HOX CLUSTER EVOLUTION IN THE HIGHLY DERIVED PIPEFISH & SEAHORSE FAMILY

by

ALLISON M. FUITEN

A DISSERTATION

Presented to the Department of Biology
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

March 2019

DISSERTATION APPROVAL PAGE

Student: Allison M. Fuiten

Title: *Hox* Cluster Evolution in the Highly Derived Pipefish & Seahorse Family

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Biology by:

Patrick Phillips	Chairperson
William Cresko	Advisor
John Postlethwait	Core Member
Matthew Streisfeld	Core Member
Kirstin Sterner	Institutional Representative

and

Janet Woodruff-Borden	Vice Provost and Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded March 2019

DISSERTATION ABSTRACT

Allison M. Fuiten

Doctor of Philosophy

Department of Biology

March 2019

Title: *Hox* Cluster Evolution in the Highly Derived Pipefish & Seahorse Family

A central question in evolutionary biology is how organisms evolve highly derived and novel morphologies. More specifically, what changes to conserved developmental genes lead to the evolution of divergent morphologies? Here, I investigate the genetic and genomic changes to the developmentally important *Hox* genes using comparative genomics, gene expression and gene editing approaches. *Hox* genes code for homeodomain transcription factors that are responsible for determining the body plan of an embryo along the anterior-posterior axis, and changes to these genes have paralleled the rise of morphological diversity in the vertebrate animals. I focus my studies in a group of fish that exhibit a striking departure from the typical fish body plan: the pipefish and seahorse family, Syngnathidae. The evolution of syngnathid fish involved major modifications to their vertebrate body plan, but the developmental genetic basis of those changes is largely unknown.

I describe the genomic organization of *Hox* clusters in a species of syngnathid pipefish—the Gulf pipefish (*Syngnathus scovelli*). I present an initial investigation on phenotypic consequences to the loss of *hox7* genes in teleost fish—a group of *Hox* genes that are missing in syngnathids—using of the CRISPR/Cas9 system to induce indels in all *hox7* genes (*hoxa7a*, *hoxb7a)* in the threespine stickleback (*Gasterosteus aculeatus*). In the second half of my thesis, I investigate noncoding changes in the syngnathid *Hox*

clusters. I use syngnathid representative species and compared their conserved noncoding sequences within the *Hox* clusters to other teleost fish, non-teleost fish, and non-fish vertebrates. I present a detailed study regarding the nature of the loss of one conserved non-coding element.

Results from this research indicate that the divergent syngnathid body plan is not due to rampant change in throughout *Hox* clusters. Also, these data do not argue for the absence of any role of genetic changes in *Hox* clusters. Instead, the findings presented here support the intermediate hypothesis that certain key changes to the *Hox* genes, microRNAs, and regulatory elements have probably contributed to their body plan developmental evolution in this unique family of fish.

This work includes published co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Allison M. Fuiten

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon
University of Kansas, Lawrence, Kansas
University of San Francisco, San Francisco, California

DEGREES AWARDED:

Doctor of Philosophy, Biology, 2019, University of Oregon
Master of Arts, Ecology and Evolutionary Biology, 2012, University of Kansas
Bachelor of Science, Biology, 2008, University of San Francisco

AREAS OF SPECIAL INTEREST:

Evolutionary Genetics and Genomics
Evolutionary Developmental Biology

PROFESSIONAL EXPERIENCE:

Graduate Employee, Department of Biology, University of Oregon, Eugene, September 2017 – June 2018

Graduate Teaching Fellow, Department of Biology, University of Oregon, Eugene, September 2012 – June 2014

Teaching Assistant, Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, January 2011 – May 2012

Curatorial Assistant, Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, August 2009 – December 2010

Research Assistant, University of San Francisco, San Francisco, August 2007 – March 2008

Teaching Assistant, University of San Francisco, San Francisco, August 2006 – May 2007

Laboratory Technician, Triple Point Biologics, Inc., Forest Grove, Oregon, Summer 2005 & 2006

GRANTS, AWARDS, AND HONORS:

Doctoral Dissertation Improvement Grant, National Science Foundation, 2017 – 2019

Developmental Biology Training Grant Fellow, National Institutes of Health, 2015 – 2016

UO Under-Represented Minority Fellow, University of Oregon, 2016 – 2017

Genetics Training Grant Fellow, National Institutes of Health, 2014 – 2015

Donald E. Wimber Fund Award, University of Oregon, 2016

GrEBES Travel Award, University of Oregon, 2014 & 2017

Summa Cum Laude, University of San Francisco, 2008

Valedictorian Finalist, University of San Francisco, 2008

Robert Bellarmine Award for Academic Excellence, University of San Francisco, 2008

USF Biology Semester Award for Academic Excellence, University of San Francisco, 2006

Tri Beta Biological Honor Society Student of the Year, University of San Francisco, 2006

Tri Beta Biological Honor Society Student of the Semester, University of San Francisco, 2005

University Scholar, University of San Francisco, 2004 – 2008

Dean's List, University of San Francisco, 2004 – 2008

PUBLICATIONS:

Dever, J. A., A. M. Fuiten, Ö. Konu, and J. A. Wilkinson. 2012. "Cryptic torrent frogs of Myanmar: an examination of the *Amolops marmoratus* species complex with the resurrection of *Amolops afghanus* and the identification of a new species." *Copeia* 2012 (1): 57–76.

Fuiten, A. M., L. J. Welton, A. C. Diesmos, A. J. Barley, B. Oberheide, M. V. Duya, E. L. B. Rico, and R. M. Brown. 2011. "A new species of stream frog (*Sanguirana*) from the mountains of Luzon Island, Phillipines" *Herpetologica* 67 (1): 89–103.

Siler, C. D., A. M. Fuiten, R. M. Jones, A. C. Alcala, and R. M. Brown. 2011. "Phylogeny-based species delimitation in Philippine slender skinks (Reptilia: Squamata: Scincidae) II: taxonomic revision of *Brachymeles samarensis* and description of five new species." *Herpetological Monographs* 25 (1): 76–112.

