributions to lamprey research and recent advances in the field. *Guelph Ichthyology Reviews* **7**, 1–52.

Docker, M. F. (2009). A review of the evolution of nonparasitism in lampreys and an update of the paired species concept. *American Fisheries Society Symposium* **72**, 71–114.

Docker, M. F., Mandrak, N. E. and Heath, D. D. (2012). Contemporary gene flow between "paired" silver (*Ichthyomyzon unicuspis*) and northern brook (*I. fossor*) lampreys: implications for conservation. *Conservation Genetics* **13**, 823–35.

Duboule, D. (ed.) (1994). Guidebook to homeobox genes. Oxford University Press, Oxford

Duboule, D. and Morata, G. (1994). Collinearity and function hierarchy among genes of the homeotic complexes. *Trends in Genetics* **10**, 358-364.

Duboule, D. (2007). The rise and fall of *Hox* gene clusters. *Development* **134**, 2549–60.

Durston, A. J., Jansen, H. J., In der Rieden, P. and Hooiveld, M. H. W. (2011). *Hox* collinearity - a new perspective. *The International Journal of Developmental Biology* **55**, 899–908.

Farmer, G., J. (1980). Biology and physiology of feeding in adult lampreys. *Canadian Journal of Fisheries and Aquatic Sciences* **37**, 1751-61

Ferrier, D. E. K. (2004). *Hox* genes: Did the vertebrate ancestor have a Hox14? *Current Biology* **14**, R210–1.

Force, A., Amores, A. and Postlethwait, J. H. (2002). *Hox* cluster organization in the jawless vertebrate *Petromyzon marinus*. *Zoology* **46**, 30–46.

Forey, P. and Janvier, P. (1993). Agnathans and the origin of jawed vertebrates. *Nature* **361**, 129–134.

Fried, C., Prohaska, S. J. and Stadler, P. F. (2003). Independent *Hox*-cluster duplications in lampreys. *Journal of Experimental Zoology* **25**, 18–25.

Frohman, M. A., Dush, M. K. and Martin, G. R. (1988). Rapid production of full-length cDNAs from rare transcripts: Amplification using single gene-specific oligonucleotide primer. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 8998–9002.

Gess, R. W., Coates, M. I. and Rubidge, B. S. (2006). A lamprey from the Devonian period of South Africa. *Nature* **443**, 981–4.

Gérard, M., Duboule, D. and Zákány, J. (1993). Structure and activity of regulatory elements involved in the activation of the *Hoxd-11* gene during late gastrulation. *The EMBO Journal* **12**, 3539–50.

Gill, H. S., Renaud, C. B., Chapleau, F., Mayden, R. L. and Potter, I. C. (2003). Phylogenetic of living parasitic lampreys (Petromyzontiformes) based on morphological data. *Copeia* **2003**, 687–703.

Golpon, H. A., Geraci, M. W., Moore, M. D., Miller, H. L., Miller, G. J., Tuder, R. M. and Voelkel, N. F. (2001). *Hox* genes in human lung: Altered expression in primary pulmonary hypertension and emphysema. *American Journal of Pathology* **158**, 955–66.

Goodman, D. H., Kinziger, A. P., Reid, S. B. and Docker, M. F. (2009). Morphological diagnosis of *Entosphenus* and *Lampetra* ammocoetes (Petromyzontidae) in Washington, Oregon, and California. *American Fisheries Society Symposium* **72**, 223–32.

Harrison, C. J. and Langdale, J. A. (2006). A step by step guide to phylogeny reconstruction. *The Plant Journal* **45**, 561–72.

Heimberg, A. M., Cowper-Sal-lari, R., Sémon, M., Donoghue, P. C. J. and Peterson, K. J. (2010). microRNAs reveal the interrelationships of hagfish, lampreys, and gnathostomes and the nature of the ancestral vertebrate. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 19379–83.

Hill, B. J. and Potter, I. C. (1970). Oxygen consumption in ammocoetes of the lamprey *Ichthyomyzon hubbsi* Raney. *Journal of Experimental Biology* **53**, 47–57.

Henkel, C. V, Burgerhout, E., de Wijze, D. L., Dirks, R. P., Minegishi, Y., Jansen, H. J., Spaink, H. P., Dufour, S., Weltzien, F., Tsukamoto, K., van den Thillart, G. E. E. J. M. (2012). Primitive duplicate *Hox* clusters in the European eel's genome. *PloS ONE* **7**, e32231, 1-9.

Hoegg, S., Boore, J. L., Kuehl, J. V and Meyer, A. (2007). Comparative phylogenomic analyses of teleost fish Hox gene clusters: lessons from the cichlid fish *Astatotilapia burtoni*. *BMC Genomics* **8**, 317.

Holland, P. W. and Garcia-Fernàndez, J. (1996). *Hox* genes and chordate evolution. *Developmental Biology* **173**, 382–95.

Holland, P. W., Holland, L. Z., Williams, N. a and Holland, N. D. (1992). An amphioxus homeobox gene: sequence conservation, spatial expression during development and insights into vertebrate evolution. *Development* **116**, 653–61.

Hubbs, C. L. and Trautman, M. B. (1937). A revision of the lamprey genus *Ichthyomyzon*. *Ann Arbor Univeristy of Michigan Press* **85**, 1–119.

Hubert, N., Hanner, R., Holm, E., Mandrak, N. E., Taylor, E., Burridge, M., Watkinson, D., Dumont, P., Curry, A., Bentzen, P., Zhang, J., April, J. and Bernatchez, L. (2008). Identifying Canadian freshwater fishes through DNA barcodes. *PLoS ONE* **3**, 1-8.

Hueber, S. D., Weiller, G. F., Djordjevic, M. a and Frickey, T. (2010). Improving *Hox* protein classification across the major model organisms. *PloS ONE* **5**, 1-12.

Ikuta, T., Satoh, N. and Saiga, H. (2010). Limited functions of *Hox* genes in the larval development of the ascidian *Ciona intestinalis*. *Development* **137**, 1505–13.

Irvine, S. Q., Carr, J. L., Bailey, W. J., Kawasaki, K., Shimizu, N., Amemiya, C. T. and Ruddle, F. H. (2002). Genomic analysis of *Hox* clusters in the sea lamprey *Petromyzon marinus*. *Journal of Experimental Zoology* **62**, 47–62.

Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences* **8,** 275-282.

Jordan, D. S. (1891). Relations of temperature to vertebrae among fishes. *Science* **18**, 104–7.

Koh, E. G. L., Lam, K., Christoffels, A., Erdmann, M. V, Brenner, S. and Venkatesh, B. (2003). *Hox* gene clusters in the Indonesian coelacanth, *Latimeria menadoensis*. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 1084–8.

Kuraku, S., Takio, Y., Tamura, K., Aono, H., Meyer, A. and Kuratani, S. (2008). Noncanonical role of *Hox14* revealed by its expression patterns in lamprey and shark. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 6679–83.

Kuraku, S., Meyer, A. and Kuratani, S. (2009). Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? *Molecular Biology and Evolution* **26**, 47–59.

Kuratani, S. (2004). Evolution of the vertebrate jaw: comparative embryology and molecular developmental biology reveal the factors behind evolutionary novelty. *Journal of Anatomy* **205**, 335–47.

Kuratani, S. (2005). Developmental studies of the lamprey and hierarchical evolutionary steps towards the acquisition of the jaw. *Journal of Anatomy* **207**, 489–99.

Lakiza, O., Miller, S., Bunce, A., Lee, E. M. J. and McCauley, D. W. (2011). *SoxE* gene duplication and development of the lamprey branchial skeleton: Insights into development and evolution of the neural crest. *Developmental Biology* **359**, 149–161.

Lang, N. J., Roe, K. J., Renaud, C. B., Gill, H. S., Potter, I. C., Freyhof, J., Naseka, A. M., Cochran, P., Pérez, H. E., Habit, E. M., Kuhajda, B. R., Neely, D. A., Reshetnikov, Y. S., Salnikov, V. B., Stoumboudi, M. Th. and Mayden, R. L.

(2009). Novel relationships among lampreys (Petromyzontiformes) revealed by a taxonomically comprehensive molecular data set. *American Fisheries Society Symposium* **72**, 41–55.

Langille, R. M. and Hall, B. K. (1988). Artificial fertilization, rearing, and timing of stages of embryonic development of the anadromous sea lamprey, *Petromyzon marinus* L. *Canadian Journal of Zoology* **66**, 549–554.

Leoni, C., Volpicella, M., De Leo, F., Gallerani, R. and Ceci, L. R. (2011). Genome walking in eukaryotes. *FEBS Journal* **278**, 3953–77.

Liang, D., Wu, R., Geng, J., Wang, C. and Zhang, P. (2011). A general scenario of *Hox* gene inventory variation among major sarcopterygian lineages. *BMC Evolutionary Biology* **11**, 25.

Madenjian, C. P., Chipman, B. D., Marsden, J. E. (2008). New estimates of lethality of sea lamprey (*Petromyzon marinus*) attacks on lake trout (*Salvelinus namaycush*): implications for fisheries management. *Canadian Journal of Fisheries and Aquatic Sciences* **65**, 535-542.

Mallatt, J. and Sullivan, J. (1998). 28S and 18S rDNA sequences support the monophyly of lampreys and hagfishes. *Molecular Biology and Evolution* **15**, 1706–18.

Mallo, M., Wellik, D. M. and Deschamps, J. (2010). *Hox* genes and regional patterning of the vertebrate body plan. *Developmental Biology* **344**, 7–15.

Manion, P. J. and Hanson, L. H. (1980). Spawning behavior and fecundity of lampreys from the upper three great lakes. *Canadian Journal of Fisheries and Aquatic Sciences* **37**, 1635–40.

Mateus, C. S., Alves, M. J., Quintella, B. R. and Almeida, P. R. (2013). Three new cryptic species of the lamprey genus *Lampetra* Bonnaterre , 1788 (Petromyzontiformes : Petromyzontidae) from the Iberian Peninsula. *Contributions to Zoology* **82**, 37–53.