Small, C. M., S. Bassham, J. Catchen, A. Amores, A. M. Fuiten, R. S. Brown, A. G. Jones, and W. A. Cresko. 2016. "The genome of the Gulf pipefish enables understanding of evolutionary innovations." *Genome biology* 17 (1): 258.

Stankowski, S., M. A. Chase, A. M. Fuiten, P. L. Ralph, and M. A. Streisfeld. *In review*. "The tempo of linked selection: rapid emergence of a heterogeneous genomic landscape during a radiation of monkeyflowers." *PLOS Biology*.

Wiley, E. O., A. M. Fuiten, M. H. Doosey, B. K. Lohman, C. Merkes, and M. Azuma. 2015. "The caudal skeleton of the zebrafish, *Danio rerio*, from a phylogenetic perspective: a polyural interpretation of homologous structures." *Copeia* 103 (4):740–750.

ACKNOWLEDGMENTS

Many people have helped make this work possible. I would first like to thank my advisor Bill Cresko for his guidance, advice, and support throughout the entire graduate school process. I would also like to thank the other members of my committee: Patrick Phillips, John Postlethwait, Matthew Streisfeld, and Kirstin Sterner for their support and guidance over the years.

A special thanks to Susie Bassham, Clay Small, and Mark Currey who have been members of the Cresko lab for my entire stint in the Cresko lab and have acted as additional mentors and answered many questions over the years. I would also like to thank the following past and present members of the Cresko lab for their support over the years: Kristin Alligood, Thom Nelson, Julian Catchen, Jason Sydes, Ann Petersen, Emily Beck, Martin Stervander, Hannah Tavalire, Kate Ituarte, John Crandall, Andrew Nishida, and Stacey Wagner. Thank you to undergraduate volunteers Nicole S Zavoshy, Mikaeli Dirling, Nia Harper, Jade Kast, and Starla Chambrose.

I would also like to acknowledge the following past and present members of the Postlethwait Lab who have also provided guidance and support at various times over the years: Angel Amores, Thomas Desvignes, Pete Batzel, Yi-Lin Yan. Ingo Braasch, Braedan McCluskey, Trevor Enright. Adam Jones and his former lab members Laura Edelstein and Andrew Anderson helped make all my fieldwork needs possible, along with Heather Mason-Jones and especially Emily Rose and their undergraduate workers Julia Skowronski, Jessica Elson, Laura Bellato, Breeann Roberts, Emily Craft, and Alana Boyles.

I would like to thank the research community in the Institute of Ecology and Evolution and in the Department of Biology, in particular Christine O'Connor, Precious

De Verteuil, and Rudyard Borowczak. I am grateful to Sara Nash and Arlene Deyo for all their office and administrative support.

Finally, this work would not have been possible without multiple sources of generous support: grants from the National Science Foundation (DEB-1701854 to W. A. Cresko and A. M. Fuiten), fellowships to A. M. Fuiten from the National Institutes of Health for both the Developmental Biology Training Grant Fellow (T32 HD007348) and the Genetics Training Grant Fellow (T32 GM007413), and from University of Oregon UO Under-Represented Minority Fellow.

To my parents, who provided continuous love and support
throughout my *many* years of study.

TABLE OF CONTENTS

Chapter Page

Chapter Page

Chapter Page

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

EVOLUTION OF DEVELOPMENT OF HIGHLY DERIVED ORGANISMS

As far back as Aristotle and Pliny the Elder, people have been fascinated by the morphological diversity found in animals in the natural world. As can be seen from the pages of ancient texts or edges of Middle Age maps, fantastical animals (some even real) have held a special place in the human imagination. Many explanations have been presented over the ages for strange or extreme animal forms, some of them natural, but most supernatural. With the publication of *On the Origin of Species* by Darwin, the genesis of diversity in organismal form and function finally had a modern evolutionary explanation (Darwin 1859, Elder 2012, Aristotle 2014).

Highly derived and novel characters litter the history of animal evolution and the appearance of these characters play an important role in the expansion of animal forms. For example, the notochord, jaws, and neural crest cells are some of the novel characters that contributed to the success of our vertebrate lineage. Other familiar examples of novel morphologies within animals range from bird feathers to turtle shells, from panda thumbs to insect wings.

In the nineteenth century, scientists began to recognize the important role development plays in the evolution of morphologies. Of note, von Baer and Haeckel used comparative embryology to present the idea of that development of form reflects evolutionary descent (Haeckel 1866, 1896, von Baer 1828). Yet, in order to understand the origin of derived and novel characters, the genetic control of development must be examined. In the twentieth century, the modern synthesis incorporated genetics into evolutionary theory (Huxley 1942).

At the advent of molecular biology in the 1950s, scientists expected to find that the genetic content of different species would be abundant and highly different among divergent organisms given the great degree of morphological variation present within animals. Despite expectations by these early evolutionary biologists, it is now known that many developmental genetic pathways have remained surprisingly conserved across the different animal lineages over the course of metazoan evolution in terms of both sequence and function (Duboule and Dollé 1989, Carroll, Grenier, and Weatherbee 2013, McGinnis et al. 1984, Graham, Papalopulu, and Krumlauf 1989, Quiring et al. 1994, King and Wilson 1975).

Starting with Kimura's neutral theory of molecular evolution, and continuing until today, scientists are still parsing out how a genome's noncoding and coding content are differently affected by various evolutionary pressures and have different degrees of conservation and rates of change (Kimura 1968, Kern and Hahn 2018). Under the context of development, starting with the discovery of *Hox* genes in the 1980s and extending to numerous different developmental regulators and cell signaling molecules, entire gene families were found to be preserved over very great evolutionary differences at the sequence level and—amazingly—functional level in some cases (Duboule and Dollé 1989, Carroll, Grenier, and Weatherbee 2013).

Nevertheless, animals do vary phenotypically, and sometimes in radical ways. So where in the conserved developmental genetic pathways does this genetic diversity reside? King and Wilson (1975) compared for the first time a large set of proteins between human and chimpanzees, producing one of several papers to first propose that evolutionary changes can be more often attributed to the change in gene expression rather than the changes of the protein sequences (King and Wilson 1975, Zuckerkandl and

Pauling 1965, Britten and Davidson 1971). More recent studies have shown connections between changes in developmental gene expression and the evolution in derived morphological features (reviewed by (Carroll 2008)).