McBride, R. S. and Horodysky, A. Z. (2004). Mechanisms maintaining sympatric distributions of two ladyfish (Elopidae: *Elops*) morphs in the Gulf of Mexico and western North Atlantic Ocean. *Limnology and Oceanography* **49**, 1173–81.

McEwen, G. K., Goode, D. K., Parker, H. J., Woolfe, A., Callaway, H. and Elgar, G. (2009). Early evolution of conserved regulatory sequences associated with development in vertebrates. *PLOS Genetics*. **5**, 1-9.

McFarlane, C. 2009. *Testing the phylogenetic and biological species concepts in the paired lamprey species, Ichthyomyzon unicuspis and I. fossor*. Honours Thesis. University of Manitoba. Manitoba, Canada

McPhee, M.V., Noaker, D.L.G. and Allendorf F.W. (2012). Developmental rate: A unifying mechanism for sympatric divergence in postglacial fishes? *Current Zoology*. **58**, 21-34.

Meyer, A. (1998). Developmental biology - *Hox* gene variation and evolution. *Nature* **391**, 225–8.

Monette, S. N. and Renaud, C. B. (2005). The gular pouch in northern hemisphere parasitic lampreys (Petromyzontidae). *Canadian Journal of Zoology* **83**, 527–535.

Mungpakdee, S., Seo, H.-C., Angotzi, A. R., Dong, X., Akalin, A. and Chourrout, D. (2008). Differential evolution of the 13 Atlantic salmon *Hox* clusters. *Molecular Biology and Evolution* **25**, 1333–43.

Murtha, M. T., Leckman, J. F. and Ruddle, F. H. (1991). Detection of homeobox genes in development and evolution. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 10711–5.

Near, T. J. (2009). Conflict and resolution between phylogenies inferred from molecular and phenotypic data sets for hagfish, lampreys, and gnathostomes. *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution* **312**, 749–61.

Nei, M. and Kumar, S. (2000). Molecular evolution and phylogenetics. Oxford University Press, New York (333 pp).

Neira, F. J. (1984). Biomorphology of the chilean parasitic lampreys *Geotria australis* and *Mordacia lapicida* Petromyzoniformes. *Gayana Zoologia* **48**, 3-40.

Nelson, J. S. (2006). Fishes of the world. 4th Edition. Wiley, New York. 1-601

Osório, J. and Rétaux, S. (2008). The lamprey in evolutionary studies. *Development Genes and Evolution* **218**, 221–35.

Pendleton, J. W., Nagai, B. K., Murtha, M. T. and Ruddle, F. H. (1993). Expansion of the *Hox* gene family and the evolution of chordates. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 6300–4.

Piavis, G. W. (1961). Embryological stages in the sea lamprey and effects of temperature Development. *Fish and Wildlife Service* **61**, 111–43.

Pick, L. and Heffer, A. (2012). *Hox* gene evolution: multiple mechanisms contributing to evolutionary novelties. *Year in Evolutionary Biology* **1256**, 15–32.

Potter, I. C. and Osborne, T. S. (1975). The systematics of British larval lampreys. *Journal of Zoology* **176**, 311–329.

Potter, I. C. (1980). The Petromyzoniformes with particular reference to paired species. *Canadian Journal of Fisheries and Aquatic Sciences* **37**, 1595–1615.

Ravi, V., Lam, K., Tay, B.-H., Tay, A., Brenner, S. and Venkatesh, B. (2009). Elephant shark (*Callorhinchus milii*) provides insights into the evolution of *Hox* gene clusters in gnathostomes. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 16327–32.

Reid, S. B., Boguski, D. A., Goodman, D. H. and Docker, M. F. (2011). Validity of *Lampetra pacifica* (Petromyzontiformes: Petromyzontidae), a brook lamprey described from the lower Columbia River Basin. *Zootaxa* 42–50.

Renaud, C. B. (2011). Lampreys of the world. An annotated and illustrated catalogue of lamprey species known to date. FAO Species Catalogue for Fishery Purposes No. 5. FAO, Rome. 1-109.

Richardson, M. K., Admiraal, J. and Wright, G. M. (2010). Developmental anatomy of lampreys. *Biological Reviews* **85**, 1–33.

Seversmith, H. F. (1953). Distribution, morphology and life history of *Lampetra aepyptera*. *Copeia* **1953**, 225–32.

Shanmugam, K., Featherstone, M. S. and Saragovi, H. U. (1997). Residues flanking the *Hox* YPWM motif contribute to cooperative interactions with *Pbx*. *Journal of Biological Chemistry* **272**, 19081–7.

Sharman, A. C. and Holland, P. W. H. (1998). Estimation of *Hox* gene cluster number in lampreys. *The International Journal of Developmental Biology* **42**, 617–20.

Shimeld, S. M. and Donoghue, P. C. J. (2012). Evolutionary crossroads in developmental biology: cyclostomes (lamprey and hagfish). *Development* **139**, 2091–9.

Sifuentes-Romero, I., Merchant-Larios, H. and García-Gasca, A. (2010). *Hox* gene expression in the embryonic genital system of the sea turtle *Lepidochelys olivacea* (Eschscholt, 1829), a species with temperature-dependent sex determination. *Gene Expression Patterns* **10**, 290–8.

Smith, B. R. and Tibbles, J. J. (1980). Sea lamprey (Petromyzon marinus) in Lakes Huron, Michigan, and Superior: History of invasion and control, 1936-78. *Canadian Journal of Fisheries and Aquatic Science* **37**, 1780-1801.

Smith, J. J., Antonacci, F., Eichler, E. E. and Amemiya, C. T. (2009). Programmed loss of millions of base pairs from a vertebrate genome. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 11212–7.

Smith, J. J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M. S., Yandell, M. D., Manousaki, T., Meyer, A., Bloom, O. E., Morgan, J. R., Buxbaum, J. D., Sachidanandam, R., Sims, C., Garruss, A. S., Cook, M., Krumlauf, R., Wiedemann, L. M., Sower, S. A., Decatur, W. A., Hall, J. A., Amemiya, C. T., Saha, N. R., Buckley, K. M., Rast, J. P., Das, S., Hirano, M., McCurley, N., Guo, P., Rohner, N., Tabin, C. J., Piccinelli, P., Elgar, G., Ruffier, M., Aken, B. L., Searle, S. M. J., Muffato, M., Pignatelli, M., Herrero, J., Jones, M., Brown, C. T., Chung-Davidson, Y-W., Nanlohy, K. G., Libants, S. V., Yeh, C-Y., McCauley, D. W., Langeland, J. A., Pancer, Z., Fritzsch, B., de Jong, P. J., Zhu, B., Fulton, L. L., Theising, B., Flicek, P., Bronner, M. E., Warren, W. C., Clifton, S. W., Wilson, R. K. and Li, W. (2013). Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nature Genetics* 1–9.

Stadler, P. F., Fried, C., Prohaska, S. J., Bailey, W. J., Misof, B. Y., Ruddle, F. H. and Wagner, G. P. (2004). Evidence for independent *Hox* gene duplications in the hagfish lineage: a PCR-based gene inventory of *Eptatretus stoutii*. *Molecular Phylogenetics and Evolution* **32**, 686–94.

Strahan, R. (1960). A comparison of the ammocoete and macrophthalmia stages of *Mordacia mordax* and *Geotria australis* (Petromyzonidae). *Pacific Science* **14**, 416–20.

Taggart, J. B., Hynes, R. A. and Ferguson, A. (1992). A simplified protocol for routine total DNA isolation from salmonid fishes. *Journal of Fish Biology* **40**, 963–65.

Tahara, Y. (1988). Normal stages of development in lamprey, *Lampetra reissneri* (Dybowski). *Zoological Science* **5**, 109–18.

Takio, Y., Pasqualetti, M., Kuraku, S., Hirano, S., Rijli, F. M. and Kuratani, S. (2004). Evolutionary biology: Lamprey *Hox* genes and the evolution of jaws. *Nature* **429**, 1–3.

Takio, Y., Kuraku, S., Murakami, Y., Pasqualetti, M., Rijli, F. M., Narita, Y., Kuratani, S. and Kusakabe, R. (2007). *Hox* gene expression patterns in *Lethenteron japonicum* embryos—Insights into the evolution of the vertebrate *Hox* code. *Developmental Biology* **308**, 606–20.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* **28**, 2731–9.

Vladykov, V. D. and Folleett, W. I. (1958). Redescription of *Lampetra ayresii* (Gunther) of western North America, a species of lamprey (Petromyzontidae) distinct from *Lampetra fluviatilis* (Linnaeus) of Europe. *Journal Fisheries Research Board of Canada* **15**, 47–77.

Vladykov, V. D. and Follett, W. I. (1965). *Lampetra richardsoni*, a new nonparasitic species of lamprey (Petromyzonidae) from western North America. *Journal of the Fisheries Research Board of Canada* **22**, 139–58.

Vladykov, V. D. and Kott, E. (1980). Description and key to metamorphosed specimens and ammocoetes of Petromyzonidae found in the great lakes region. *Canadian Journal of Fisheries and Aquatic Sciences* **37**, 1616–1625.