Another source of genetic diversity within conserved developmental pathways can be through gene duplication. Mechanisms that promote the retention of copied genes include neofunctionalization (Ohno 1970, Sidow 1996) and subfunctionalization (Force et al. 1999). Extra copies of genes allow for emergence of new gene functions within conserved developmental pathways and therefore, allow for novel morphological evolution. It has been found in past studies that there is a bias towards transcriptional and developmental genes being retained in duplicate after all genome duplications in plants, vertebrates, fish and yeast (reviewed in (Van de Peer, Maere, and Meyer 2009); also see (Putnam et al. 2008, Maere et al. 2005, Seoighe and Wolfe 1999, Seoighe and Gehring 2004, Blanc and Wolfe 2004, Blomme et al. 2006, Brunet et al. 2006, Davis and Petrov 2004)).

HOX GENES AND MORPHOLOGICAL EVOLUTION

Hox genes are prime examples of core developmental genes that have maintained a great level of conservation throughout the animal kingdom despite the large amount of body plan diversity found in animals (reviewed in (Gehring, Affolter, and Bürglin 1994, Burglin and Affolter 2016, Holland 2013)). *Hox* genes code for homeodomain transcription factors that are responsible for determining the body plan of an embryo along the anterior-posterior axis. They are made up of two exons that contain a homeobox DNA sequence. Although there is some divergence in sequence and content,

the core functionality of *Hox* genes in A-P axis determination has been conserved across nearly all animals examined to date.

This level of conservation in *Hox* genes, and in other core developmental gene families, has been hypothesized to occur because major changes will be detrimental to the development of the organism because of antagonistic pleiotropy and therefore will be removed by selection (Carroll 2008, Hoekstra and Coyne 2007). Alternatively, slight shifts in gene copy number and gene regulation of these conserved developmental genes may create traits that can evolve adaptively because they are producing different morphologies while still working within developmental constraint (Wilkins 2002, Raff 2012). This question remains central to the field of evolution of development.

The ancestral set of *Hox* genes consisted of a single cluster of genes, resulting from tandem duplications of an ancestral proto-*Hox* gene (Garcia-Fernandez 2005). Invertebrates, for the most part, still maintain just a single *Hox* complex. Due to subsequent rounds of whole genome duplications, vertebrates have duplicate copies of the *Hox* complex (Pascual-Anaya et al. 2013) (Figure 1.1). In vertebrates, tetrapods have four *Hox* gene clusters (denoted as *Hox* clusters A, B, C, and D), while teleost fish have eight clusters of *Hox* genes due to the whole teleost genome duplication (*Hox* clusters Aa, Ab, Ba, Bb, Ca, Cb, Da, Db) (Amores et al. 1998). The majority of teleost fish have lost their *HoxCb* cluster, while a smaller subset has lost their *HoxDb* cluster.

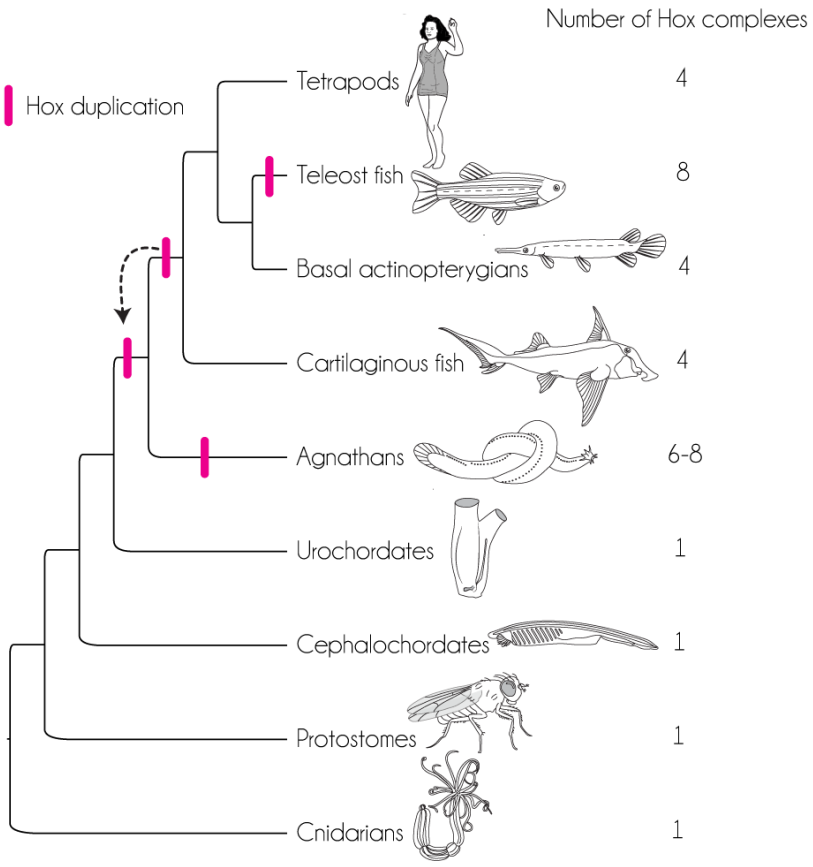


Figure 1.1: Evolution of *Hox* complex. Evolutionary timing of *Hox* complex duplications are denoted on the animal phylogeny based on (Carroll, Grenier, and Weatherbee 2013), with updates from (Ravi et al. 2009, Pascual-Anaya et al. 2018). Dashed arrow indicates current uncertainty where the second vertebrate *Hox* cluster duplication occurred relative to agnathans.

In vertebrates, *Hox* genes are organized into 13 paralogous groups that are arranged into these multiple gene clusters (Scott 1992). Often, *evenskipped* (*evx*) genes are included as a member of the *Hox* clusters, as they are closely related homeodomain transcription factors found immediately upstream of the *hox13* genes. More recently, microRNAs have been annotated within the *Hox* clusters. These microRNAs—a class of noncoding RNA gene—serve as important post transcriptional regulators for expression of surrounding *Hox* genes. The *mir196* microRNAs are located between certain *hox10*

and *hox9* genes and *mir10* microRNAs are located between certain *hox5* and *hox6* genes (Tanzer et al. 2005) (Figure 1.2).