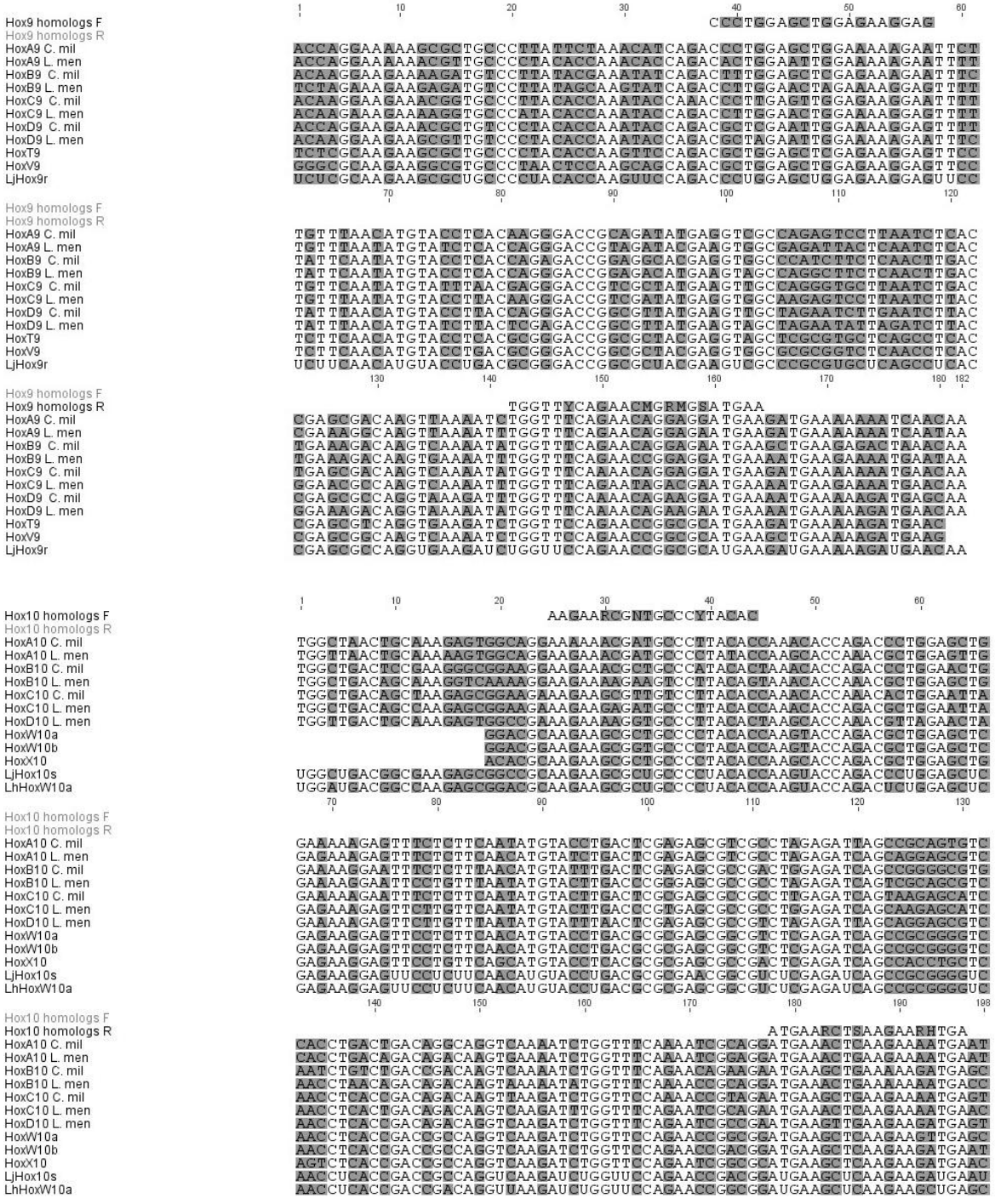
Yamahira, K. and Nishida, T. (2009). Latitudinal variation in axial patterning of the medaka (Actinopterygii: Adrianichthyidae): Jordan's rule is substantiated by genetic variation in abdominal vertebral number. *Biological Journal of the Linnean Society* **96**, 856–66.

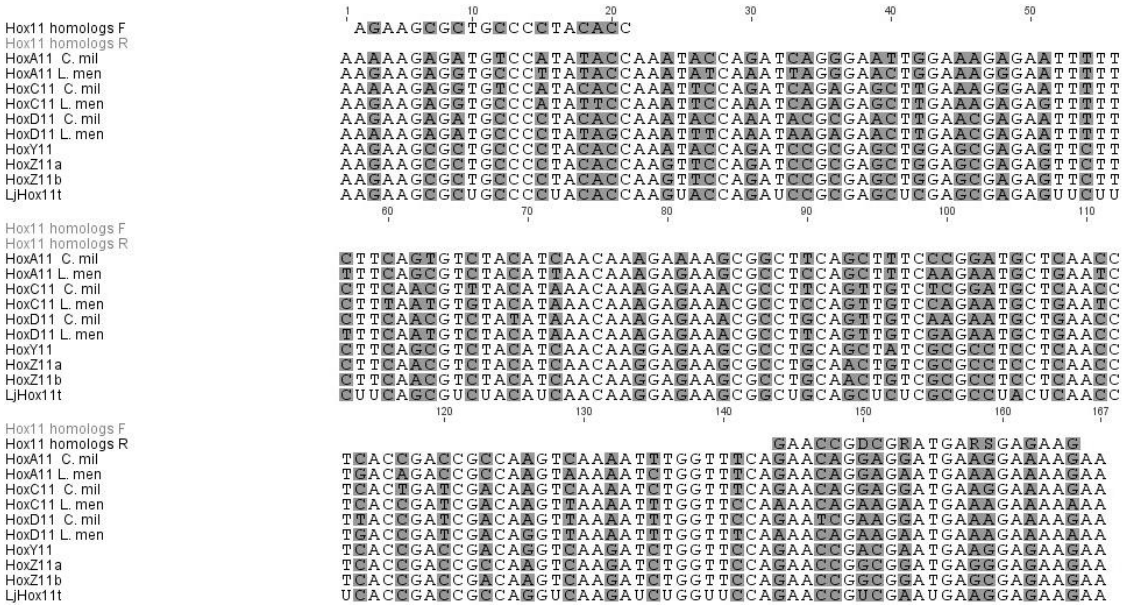
Wolpert, L., Jessell, P., Meyerowitz, E., Robertson, E. and Smith, J. (2007). Principles of development: Third Edition. Oxford university press. Chapters 4 and 9.

Youson, J. H. (1997). Is lamprey metamorphosis regulated by thyroid hormones? *Integrative and Comparative Biology* **37,** 441–60.

Appendix A

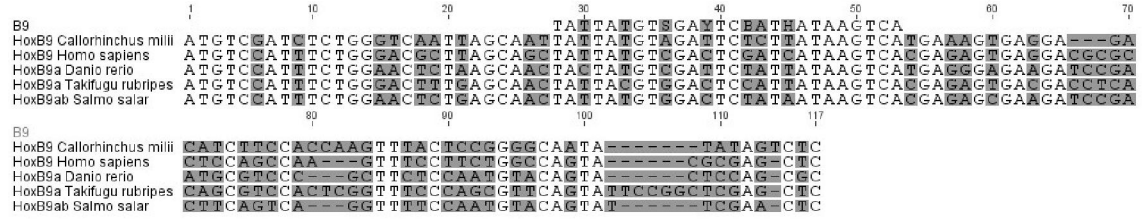
Hox9, Hox10 and *Hox11* homeobox alignments and primer design. (Aligned sequences from Table 2.2 and primers from Table 2.3)

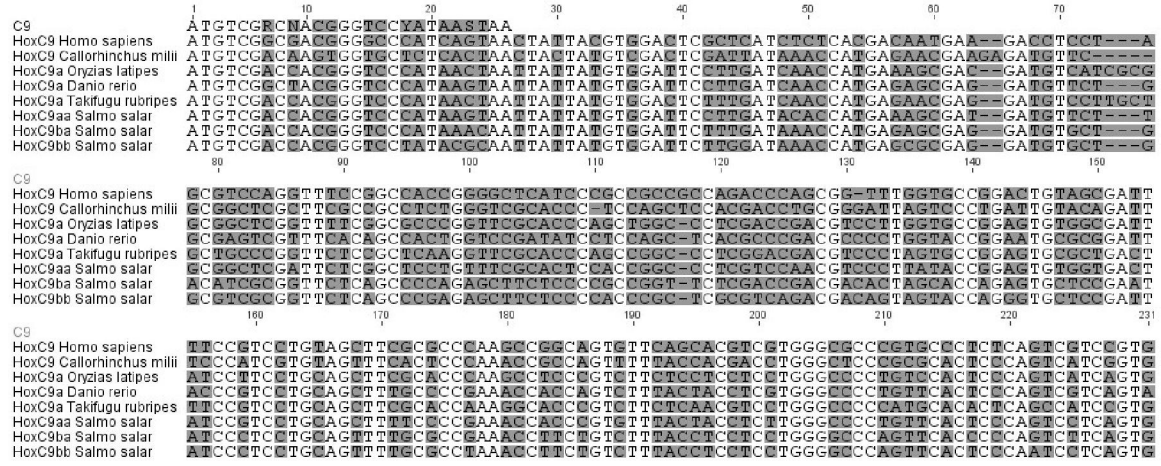


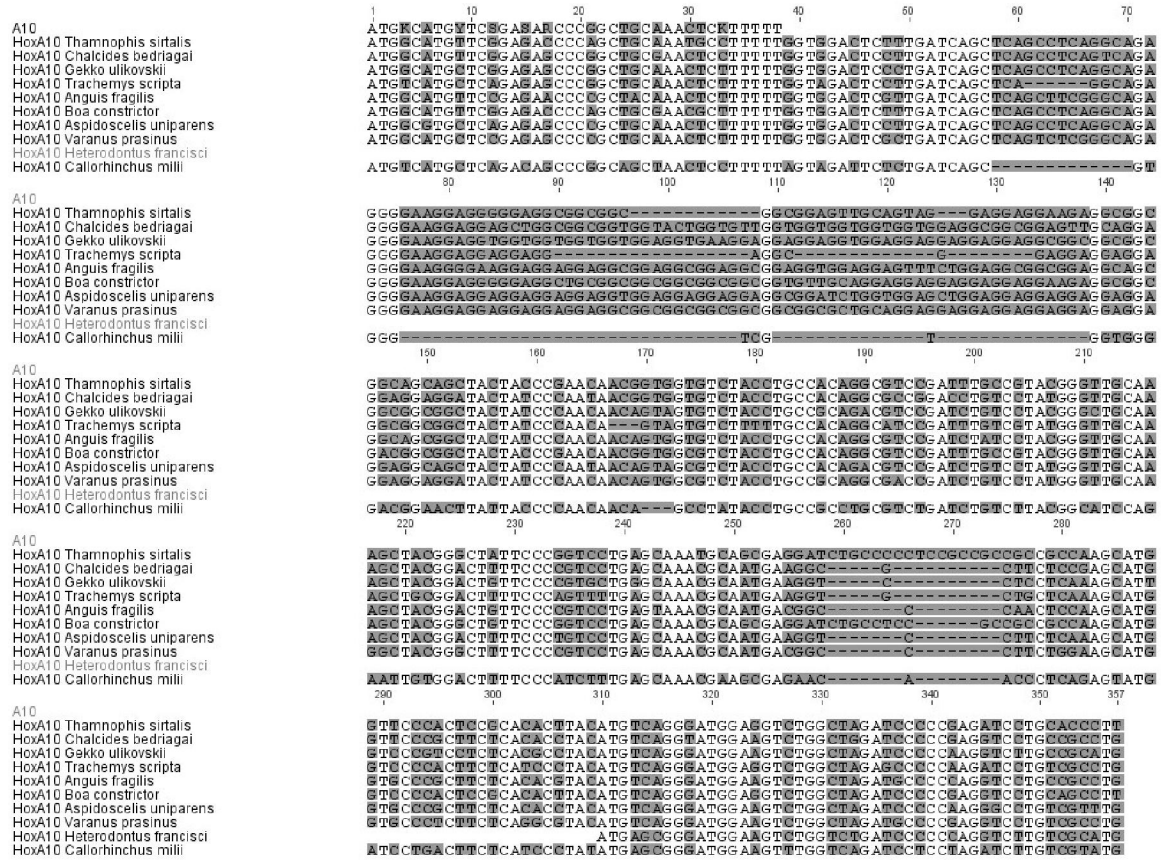

Hox11 homologs F
Hox11 homologs R
HoxA11 C. mil
HoxA11 L. men
HoxC11 C. mil
HoxC11 L. men
HoxD11 C. mil
HoxD11 L. men
HoxY11
HoxZ11a
HoxZ11b
LjHox11t

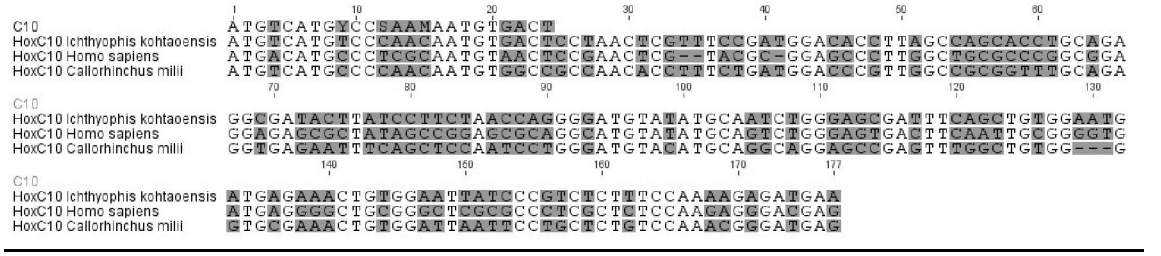
Appendix B

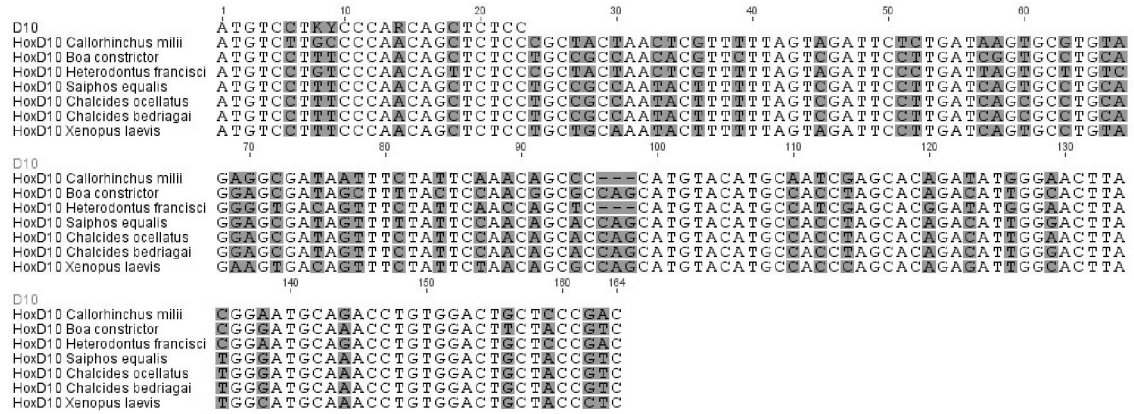
Nucleotide alignments of *Hox9, Hox10* and *Hox11* genes exon1 for exon1 primer design. (Aligned sequences from Table 2.9 and primers from Table 2.10)

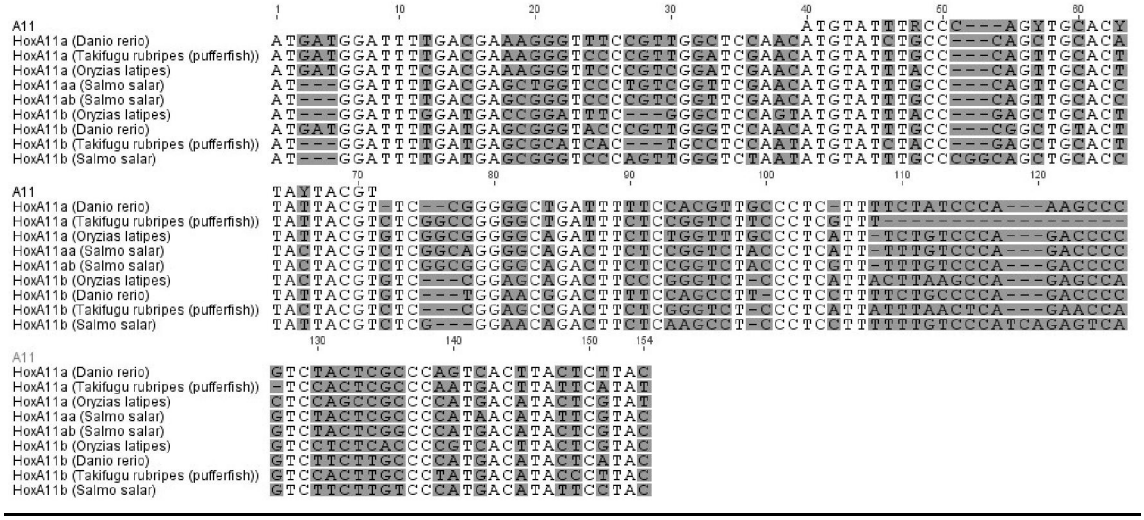


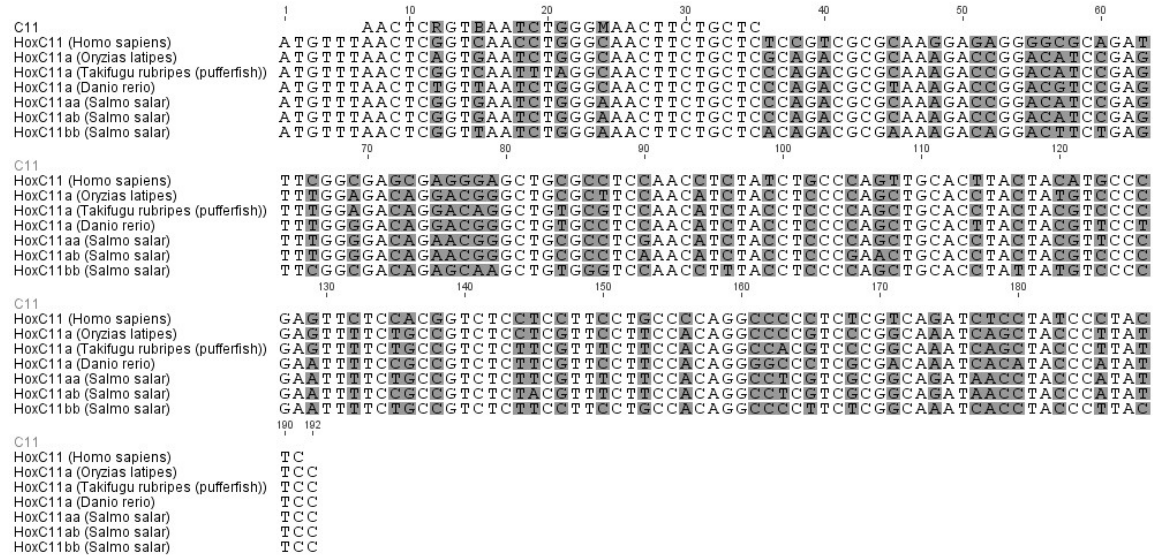


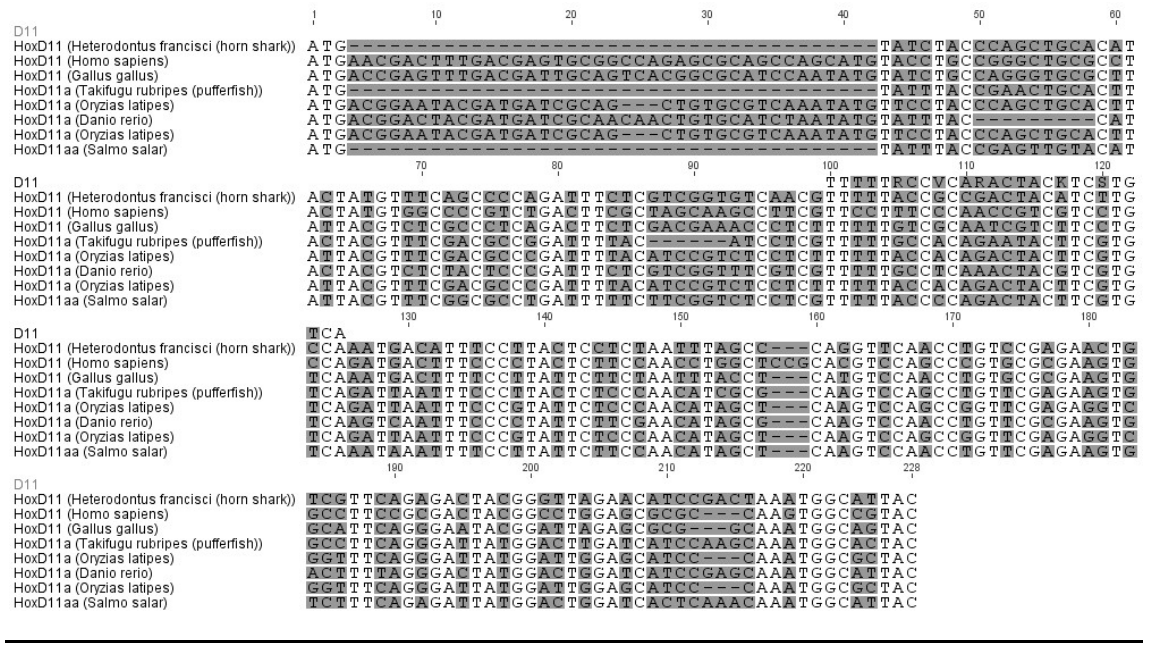

A10
HoxA10 Thamnophis sirtalis
HoxA10 Chalcides bedriagai
HoxA10 Gekko ulikovskii
HoxA10 Trachemys scripta
HoxA10 Anguis fragilis
HoxA10 Boa constrictor
HoxA10 Aspidoscelis uniparens
HoxA10 Varanus prasinus
HoxA10 Heterodontus francisci
HoxA10 Callorhinchus milii