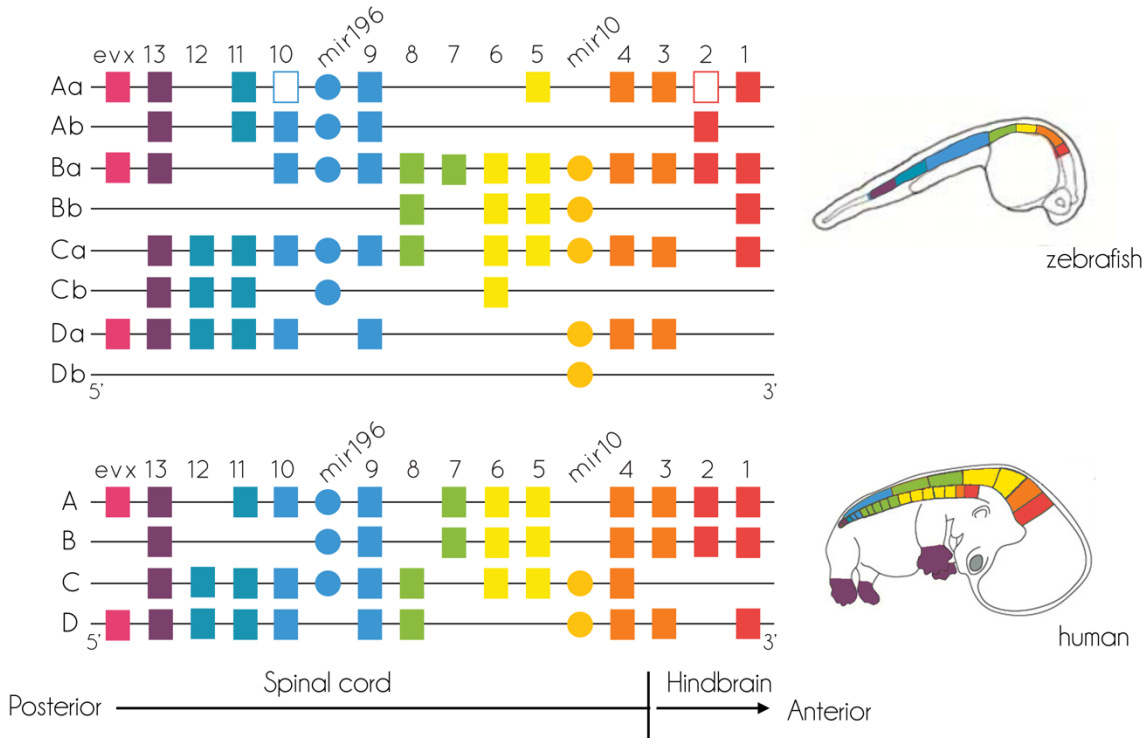


Figure 1.2: *Hox* clusters are important in body plan development. A cartoon of the *Hox* clusters in a representative tetrapod (human) and a representative teleost fish (zebrafish) with boxes representing genes and circles representing microRNAs arranged along chromosome segments oriented left to right 5' to 3'. Colors of the genes correspond to where they are expressed along the A-P axis during development as indicated with the matching colors on the cartoons. Human embryo and zebrafish embryo cartoons respectively modified from (Goodman 2003, Swalla 2006).

Early studies looking at expression patterns of these genes noted that *Hox* genes in the same paralogous groups have overlapping expression along the axis. These studies also show that *Hox* genes exhibit collinearity. This means that the order they appear in the genome reflects the order they are expressed along the anterior-posterior body axis (Gaunt, Sharpe, and Duboule 1988, Graham, Papalopulu, and Krumlauf 1989, Peterson et

al. 1994, Duboule and Dollé 1989, Dekker et al. 1993, Godsave et al. 1994), with the *hox3* to *hox11* genes expressed along the axial skeleton and the *hox1* to *hox2* genes expressed in the hindbrain during development (reviewed in (Wellik 2009)) (Figure 1.2). Later experiments using gain-of- and loss-of-function experiments further demonstrated that *Hox* genes in the same paralogous groups have redundant functions—where knocking out all members of a single paralog groups would confer a stronger phenotype that knocking out a single member of a paralogous group (reviewed in (Wellik 2009)).

TELEOST FISH AS MODELS

Teleost fish make ideal models for studying whether variation in *Hox* genes contributes to morphological evolutions for several reasons. In general, teleost fish are recognized as important models for vertebrate evo-devo (evolutionary development biology) in the genomics era (Braasch et al. 2015). Overall, this class of fish make up around 40% of all vertebrate diversity with over 27,000 described species (Hoegg et al. 2007, Nelson 2006). Because of their great diversity, scientists have used the teleost treasure trove of adaptive phenotypes like blindness in cavefish or lack of hemoglobin in Antarctic icefish to study aspects of human diseases and disorders (Albertson et al. 2009).

This species richness of teleost fish has been correlated to the teleost specific whole genome duplication (Amores et al. 1998, Van de Peer, Maere, and Meyer 2009). Important insights into the timing of when certain genes—including *Hox* genes—evolved, along with identifying ancestral gene functions and subsequent gene subfunction partitioning can be studied because of the whole teleost genome duplication (Postlethwait et al. 2004, Force et al. 1999, Amores et al. 2004, Amores et al. 1998). This type of study has been greatly aided by the availability of a basal non-teleost fish that bridges the

comparative genomics gap between the duplicated teleost genome with other vertebrate genomes (Braasch et al. 2016, Amores et al. 2011). Therefore, the evolutionary history of the various *Hox* genes among the different lineages can be traced.

Additionally, because of the teleost whole genome duplication, fish have more copies and combinations of *Hox* genes and microRNAs than tetrapods. This makes teleost fish a robust comparative, evolutionary framework to study the significance each of the *Hox* genes play in morphological evolution (Amores et al. 2004, Hoegg et al. 2007). Examples of these studies include reporting the altered expression of *hoxd9a* corresponding to the loss of pelvic fins in pufferfish and the regulation of axial development in zebrafish by *Hox* microRNA, *mir196* (Tanaka et al. 2005, He et al. 2011). The genomes of the dwarf cyprinids from the genus *Paedocypris* have a reduced complement of *Hox* genes potentially tied to the evolution of their reduced skeletons (Malmstrom et al. 2018). Alternatively, the genome of the sunfish (*Mola mola*) have retained more *Hox* genes that would be predicted based on their phylogenetic relatedness to pufferfish and their reduced body plan (Pan et al. 2016).

SYNGNATHID FISH AS MODELS FOR MORPHOLOGICAL EVOLUTION

While the macroevolutionary studies across great phylogenetic distances in fish has been useful, it has been difficult to causatively tie changes in *Hox* genes to differences in morphology. What is needed is the ability to study *Hox* gene content, expression and function among closely related species within a single family that contains a great array of morphological diversity. One example of a teleost family that exhibits a great amount of derived and novel fish morphology is the Syngnathidae family. Studying *Hox* genes in this particular teleost family provides a unique opportunity to

explore the ways that conserved genetic pathways can be altered and how genetic changes can lead to the evolution of highly derived traits.

Syngnathids made their formal debut in evolutionary biology with their first description in *Systema Naturae* (1758), where Linnaeus described several species of pipefish and one species of seahorse—*Hippocampus hippocampus*. The family name is derived from Greek words meaning fused jaws (*syn* = together/fused, *gnathos* = jaws). Syngnathidae currently consists of 319 described species of pipefish, pipehorses, seahorses and seadragons organized into 57 genera. This includes three species of seadragon, 45 species of seahorse, 21 species pipehorses and the rest considered pipefish (Froese and Pauly 2018 , Fricke, Eschmeyer, and van der Laan 2019, Neutens et al. 2014). Pipehorses fall morphologically between seahorses and pipefish because they lack the vertical body posture of seahorses but have prehensile tails and lack caudal fins. While seadragons and seahorses reflect monophyletic clades within Syngnathidae, pipehorses are scattered across multiple clades of pipefish (Neutens et al. 2014). The family is divided into two subfamilies—Nerophinae and Syngnathinae. Males in the Nerophinae subfamily carry their eggs on the ventral side of their trunks, while males in the Syngnathinae subfamily carry their eggs under their tails. These subfamilies are also supported by molecular phylogenetics, where Nerophinae and Syngnathinae are two monophyletic sister clades that consist of 56 species and 263 species, respectively (Fricke, Eschmeyer, and van der Laan 2019, Hamilton et al. 2017) (Figure 1.3).