C10
HoxC10 Ichthyophis kohtaoensis
HoxC10 Homo sapiens
HoxC10 Callorhinchus milii


D10
HoxD10 Callorhinchus milii
HoxD10 Boa constrictor
HoxD10 Heterodontus francisci
HoxD10 Saiphos equalis
HoxD10 Chalcides ocellatus
HoxD10 Chalcides bedriagai
HoxD10 Xenopus laevis


A11
HoxA11a (Danio rerio)
HoxA11a (Takifugu rubripes (pufferfish))
HoxA11a (Oryzias latipes)
HoxA11aa (Salmo salar)
HoxA11ab (Salmo salar)
HoxA11b (Oryzias latipes)
HoxA11b (Danio rerio)
HoxA11b (Takifugu rubripes (pufferfish))
HoxA11b (Salmo salar)


C11
HoxC11 (Homo sapiens)
HoxC11a (Oryzias latipes)
HoxC11a (Takifugu rubripes (pufferfish))
HoxC11a (Danio rerio)
HoxC11aa (Salmo salar)
HoxC11ab (Salmo salar)
HoxC11bb (Salmo salar)


D11
HoxD11 (Heterodontus francisci (horn shark))
HoxD11 (Homo sapiens)
HoxD11 (Gallus gallus)
HoxD11a (Takifugu rubripes (pufferfish))
HoxD11a (Oryzias latipes)
HoxD11a (Danio rerio)
HoxD11a (Oryzias latipes)
HoxD11aa (Salmo salar)

Appendix C

Lamprey *Hox* gene sequence identified in this study (Underline indicates previously identified sequences, superscript 1 indicates sequence identified in general Hox screening Sections 3.5.3 and 4.5.2, superscript 2 indicates sequence identified in orthologous *Hox10* gene identification Sections 3.5.2 and 4.5.1, superscript 3 indicates sequence identified in *P. marinus* genome search Sections 3.3.1 and 4.3.1)

Entosphenus tridentatus

[1]*Hox3* putative, partial sequence exon2
CACTTCAACCGGTACCTCTGCCGCCCTCGACGAGTTGAAATGGCCAACCTACTTAACCTCACCGAGAGGCAGATCAAGA
TA

[1]*Hox8* putative, partial sequence exon2
CACTTCAACCGGTACCTGACGCGCAAGCGACAAATCGAGGTGTCGCACGCGCTCGGACTGACCGAGCGCCAGGTCAAG
ATC

[2]*Hox10* putative, partial sequence exon2
CCTGCCGGGTGGCTCACGGTGAAAAGCGGCCGCAAGAAACGCTGCCCCTACACCAAGTACCAGGCGCTGGAGCTCGAG
AAAGAATTCCTCTTCAACATGTACCTGACGCGCGAGCGGCGTCTCGAGATCAGCCGCGGGGTCAACCTCACCGACCGC
CAGGTCAAGATCTGGTTCCAGAACCGACGGATGAAGCTGAAGAAGACGAATCGAGAGATTCGTGTCCGCG

Eudontomyzon vladykovi

[1]*Hox4/5* putative, partial sequence exon2
CACTTCAACAGGTACCTCGCCCGGAGACGGCGGGTCGAGATCGCGCACTCGCTGTGCCTGAGCGAGCGCCAGGTCAAG
ATC

[1]*Hox4/5/6* putative, partial sequence exon2
CACTTCAACAGGTACCTCACCCGCCGGCGCCGCATCGAGATCGCGCACGCGCTCTGCCTCACCGAGCGCCAGATCAAG
ATC

[2]*Hox10* putative, partial sequence exon2
CCGGGCGGCTGGCTGACGGCGAAGAGCGGCCGCAAGAAGCGTTGCCCCTACACCAAGTACCAGACCCTGGAGCTCGAG
AAGGAGTTCCTCTTCAACGTGTACCTGACGCGCGAACGGCGTCTCGAGATCAGCCGCGGGGTCAACCTCACCGACCGC
CAGGTCAAGATCTGGTTCCAGAACCGACGGATGAAGCTCAAGAAGATGAATCGCGAGATCCGTGTCCGCG

Geotria australis

[1]*Hox2* putative, partial sequence exon2
CACTTCAACAGCTACCTGTGCCGGCCGCGGCGCGTCGAGATCGCAGCTCTGCTCGACCTCACGGAGCGCCAGGTCAAA
GTG

[1]*Hox6/7* putative, partial sequence exon2
CACTTCAACCGGTACCTGTCGCGACGACGCAGGATCGAGATCGCACACGCGCTCTCGCTGAGCGAGCGACAGATCAAG
ATC

[1]*Hox6/14* putative, partial sequence exon2
CACTTCAACAGGTACCTGACGCGAGAGCGGCGCCTCGAGATTAGCCGAGGCGTCAACCTCACGGACCGACAGGTCAAG
ATC

[2]*Hox10* putative, partial sequence exon2
CCGGGCGGGTGGCTGACGGCTAAGAGCGGGCGCAAGAAGCGCTGCCCCTACACCAAATACCAGACGCTGGAGCTGGA
GAAGGAGTTCCTCTTCAACACGTACCTGACGCGAGAGCGGCGCCTCGAGATCAGCCGAGGCGTCAACCTCACGGACCG
ACAGGTCAAGATCTGGTTCCAGAACCGGCGCATGAAGCTGAAGAAGATGAACCGCGAGATCCGTGTTCGTG

Ichthyomyzon castaneus

[1]*Hox4/5/6/7* putative, partial sequence exon2

CACTTCAACAGGTACCTCACCCGCCGGCGCCGCATCGAGATCGCGCACGCGCTCTGCCTCACCGAGCGCCAGATCAAG
ATC

[2]*Hox10* putative, partial sequence exon2

CCGGGCGGCTGGCTGACGGCCAAGAGTGGACGCAAGAAGCGCTGCCCCTACACCAAGTACCAGACCCTGGAGCTCGAG
AAGGAGTTCCTCTTCAACATGTACCTGACGCGCGAGCGGCGTCTCGAGATCAGCCGCGGGGTCAACCTCACCGACCGC
CAGGTCAAGATCTGGTTCCAGAACCGACGGATGAAGCTCAAGAAGATGAATCGCGAGATCCGTGTCCGCG

Ichthyomyzon fossor

[1]*Hox5/6/7* putative, partial sequence exon2

CACTTCAACCGGTACCTCACCCGCCGGCGCCGCGTCGAGATCGCCCACTCGCTCTGCCTCACCGAGCGCCAGATCAAGA
TC

[1]*Hox1/8* putative, partial sequence exon2

CACTTCAACAGGTACCTGACGCGCGCCCGCCGCGTCGAGATCGCCGCCGCGCTGCAACTCAACGAGACCCAGGTGAAG
ATC

[1]*Hox8* putative, partial sequence exon2

CACTTCAACAGGTACCTGACGCGCAAGCGGCGCATCGAGGTGTCCCACGTGCTCGGCCTCAGCGAGCGCCAGGTCAAG
ATC

[2]*Hox10* putative, partial sequence exon2

CCGGGYGGCTGGCTGACGGCCAAGAGTGGACGCAAGAAGCGCTGCCCCTACACCAAGTACCAGACCCTGGAGCTCGA
GAAGGAGTTCCTCTTCAACATGTACCTGACGCGCGAGCGGCGTCTCGAAATCAGCCGCGGGGTCAACCTCACCGACCG
CCAGGTCAAGATCTGGTTCCAGAACCGACGGATGAAGCTCAAGAAGATGAATCGCGAGATCCGTGTCCGCG

Ichthyomyzon gagei

[1]Putative *Hox5/6* putative, partial sequence exon2

CACTTCAACCGCTACCTGACGAGGCGACGCCGCATCGAGGTGGCGAACGCGCTCTGCCTGAGCGAGCGCCAGATCAAG
ATC

Ichthyomyzon unicuspis

[1]*Hox4/5* putative, partial sequence exon2

CACTTCAACCGGTACCTGACGAGGCGACGCCGCGTGGAGGTGGCGAACGCGCTCTGCCTGAGCGAGCGCCAGATCAAG
ATC

[2]*Hox10* putative, partial sequence exon2

CCGGGCGGCTGGCTGACGGCCAAGAGTGGGCGCAAGAAGCGCTGCCCCTACACTAAGTACCAGACCCTGGAGCTCGAG
AAGGAGTTCCTCTTCAACATGTACCTGACGCGCGAGCGGCGTCTCGAAATCAGCCGCGGGGTCAACCTCACCGACCGC
CAGGTCAAGATCTGGTTCCAGAACCGACGGATGAAGCCCAAAAAGATGAATCGCGAGATCCGTGTCCGCG