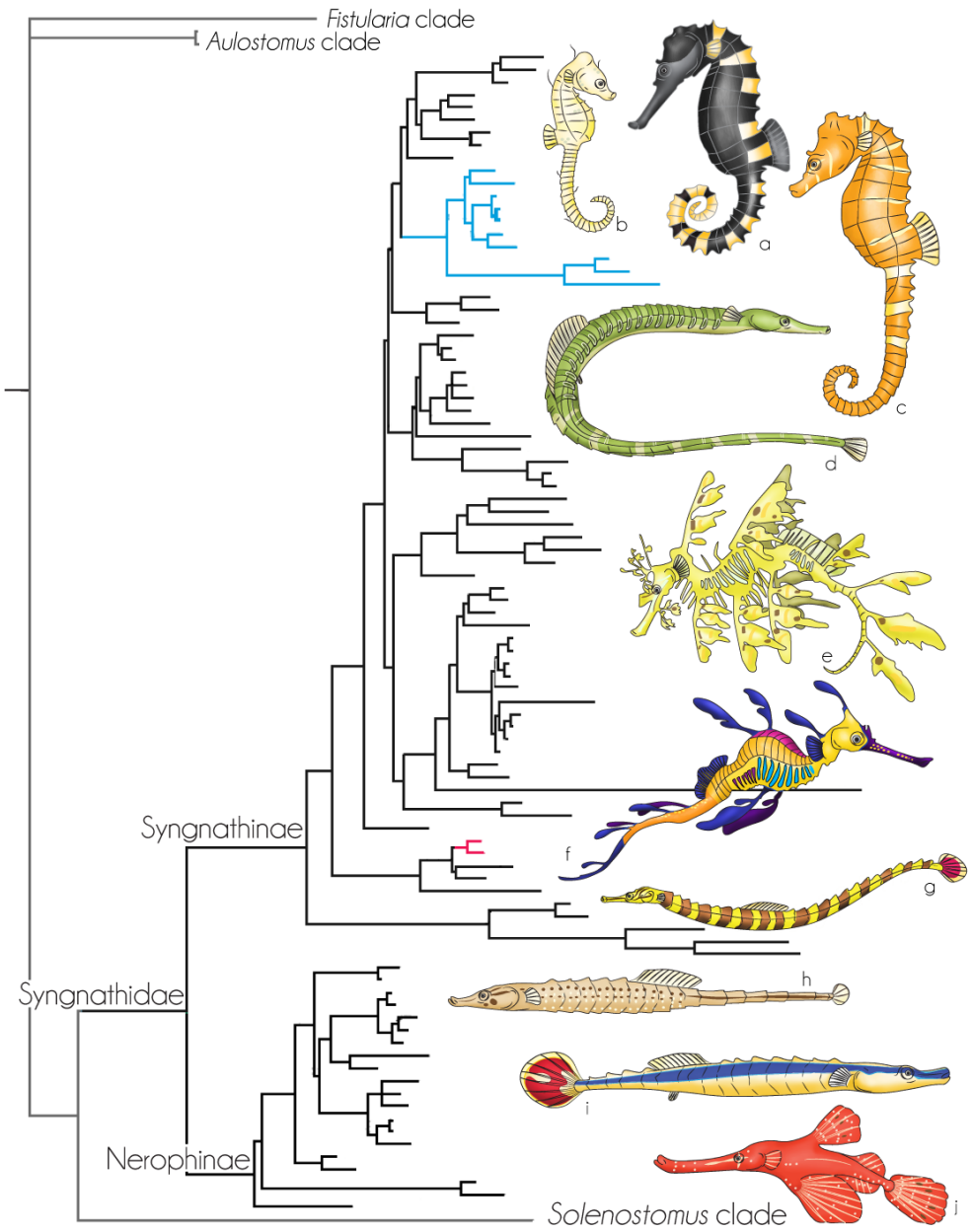


Figure 1.3: The Syngnathidae family contain morphologically diverse fish encompassing pipefish, seahorses, seadragons and pipehorses. Illustrations depict representative species: (a) *Hippocampus zostrae* (b) *H. comes* (c) *H. erectus* (d) *Syngnathus scovelli* (e) *Phycodurus eques* (f) *Phyllopteryx taeniolatus* (g) *Corythoichthys haematopterus* (h) *Choeroichthys sculptus* (i) *Doryrhamphus excisus* (j) *Solenostomus cyanopterus.* Syngnathidae is divided into two subfamilies—the tail brooding Syngnathinae and the trunk brooding Nerophinae. Seadragon clade highlighted in pink, seahorse clade in blue, with black indicating pipefish and pipehorses. Cladogram based on molecular phylogeny published by Hamilton et al. 2017.

The immediate outgroups to Syngnathidae are several less speciose families and are united with Syngnathidae under the broader order of Syngnathiformes. These fish also exhibit a certain degree of elongated and unusual morphologies and are comprised of the ghost pipefish (Solenostomidae), shrimpfish (Centriscidae), trumpetfish (Aulostomidae), and cornetfish (Fistulariidae). The most closely related fish that exhibit more standard teleost fish body plans is a monophyletic clade that contains seamoths (Pegasidae), goatfish (Mullidae), flying gurnards (Dactylopteridae), and dragonets (Callionymidae) (Longo et al. 2017).

Syngnathid fish have a worldwide distribution in both temperate and tropical waters. They are mostly found in shallow marine water but can also be found in fresh and brackish water. Their habits can range from seagrass beds and mangrove forests to reefs to estuaries and rivers to sandy and silty bottom habitats (Allen et al. 2006 , Howard and Koehn 1985, Pollard 1984, Whitfield 1999, York et al. 2006).

So unusual is their body plan that syngnathid fish were once thought of as marine insects and were even categorized as amphibians in one edition of *Systema Naturae* (1766). Syngnathid fishes are known for their highly divergent body plans, including the elongate form of many pipefishes and seadragons and the vertical body axis and reduced craniovertebral angle of seahorses (Herald 1959, Teske and Beheregaray 2009, Wilson and Rouse 2010). This elongated body plan can partly be explained by an increase in number of vertebrae. The syngnathid fish lineage has undergone an expansion of the vertebral column with the total number of vertebrae ranging from 31 to 94 depending on the lineage (Hoffman, Mobley, and Jones 2006).

Derived characters such as leafy appendages, prehensile tails, bony body armor, male somatic brooding and loss of ribs, caudal, and pelvic fins are common across the

family and in many cases have evolved independently in multiple lineages (Neutens et al. 2014, Herald 1959, Wilson and Rouse 2010). Examples of novel traits from this family are the reproductive tissue found in the brood pouch of male syngnathids, and the prehensile ability of the tail of the seahorse (Small, Harlin-Cognato, and Jones 2013, Neutens et al. 2014). Both of these novel characters are tied to the elongated body of the pipefish which provides room for the brood pouch, and allows for the specialized flexing and bending necessary for the prehensile grasping of the tail in seahorses and pipehorses (Neutens et al. 2014, Bruner and Bartolino 2008). The position of where males carry their embryos is thought to be a selective pressure that results in a shift in relative proportion of tail and trunk vertebrae (Hoffman, Mobley, and Jones 2006).

In addition, syngnathids have a highly modified cranium that is the result of an elongation in a series of bones in the ethmoid craniofacial region (Leysen et al. 2010). Their unique cranial elements (that overall give syngnathid fish their signature equine look) are highly adapted for suction feeding—making them the fastest recorded suction feeders among teleost fish (Van Wassenbergh et al. 2011, Van Wassenbergh, Roos, and Ferry 2011). These skeletal changes happen early in development (Brown 2010) (Figure 1.4).

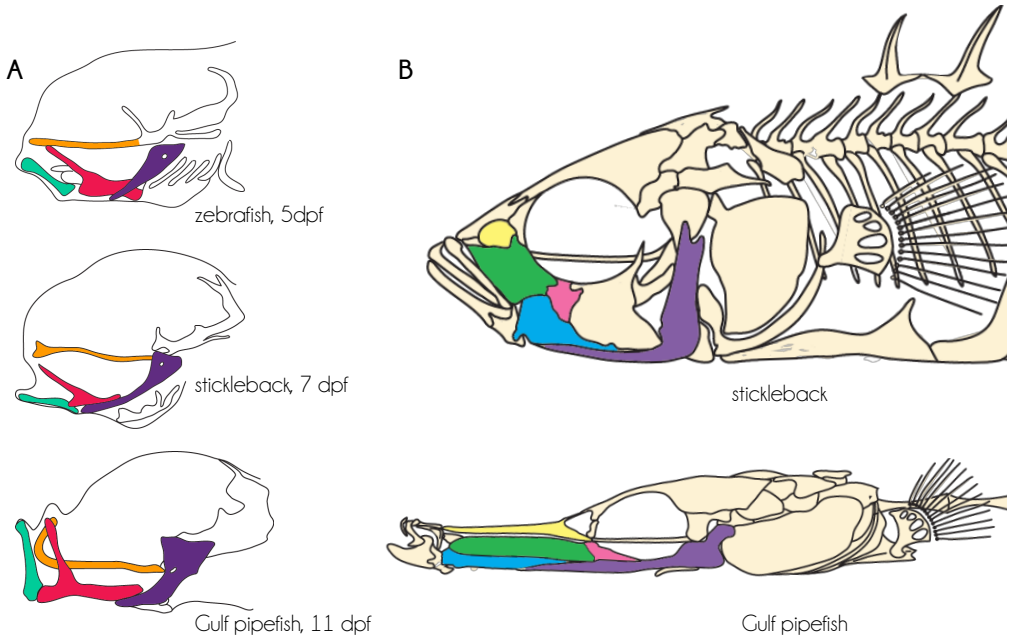


Figure 1.4: Modifications to the syngnathid skull happens early in development. a) illustrations highlighting homologous bones in the developing skull between Gulf pipefish, threespine stickleback and zebrafish: hyosymplectic (purple), Meckel's cartilage (teal), palatoquadrate (pink), ethmoid plate (orange). Drawing of zebrafish modified from (Schilling and Kimmel 1997). b) Homologous bones are highlighted in the adult skull of the threespine stickleback and Gulf pipefish in the elongated ethmoid region of the cranium.