Lampetra aepyptera

[2]*Hox10* putative, partial sequence exon2

CCGGGCGGCTGGCTGACGGCGAAGAGCGGCCGCAAGAAGCGCTGTCCCTACACCAAGTACCAGACGCTGGAGCTCGA
GAAGGAGTTCCTCTTCAACATGTACCTGACGCGCGAACGGCGTCTCGAGATCAGCCGCGGGGTCAACCTCACCGACCG
CCTGGTCAAGATCTGGTTCCAGAACCGACGGATGAAGCTCAAGAAGATGAATCGCGAGATCCGTGTCCGCG

Lampetra pacifica

[1]*Hox4/5/6/7* putative, partial sequence exon2

CACTTCAACAACTACCTCACCCGCCGGCGCCGCATCCAGATCGCGCACGCGCTCTGCCTCACCGAGCGCCAGATCAAGA
TC

[1]*Hox10a* putative, partial sequence exon2

CTCTTCAACATGTACCTCACGCGCGAGCGGCGGCTCGAGATCAGCCGCGGGGTCAACCTCACCGACCGCCAGGTCAAG
ATC

[2]*Hox10b* putative, partial sequence exon2
CCGGGCGGCTGGCTGACGGCGAAGAGCGGCCGCAAGAAGCGCTGCCCCTACACCAAGTACCAGACGCTCGAACTCGA
GAAGGAGTTCCTTTTCAACATGTACCTGACGCGCGAACGGCGTCTCGAGATCAGCCGCGGGGTCAACCTCACCGACCG
CCAGGTCAAGATCTGGTTCCAGAACCGACGGATGAAGCTCAAGAAGATGAATCGCGAGATCCGTGTCCGCG

Lampetra richardsoni
[1]*Hox3* putative, partial sequence exon2
CACTTCAACCGGTACCTCTGCCGGCCGCCTCGCGTCGAGATGGCCAACCTGCTGAACCTCACCGAGCGCCAGATCAAGA
TC

[1]*Hox11* putative, partial sequence exon2
TTCTTCAACGTCTACATCAACAAGGAGAAGCGACTGCAGCTTTCGCGCCTGCTCAACCTCACCGACCGACAGGTCAAGA
TC

[2]*Hox10* putative, partial sequence exon2
CCGGGCGGCTGGCTGACGGCAAAGAGCGGCCGCAAGAAGCGTTGCCCCTACACCAAGTACCAGACGCTCGAACTCGAG
AAGGAGTTCCTCTTCAACATGTACCTGACGCGCGAACGGCGTCTCGAGATCAGCCGCGGGGTCAACCTCACCGACCGC
CAGGTCAAGATCTGGTTCCAGAACCGACGGATGAAGCTCAAGAAGATGGATCGCGAGATCCGTGTCCGCG

Petromyzon marinus
[3]*Hox1w* putative extended, partial sequence exon2
ATGGACAATGCGGCCATGAGTTTCCTGGAGTATTCGTCATTCTGCGGCGGCGGCGGCGGTGGCGGCGGCGGCGGCGAG
GCGGGCGGCTTCACCCCTCGCTCGTTCGCTGCGGCAGCCGCGGGAGACCCATGTCTCGCTCCCTTCCAGTCGTGCGCCG
TGAGCGCTGCGGCGGGCGCGAGCGGCGTCGGCGATGGGGAAGGGCGCTACCACAACCATCACCACCACCACCATCACC
ACCCTCATCATCATCACCACCATCATCAACAGCAACAGCAGCAGAATGAGCAGCAGCAGCATCAGCAGCAGCAGCAGA
GTGGCTACCAGCCGCACAACTTCACCTTCACTGGCTCGTACGCCCCCGCGACGAACTCGGGGTGCTATTCGCCCGGCCA
AGCCTATGCGGGGGCGGCGAGTTACGGGGTCTACTGTTCGCAGGGGGCCGACTTCGGTGGCGGCGGCGGTGGCAGTGG
CGTTGGTGGAGCCGTGAGCGGGTTGCAGGCGACGCACGGCTCGCTCGTCGGTGGTTACGGCGCCTCGGTGTTCCACCAC
GACTCGCTGCAGGCTTCAGACCTCGGCGATCTTCAGCCGCTCGCCAGCTGCTGCCTGCAGCAAGCGGGCCCTGTGTCGC
ACACTCAGCATCATCATCAGCAGCAGCAGCATCATCAGCAGCAGCAGCAACACGGGTCGCCGAATGCGTCGGAGCCCA
CGCCGCCTTCGCACTGCACGTTCGAGTGGATGCGAGTGAAGAGAAATCCGCCGAAAACAGCGAGGATCGGCGACATGA
GCTGCGCGCAGGTGGGCGCCCTCCCGTGCGGCTCGGGAGGAGCGGGAGGCGGCGGAGTCGGACTGAGCATCGGCGGG
GGCAACAACGGCCTCGCGATGGGCGGAGGCATCGCGACCCATCGGACCAACTTCAGCACCAAGCAGCTGACGGAGCTG
GAGAAGGAGTTCCACTTCAACAAGTACCTGACGCGCGCGCGCCGCGTCGAGATCGCCGCCGCGCTGCAGCTCAACGAG
ACCCAGGTGAAGATCTGGTTCCAGAACCGGCGCATGAAGCAGAAGAAGCGAGAGAAGGAGGGCCACAGCCTGCAGAC
GCTCGCGTCTCCTATCGCATCCTCTGCAGCTGCTGCTGCTGCTGCTACTGCTGCTGCTGGACCTTCAGTCTCGTCCG
CGGCCTCCTCCGAGGGGTCCTCGCCGAGTCGCTCGCCGGCCTCAGAGCCACCGGCTGATGACGCCTCGCCCTGA

[3]*Hox1* putative, partial sequence exon2
CAGCAGCAGCAGCGAACCAACTTCACCACCAAGCAGCTGACCGAGCTGGAGAAAGAGTTCCACTTCAGCAAGTACCTG
ACGCGCGCGCCGCGTCGAGATCGCCGCCGCCCTGCAGCTGAACGAGACCCAGATCAAGATCTGGTTCCAGAACCGG
CGCATGAAGCAGAAGAAGCGGGAGCGCGAGTGTCGCGACCCCGCCAGCAAGGGGCTCTCGGCGGCCGCCGGAGAAGG
GCACGCCGAGGTCGGCGGCGTCTCGGCGAAGGGGTTCGCGTTGGGAGTCGCGATCGGGGCCCGTGAGAGATCGTCGGG
TCTCGGGAGCGGGCCGACGTCGCCTGCTGAGCCAGACTCGGCCGCGGTTGCTGCCGCCGCTGCTGCTGATGATGATGAC
GACGACGACGATCATCACGATCATCGCGATGATGATGATGATGATCGTGATGATGACGATTATGACATATCGCCCT
GA

[3]*Hox2* putative, partial sequence exon1
ATGAACTACGAGTTTGAGCGGGAAACTGGGTTCATCAACAGCCAGCCCTCGCTCGCCGAGTGCCTGACAGCCTTTCCCC
CTGTCGCCGCCGAGACATTTCAAAGTTCATCAATCAAGAGCTCGACGCTTTCACCCCCGACGACACTCAGTCCTCCTCC
TCCTTTCGAGTCCATCATCCCGGCTCTGCTGCAGCACGGCACTCGGCCAGCGAGCCGGGCGCGGCCGGCTCCCTCCGGC
CGTCTCCCGTCGCCCGCTCAGCCTCCCGAGTACCCGTGGATGAGGGAGAAAAAATCGTCCAAGAAACAACAACAACAG
CAGCAGCAGCAGCAGCAACAGCAGCAGCATCAGCAGCAGCAGCATCTACAGCCCGCGCTTCCAGTGCCATCGTCGGTG
GCCGGGCCCGTCTCGCCG

[3]*Hox3* putative, partial sequence exon2
GCATCCAAGCGGGCGCGCACCGCCTACACCAGTGCCCAGCTGGTGGAGCTCGAGAAGGAGTTCCACTTCAACCGCTAC
CTCTGCCGTCCGCGCCGCATCGAGATGGCGAACCTGCTGAGCCTCACCGAGCGCCAGATCAAGATCTGGTTCCAGAACC
GCCGCATGAAGTACAAGAAGGACCAGAAGATTAAAGGTTTCGGTGGCTGCGGTGTTGGTGTGGGCGTTGGCCCACTGT
CGCCGCCGTGCGTCGCTGCGGGTGGCGTGGGTTTGGGGTCGAGGCTGTCACCCCCTCGCGCGTCACCGCCGGGGCTCCC
CCCTTCGTGTCACTTCGCCCTGGGAGGTCACCCGCTGCACGCCTACGACCACCGCTCGTCGCCGCACTTCTTCGGCAAGT
GCCACGAGGCCGCGGCCTTCGCGACGTCTCCGGGTTTCGTGGCTCACGCGGGAGACGGCGGCGGCGGCGGCGGCGGCG
GATCTCCATCATCGCCATCGTCGTCTTCGTCGTCGTCGCTGATGCTCAGACGAGGCTACTCGGGCGGCGGCGGCGGCGG

CGGCGGTGGTGTCGGTGGCCGGCGATGCCCCCAAAAGTCTCGTCAGGGGCCACGATTTTCACGCCGTACTTTGCCGGAG
GGTCTGACGGGTTGCCCTC

[3]*HoxG4* putative extended, partial sequence exon2
TCGAAGCGCTCCCGCACGGCCTACACGCGCCAGCAGGTGCTGGAGCTGGAGAAGGAGTTCCACTTCAACCGCTACCTC
ACCAGGAGACGGCGAATCGAGATCGCGCACTCGCTGTGCCTGAGCGAGCGCCAGATCAAGATCTGGTTCCAGAACCGG
CGCATGAAGTGGAAGAAGGACCACAAGTTGCCCAACACCAAGATCCGTTCGGCCACGGCCACCGCCGCCGCGGCCATC
GCCACCGCCGGCGGAGGCTCCTCGTCCTCGGCCGGCGGAGGAGGCGGAGGCCCCGACACTCCGGGCGGTAGCGGCATG
GCGGTCACGGTGGCGCCGGGCTCCGTGGCGGCCCACGCCGCGTCGGGCCAGAGCGGTCCCACCGGTGGGCGTGGCCTC
AACCGCCGCCGCTTGCAAAGGCGGTGGCCACCAGGCCCGCACGGCCGATGGCTTGAGCAGGCCCTTGTGAGGAGGACG
AGGCGAGCAAGCCGTGGCCGTGATGGCTGCTGCTGCTGGTGGTGGTGGTTGGTGATGATGATGTTGTTGTTGTGTA
GAGCCAGAGGGACTGTAGCTGTGCAGCGGGATAGCAGCAAGTTGAAGTAG