In total, these remarkable characters make syngnathids an exceptional clade for the study of evolutionary novelty. Connections between the highly divergent body plan seen in this family of fish and modification to the *Hox* gene fish has remained an open question for curious biologists since many of these modifications happen early in develop at the time *Hox* genes are at work.

DISSERTATION OUTLINE

Previous studies have explored the functional role and adaptive significance of these unusual syngnathid traits, but their genetic basis remains unclear (Neutens et al. 2014, Porter et al. 2015, Flammang et al. 2009, Leysen et al. 2011, Van Wassenbergh et al. 2011, Van Wassenbergh, Roos, and Ferry 2011). Unlike previous studies, my

proposed research addresses the identification of the genetic changes that are responsible for the evolution of unique syngnathid morphology.

My dissertation work aims to determine changes to core genetic pathways that contribute to the evolution of highly derived morphologies. More specifically, I am using comparative genomics, gene editing, and gene expression approaches to investigate the coding and noncoding genetic changes to the developmentally important *Hox* genes and studying how these changes might contribute to the divergent body axis of syngnathid fish.

In Chapter II, I include the Gulf pipefish genome publication (Small et al. 2016). I was a co-author on this large, collaborative research paper. Production of a reference genome from this family of Syngnathidae was necessary for my proposed dissertation research. Therefore, I ended up significantly contributing to the production of the Gulf pipefish genome and its publication. The Gulf pipefish (*Syngnathus scovelli*) are a great representative of this group because this species has been the subject of recent evolutionary genetic and behavioral studies, can be kept in a lab for experimental studies, and it has many of the derived traits that define the family (Hoffman, Mobley, and Jones 2006, Jones, Walker, and Avise 2001, Paczolt and Jones 2010, Flanagan et al. 2014). Furthermore, a subset of my *Hox* gene dissertation research—restricted to only presenting the coding genes and microRNA contents of the *Hox* cluster—are included in this chapter. It was the first time that the *Hox* clusters were described from a member of the Syngnathidae family. Notably, shortly after this research was published, two more Syngnathid genomes were also published for the tiger tail seahorse (*Hippocampus comes*) and the lined seahorse (*H. erectus*) along with their *Hox* content (Lin et al. 2016, Lin et al. 2017). As part of my work in this publication, I assess the phylogenetic placement of

syngnathid fish relative to other representative fish taxa using ultraconserved elements and I compare the *Hox* cluster gene content of the Gulf pipefish against other teleost fish species. Given their phylogenetic position, I find that the *Hox* gene content has remained largely conserved relative to other teleost fish with annotated *Hox* clusters. Nevertheless, some interesting losses include the convergent losses of *hox7* genes and *mir196b*, and the unique loss of *eve1*.

In Chapter III, in an attempt to determine possible effects on the evolution of the syngnathid body plan of the loss of *hox7* genes, I describe the creation of mutations in the orthologous genes in threespine stickleback fish (*Gasterosteus aculeatus*). In this chapter, I discuss the experimental design of using the CRISPR/Cas9 system to induce indels in all *hox7* genes (*hoxa7a*, *hoxb7a)* in stickleback and the successful establishment of transgenic lines for the *hox7* gene knockouts. I also describe some initial results that indicate the possible role for *hox7* genes in rib and vertebrae development.

Both Chapters II and III focus on exploring the *Hox* gene content and the phenotypic impact of the evolutionary loss of some of these *Hox* genes. For Chapter IV, I examine the conserved noncoding elements within the boundaries of the syngnathid *Hox* clusters. I use the lined seahorse, tiger tail seahorse and the Gulf pipefish (*Hippocampus erectus, H. comes* and *Syngnathus scovelli*, respectively) as the syngnathid representatives and compared their CNE content to percomorph teleost fish (*Gasterosteus aculeatus, Takifugu rubripes*, *Oryzias latipes*, *Thunnus orientalis*), non-percomorph teleost fish (*Boleophthalmus pectinirostris*, *Gadus morhua*, *Danio rerio*), non-teleost fish (*Lepisosteus oculatus)*, and two non-fish vertebrates (*Mus musculus* and *Homo sapiens*). I catalog 718 CNEs, of which 388 elements are specific to the Gulf pipefish, tiger tail and lined seahorse genomes. I find five instances of syngnathid specific losses

of CNE among the species examined. This includes two independent losses of *Hox* cluster microRNAs—*mir19b* and *mir10a*—and three unique CNE losses only found among the syngnathid species. In two of these three losses, it is unknown whether these CNEs serve a functional role or are merely the result of neutral sequence conservation. The third unique loss is located in the intron of *hoxa2b* in the *HoxAb* cluster. It is highly conserved in that it is present in all other species examined. This element is a known enhancer element for *hoxa2b* and is scrutinized in greater detail in the next chapter of this thesis.

In Chapter V, I further research the surprising loss of the *hoxa2b* enhancer element. For this study, I expand my syngnathid sampling to include two species of the Nerophinae subfamily—*Doryrhamphus excisus* and *Choeroichthys sculptus* and five species from the Syngnathinae subfamily—*Corythoichthys haematopterus*, *Syngnathus scovelli*, *Hippocampus erectus*, *H. comes*, and *H. zostrae*. I also included *Solenostomus cyanopterus*—the robust ghost pipefish. The *Solenostomus* genus is the immediate outgroup to Syngnathidae. I find that the Pbx/Hox binding element sequence motifs and spacing between the binding elements have been modified for this enhancer. One Prep/Meis binding motif has been lost in Syngnathidae. Subsequently, I show expression of this gene in rhombomere 4 of the hindbrain is lower relative to the surrounding rhombomeres in the Gulf pipefish and this change in expression is consistent with it causing effects on the cranial neural crest. Ghost pipefish, the immediate outgroup to the teleost family Syngnathidae, has all the expected binding sites for this enhancer element, which means that the total loss of the Prep/Meis binding site must have occurred after ghost pipefish split from Syngnathidae.