[3]*Pm88-H* putative extended, partial sequence exon2
ACCAAACGGGCGCGCACGGCCTACACGCGTCAGCAGGTGCTGGAGCTCGAGAAGGAGTTCCACTTCAACCGCTACCTC
ACGCGGCGGCGCCGCCTGGAGATCGCTCACTCGCTGGCGCTGAGCGAGCGCCAGGTCAAGATCTGGTTCCAGAACCGG
AGGATGAAGTGGAAGAAAGACCACAAACTGCCCAACACCAAGATGAGGCCGGGTCCTCCCGGACTTCCCCAGGGTCTG
TCTCATCGCCCCGACTGCAGTAGAACATAA

[3]*HoxN5* putative extended, partial sequence exon2
GACGGCATGAGCGGACCGGAGGGCAAGCGCTCGCGCACCGCCTACACGCGCTACCAGACGCTGGAGCTGGAGAAGGA
GTTCCACTTCAACCGCTACCTCACCCGCCGGCGTCGCATCGAGATCGCGCACGCGCTCTGCCTCACCGAGCGCCAGATC
AAGATCTGGTTCCAGAACCGGCGCATGAGGTGGAAGAAGGACAACAAGCTTAAGAGCCTGAGCATGGCTAACCAAGC
CGGCGCCTACCACGGCTAA

[3]*HoxF5* putative extended, partial sequence exon2
GGACACGACGGTAGGAAGGGCAGGAGGAGCTACTCGCGCCACCAGAGCCTGGAGCTGGAGAAGGAGTTCCACTTCAA
CCGCTACCTCGCCCGGAGACGTCGAGTCGAGATCGCGCACTCGCTGTGCCTGAGCGAGCGCCAGGTCAAGATCTGGTTC
CAGAACCGGCGCATGAAGTGGAAGAAGGAGCGTCGAGGCATTGACGCGCGGCGTGAAGAGGTGGAGGAAGGGGAGG
AGATGGTGGAGGACGAGGTGGAGGAGAGTCAAGACGGCGAGAAGACGGTCGACATCAAGGAGTGA

[3]*HoxK6* putative extended, partial sequence exon2
GGCCACGATGGCAAGCGAGGCCGGCAGACCTACTCGCGCTACCAGACCCTGGAGCTGGAGAAGGAGTTCCACTTCAAC
CGCTACCTGACGAGGAGGCGCCGCGTCGAGATCGCCCACTCGCTCTGCCTCACCGAGCGCCAGATCAAGATCTGGTTCC
AGAACCGCCGCATGAAGTGGAAGAAGGAGAACAGGGGTCCGAACCCATCGCCCGAGGAGGCTCACGAGGAGGAAGAC
GACGAGGAGGAGGAAGAAGAAGAGGAGAAGAAGGCCGAATGA

[3]*Hox7* putative, partial sequence exon1
ATGAGTTCGTCGTATTATGTGAACTCTCTTTTCCCCAAATATACCGCGGGAGCCTCGGTCTTCCAGCCCGGCGCGGGGCT
GGTGGAGCCTCGGCCCGCGTGCGAATTCTCCGCCGACTCGTTCAGCTCGGGCGCTGCCGCAGCCGCAGCAGCAGCAGC
AGCGGCCGCCGCGGCCTCGTTCCCGGGCTCCGTGTCGGGGCTCTACGGACTCTATCAGCAGGGCGGCGGCGGCGGTCC
ACCGGGCTCCGTCTACGGCGCGGGCTACGCGCTCGGCGGCTCGGCGCCGCTCGGGCTCGCGTGCACCGCCTACGAGCA
GCCGTTCGGGCATAGCCCCTACGGGAAGGAGGCGGCCGCGGCGGCGGCGGTGGCGGAGGCGGACGGGGAGAGGGCGC
TGGCGGCCAGGAGCGACGCGGGGCTGCGCATTTACCCGTGGATGCGCTCCACGGCAGGTAGT

[3]*HoxQ8a* putative extended, partial sequence exon2
GGGCCGGGCAGGCGACGGGGCCGCCAGACCTACTCGCGCTTCCAGACGCTGGAGCTGGAGAAGGAGTTCCTCTTCAAC
CCGTACCTGACGCGCAAGCGGCGCATCGAGGTGTCGCACGCGCTGGGCCTCACCGAGCGCCAGGTCAAGATCTGGTTC
CAGAACCGGCGCATGAAGTGGAAGAAGGAGAACAACAAAGACAAATTCCCGGCCAGGAAGGAGGACATTGAAGAGC
CCGCGAAGAAGGCGGAGGAAAACTCGGGCGACGAGGAGTCGCAAGGCACCTCCAAGGAGTGA

[3]HoxR8 putative extended, partial exon1
ATGAGCTCGTACTTTGAAGATTCGCTCCACGCGGGATACTTCGACTGCCGCTTCGCCGCTCACGAAGCCGCGATGCGCA
GCGCGGCCGCCGGCGACCACTCCGGAGGGACCTTCGACAGGGCCGGGGGCTTCGGCGTGGGGGGCTTCAACGGGGCCC
CCTACGGACTCGGCTACCACGGGGTGCCGCCCTTCGAGGGCGAAACGTCGGACTCCGGCTACCCGAGGTTGCTCGACC
CGACCGCCCCCTACGGAGCCGCCTCGCCCGGCGCACGGGGGCGAGGCGGAGCGGGGGTCGAGGGGGTCGAGGTGGGC
GCCGCGCGGGGCGGAGACGCGGGCGGCTCGGTCTTCCCGTGGATGAGACCCCAAGGT

[3]HoxR8 putative extended, partial exon 2
CCTGCGCGCAGGCGAGGCCGCCAGACCTACAGCCGCTACCAGACCCTGGAGCTGGAGAAGGAGTTCCTCTTCAATCCG
TACCTGACGCGGAAGCGGCGCATCGAGGTGTCCCACGTGCTGGGCCTCAGCGAGCGCCAGGTCAAGATCTGGTTCCAG
AACCGGCGCATGAAGTGGAAGAAGGAGAACAACAGGGACAAGTTCCCCAGCACCTGCCGGCCCAGTGGCAGCACGAG
TGAGGAGGGGTCATCGCATTCCCCCGATAGCCAAACTCAAGACAAACCACTCTGCGACATATAA

[3]*Hox8* putative, partial sequence exon2
ATGAATTCGTACTTCGCGAACCCTCTCTTCGCGAAGTACAAGGCCGGCGACGCGCTGCGCCCCAGCTACTATGACTGCC
GCCTCGCAACGGACGCCATGAGTGGCCGGCAGGCCGTCTTCTACGGTGCCGGCGCCGGAGCCGGAGGAGCCGGAGGA
GGAGGAGGAGGAGGCCCCCCGGCATTCGAGCACCCGGCCGGCACCGGCGCGGGCACGGGCGGAGCCGCGGGAGTTCA
CCACCAGGAGTTCTACCACAACGGAAGTTCGGCGGCGGCGGCGGCTGCCGCGTACCAAGCGGCCGGTTCGTGCGC
ACTGGACGCAGCGGGTAACTTCTACGGCTACCCGGCCTTCCCGAGGCAGCCCGTGTTCGCCCTGCAGCAGGAGTGCGGT
GGCGTGGTGCAGTACCCGGGACCCGACTTCAAGTCGTTCGGGCCCGGCGCCGAGGAGGAGGCCGAGTACCCGGGTCAG
GGATCCTCGTCGGCGCAGCTCTTCCCCTGGATGAGGTCTCAA

[3]HoxT9 putative extended, partial sequence exon2
CACGCGCGACCCTCTCGCAAGAAGCGCTGCCCCTACACCAAGTTCCAGACGCTGGAGCTCGAGAAGGAGTTCCTCTTCA
ACATGTACCTCACGCGGGACCGGCGCTACGAGGTAGCTCGCGTGCTCAGCCTCACCGAGCGTCAGGTGAAGATCTGGT
TCCAGAACCGGCGCATGAAGATGAAAAAGATGAAC AAGGACAGAAGCAAAGATCCGCGTTGCTGA

[3]*HoxV9* putative extended, partial sequence exon2
CACGCGCGCGCTGGGCGCAAGAAGGCGTGCCCTAACTCCAAGCAGCAGACGCTGGAGCTGGAGAAGGAGTTCCTCTTC
AACATGTACCTGACGCGGGACCGGCGCTACGAGGTGGCGCGCGGTCTCAACCTCACCGAGCGGCAAGTCAAAATCTGG
TTCCAGAACCGGCGCATGAAGCTGAAAAAGATGAAGAAAGAGAAGAGCGAGGATCAGGGCTGA
AAAGAGAAGAGCGAGGATCAGGGCTGA

[3]*Pm98-s* putative extended, partial sequence exon2
CACGCGCGAGCGGGCCGCAAGAAGCGCTGCCCCTACTCCAAGCAGCAGACGCTGGAGCTGGAGAAGGAGTTCCTCTTC
AACATGTACCTGACGCGGGACAGGCGCTACGAGGTGGCGCGCGGTCTCAACCTCACCGAGAGGCAGGTGAAGATC
TGGTTCCAGAACCGGCGCATGAAGCTCAAGAAGATGAAGAGAGATAAAAGTGATGATCTGTAA

[3]*HoxW10a* putative extended, partial sequence exon2
ACGGCCAAGAGTGGACGCAAGAAGCGCTGCCCCTACACCAAGTACCAGACGCTGGAGCTCGAGAAGGAGTTCCTCTTC
AACATGTACCTGACGCGCGAGCGGCGTCTCGAGATCAGCCGCGGGGTCAACCTCACCGACCGCCAGGTCAAGATCTGG
TTCCAGAACCGGCGGATGAAGCTCAAGAAGTTGAGCCGAGAGAGTCGCATCCGTGAGCTGTCTTCTGGATTCAGCTTTT
CCTAG

[3]*HoxX10* putative extended, partial sequence exon2
CACGCGAGCTCCACACGCAAGAAGCGCTGCCCCTACACCAAGCACCAGACGCTGGAGCTGGAGAAGGAGTTCCTGTTC
AGCATGTACCTCACGCGCGAGCGCCGACTCGAGATCAGCCACCTGCTCAGTCTCACCGACCGCCAGGTCAAGATCTGGT
TCCAGAATCGGCGCATGAAGCTCAAGAAGATGAAC CGAGGCCGTTGCAAGGAGTTTCTGTGA

[3]*HoxY11* putative extended, partial sequence exon2
CGGCCGCGTAAGAAGCGCTGCCCCTACACCAAATACCAGATCCGCGAGCTGGAGCGAGAGTTCTTCTTCAGCGTCTAC
ATCAACAAGGAGAAGCGCCTGCAGCTATCGCGCCTCCTCAACCTCACCGACCGACAGGTCAAGATCTGGTTCCAGAAC
CGACGAATGAAGGAGAAGAAGCTGAAC AGGGATCGCATGCAGTACTTCGCGCCTAACCCGCTCCTATGA

[3]*HoxZ11b* putative extended, partial sequence exon2
CGCGTGCGTAAGAAGCGCTGCCCCTACACCAAGTTCCAGATCCGCGAGCTGGAGCGAGAGTTCTTCTTCAACGTCTACA
TCAACAAGGAGAAGCGCCTGCAACTGTCGCGCCTCCTCAACCTCACCGACCGACAAGTCAAGATCTGGTTCCAGAACC
GGCGGATGAGCGAGAAGAAGCTAAGC CGGGACCGTCTGCAGTACTACGCAGGGAACCCATTTTTATAG

[3]*Hox11* putative, partial sequence exon2
CCCCGGAAGAAGCGCTGCCCCTACACCAAGTTCCAGACGCGTGAGCTCGAGCGAGAGTTCTTCTTCAGCGTCTACATCA
ACAAGGAGAAGCGCCTGCAGATCTCACGCCTCCTCAACCTCACCGACCGCCAGGTCAAGATCTGGTTCCAGAACAGGC
GCATGAAAGAGAAGAAGCTGAACAGGGATCGAATGCAGTACTTTGCGGCCAACCCGCTGCTGTGA

[3]*Hox13a* putative, partial sequence exon1
ATGAAGCGGCGCGGCGATAGGCACGCACGCGCTCGCGCGCGCGCCCGGCACCACGTGACCACCGGCACGACGCGTGC
ACAGGTGTACACGCGCGTGTCTGTGTGGCTTCTGTGTCCACGTGTGTACATCGGGTCGCCCAGGGTTCACGTTCGCCAC
TTCCCACCCTCCCTTCCACCTCCCTCCATCCCCGCACTTATACCTGTGGGACCCAGGTGTACGCGGGCGTGCGTACGCGA
ATGTGCACAGGTGTTCTCACGTGTATACCGGGGGCTGCTGTGTCCACACGTGCATAGCTTCCCCCCCTCCGCCGCCCCCT
CCCAGGTTCTGCGTAAGCGCTGTCCACGGTTTGCACGCGTGTACATAAGTTGTACACGCGCATTAGCACTCGTGTCTAC
ACGGGGGGAGTTGCTCCTCCGTCCGCTGGAAACGATGTTCCTGCATGGCTCTGGCGGTGGCGTTGGAGGTGGCGGCTGT
GGTGTTGGTGGCGTCCTCTGGGACGAGTCGAGGAGGCTCGGCGAGCACGATCAGGGATTTCCTGACCCCAGTCGCAAC
CCTTGGGCGTCCCCGTTCGCCGAGGCCCCTACTCCGAGCTCGCCAGGAAGGCGCTGCATAGCGAGTCACCTGTGCACCT
GCTATATCGCTCATCAAAGTCATCAAAATCATCAGCATCAACAGCAGGACGGAGCACTGAGCAGATGTGCGGATGTCC

AGGCGGTTCAACCACGCGGATACATGGGGTTTGATGTGGATGTTGTGTCCGAACAACAACAACAACAGCAGCAGCAGC
ATCATCCGCAGATCGGCACCACTCCTGCCCGACAGCTACCGCCCTGGGCTGTGTACGCGCCGGGTACCTGGCCGGCCCA
GGTGTGTGGTGGGTACAGGGAGCGCGCTCAGCCGTCCCACCTGTGGAAGGCCGCC

[3]*Hox13a* putative, partial sequence exon2
TTCCTCGCAGGGGAGGAAGGGTTCCCTGTGTTCGCGGGGGTCCAGCAGGAAGGGGGTCTCCCGTACCCCGCGGACCCC
GTGCCCCGGCGTTCCCGCAAGAGGCGGGTCCCCTACAGCAAGGCGCAGCTGCGGGAGCTCGAGGCGGAGTTCGGGGCG
AGTCGCTTCGTGAGCCGCGAGAGGCGGCGCGGCGTGGCGGCCAGTACCCAGCTCAACGAGAGGCAG

[3]*Hox13a* putative, partial sequence exon3
GTCACCATCTGGTTCCAGAACCGGCGCGTCAAGGAGAAGAAGATTGCCGTGCGACGAGTGGGGTCCGCTTCCGGGGTC
AAATCGAGTTCCCAGGGAATCGGGTCCCCATGA

[3]*Hox13b* putative, partial sequence exon2
GCACGCAGCAAACGCGTTCCCTACACGAAGCTCCAACTGAAGGAGCTCGAGAAGGAATTCGAGACGAGCAGATTCGTC
AGCAAGGACAAGCGAAGGGGGATCGCCGCCTCCAGTGGACTCAGCGACAGGCAGGTCACCATCTGGTTCCAGAACCG
ACGGGTCAAGGAGAAGAGGCTGCTGGCCAAGGGCAAAACGGCCCTGGCTCTGTTGTGA

[3]*Hox13c* putative, partial sequence exon2
CATCCCGCGGACCCGGCGCCGCGCCGCTCGCGCAAGAGGCGCGTGCCCTACAGCAAGGCGCAGCTGCGGTCTCTCGAG
CGCGAGTTCGCGGCCTGCAGGTTCGTGACGAGGGAGAGGCGGCGACGCGTGGCGGCCCTGGCCGACCTCACGGAGCGC
CAGGTGACCATTTGGTTCCAGAACCGGCGCGTCAAGGAGAAGAAGCTTCTCGGCAGGGGCTCTCCTGCTGCGCGCAAG
GCGGCGGCGGCGGCGGCTTCGTCGTAG

[3]*Hox13d* putative, partial sequence exon2
CCCAACAGTCTGGTTAAGGCTACACGGGGGATCTGTTTTTGTGTGTGCGTGTGTCTCCGTGCGTGTGTCTCCGTGCGTTG
CCCTCTCGCAGACGGCTCGCAACCTCCTCACGAGATGAGCCTCTTCCGCCGGGGCCGCAAGAAGAGAGTCCCCTACACC
AAGGTGCAGCTGAAGGAACTCGAGCGGGAGTACGCCGCCAACAAGTTCATCACCAAGGACAAGCGGAGGAAAATTTC
GGCCGCCATCAGCTTATCCGAACGCCAGGTCACCATCTGGTTCCAGAACCGCAGGGTCAAAGAAAAGAAGGTTGTGTC
CAAGATGAAAACGACTCACCTGTGA

[3]*Hox13e* putative, partial sequence exon1
CGCGTGGGTCACGGCATGGGCGGCGCCCACGTTAAGGCGGCGTGCGGATCCGCGGCGCAGACGGCCGCGAGCCTCGCC
TCGGGCTACGGGCCGCACTCGCTCGTCGACAAATACGTGCAGGACGTCTCCGTCGGACTACAACCGAGCAACGGTGTC
GGCTCATCCGTTGCTGCCGGAGGTGGCGGTGGTGGTGGTCTCGGTGGTGGTGGAGGCGGTATCGGTGGTAGCGGTGGT
ATCGGTGGAGGTGGCGGCGGTGTGGGTGATGATTACACGAGCGGCCGGGCAGCCAAAGACTTCGCCTTCTACCCCAGC
TACGGCAGCGCTTATCACCACCACCACCATCATCAGCACATGCCGGGCTACTTAGACATGCCCGTGGTGCCGCACGCGT
CGCTCGCCCCGCCGAGCGAAGGTCGCCACGACCCCATCCTTCACCACGGAATGGACGGATACCACCATCAGCCATGGG
CGCTGCCCAATGGATGGAACGGGCAAATGTACTGCGCCAAGGAGCAGAGTCAGCTGTCGCACCTGTGGAAGTCGCCCC
TTGCA

[3]*Hox14* putative, partial sequence exon1
ATGGGCCACTATGGCCACCACCATCACCATCACGCGGCGCCTTACTACAGTCCGCCCTACAGCAGCGGAGCGCTCAACC
ACGGTCTCGCCAACATGTCCTCGGCGCTCGGT

[3]*Hox14* putative, partial sequence exon2
GCACTGGGCGGCATCCCGGGCCACTGGCCGCAGGCGGCGCAGGGCACTTCGTGCGCCACGCGGCCCCGCAAGAAGCGC
GTCCCCTACAGCAAGCAGCAGATCTCCGAGCTGGAGAGAGCCTTCGACGAGAACCGATTCCTCACCCCGGAGCTGCGC
CTCAGCATCTCGCACCGGCTCAGCCTCACCGAGCGACAG

[3]*Hox14* putative, partial sequence exon3
GTCAAGATTTGGTTCCAGAATCAGAGGCAGAAGGAGAAGAAGCTAATGCGAAGGCAGCAGAGTGGAGCCGCGGCCAA
CGCCACCAACATGAACAACGGGAGCGGTGGCACGACCGGCTCGGCGACTCTCCCGTGA