In Chapter VI, I summarize the results from Chapters II, III, IV, and V and discuss how they contribute to our understanding of the genetic, genomic, and developmental changes involved in the evolution of the modified morphology in a syngnathid pipefish lineage.

CHAPTER II

THE GENOME OF THE GULF PIPEFISH ENABLES UNDERSTANDING OF EVOLUTIONARY INNOVATIONS

This chapter was published in volume 17 of the journal *Genome Biology* in December 2016. Clay Small, Susan Bassham, Julian Catchen, Angel Amores, Robin Brown, Adam Jones, and William Cresko are co-authors on this publication. Production of a reference genome from this family of Syngnathidae was crucial for my proposed dissertation research. Therefore, I ended up significantly contributing to the production of the Gulf pipefish genome and its publication. Furthermore, a subset of my *Hox* gene dissertation research—restricted to only presenting the coding genes and microRNA contents of the *Hox* cluster—are included in the genome paper.

This was a large, joint collaborative project on which I was a key team member. My personal contributions to this paper included performing the genome assembly with co-authors C. Small and J. Catchen. In addition, I contributed the whole genome annotation, the teleost genome assembly statistics comparison (Table 2.1), the phylogenomic analysis (Figure 2.3), the *Hox* gene cluster description and analysis (Figure 2.4), and the conserved non-coding element analysis for the *Dlx* gene clusters (Figure 2.5). I also created Figure 2.1.

The genetic map was produced by A. Amores, the chromosome evolution analysis was performed by J. Catchen, the *tbx4-pitx1* pathway analysis was performed by S. Bassham, the brood pouch gene expression and patristacin duplication analysis was performed by C. Small. C. Small, S. Bassham, J. Catchen, and R. Brown produced sequencing libraries. J. Catchen wrote custom software. R. Brown performed

morphological analysis of embryos. A. Jones and W. Cresko were the principal investigators for this work.

Because this was such a large, collaborative genome project, numerous authors contributed significant amounts of work. C. Small, S. Bassham, and J. Catchen were appointed as the main authors of this manuscript. Nevertheless, I contributed to writing the results for The pipefish genome assembly is of high quality and completeness, Phylogenomic analysis supports an alternative hypothesis for the position of syngnathiform fishes among the Percomorphs, Convergent and unique gene losses have occurred in the pipefish *Hox* clusters, the methods for Genome sequencing libraries and genome sequence assembly, Genome annotation, Conserved synteny analysis, Phylogenomic analysis using ultraconserved elements, Characterization of *Hox* clusters *Hox* gene content, and Characterization of *dlx* CNEs. The full supplementary material for this publication can be found under the Additional Files section at https://doi.org/10.1186/s13059-016-1126-6.

The citation for this publication is as follows:

Small, C. M., S. Bassham, J. Catchen, A. Amores, A. M. Fuiten, R. S. Brown, A. G. Jones, and W. A. Cresko. "The genome of the Gulf pipefish enables understanding of evolutionary innovations." *Genome biology* 17, no. 1 (2016): 258.

INTRODUCTION

Evolutionary novelties adorn the tree of life, and yet their genetic origins remain a problem for biologists. The Modern Synthesis sparsely addressed novel traits but rationalized their incidence with neo-Darwinian models of gradual change via accumulation of many small-effect mutations (Mayr 1960). Contemporary perspectives are more accepting of discontinuous morphological change (Muller and Wagner 1991), underlain by genetic changes diverse in nature. These changes may include point mutations as well as gross changes like gains and losses of genes or their regulatory elements, but the common thread is their effect on developmental systems. Indeed, the origin of novelties is now routinely viewed through the lens of evolutionary developmental biology, with an emphasis on how gene regulatory networks arise *de novo* or are modified from ancient ones (Shubin, Tabin, and Carroll 2009) to orchestrate novel gene expression in development (Wagner and Lynch 2010).

This modern genetic and developmental understanding of novel traits is an extremely difficult objective without quality genomic resources. Past genome sequencing efforts have been the purview of large, well-populated research communities generally focused on producing a resource beneficial for biomedical research. In the midst of the current sequencing technology revolution, however, the door is open for small research groups to produce genome resources for a variety of other questions, including those in ecology, conservation biology, evolutionary biology, and population genomics. As new evolutionary lineages are sampled, a valuable by-product is that novel reference genomes can augment the study of other existing model genomes, in the way the spotted gar (*Lepisosteus oculatus*) genome aids in bridging between the tetrapod and teleost model organisms (Braasch et al. 2016). We set out to genomically enable the study of novel

body plan and reproductive character evolution in syngnathid fishes (pipefishes, seahorses, and seadragons) by generating a high-quality reference genome for the Gulf pipefish, *Syngnathus scovelli*.

Syngnathid fishes are widely recognized for their highly divergent body plans (Herald 1959, Teske and Beheregaray 2009, Wilson and Rouse 2010), including the elongate form of many pipefishes (Figure 2.1), the upright body axis and reduced craniovertebral angle of seahorses, and the highly cryptic morphology of the seadragons. Derived characters such as leafy appendages, prehensile tails, and bony body armor are common across the family and, in many cases, have evolved independently in multiple lineages (Herald 1959, Wilson and Rouse 2010, Neutens et al. 2014). A truly striking evolutionary innovation shared by all syngnathid fishes is the somatic brooding of offspring by males, crowned by those lineages that have evolved complex, pouch-like structures for the maintenance of homeostasis during pregnancy (Carcupino 2002, Wilson et al. 2003, Ripley 2009, Ripley and Foran 2009). In total, these remarkable characters make syngnathids an exceptional clade for the study of evolutionary novelty. The Gulf pipefish represents the group well, given its recent history as a choice subject for evolutionary genetic and behavioral studies (Jones, Walker, and Avise 2001, Hoffman, Mobley, and Jones 2006, Paczolt and Jones 2010, Flanagan et al. 2014), its abundance and amenability to experimental work, and its embodiment of many of the derived syngnathid traits.

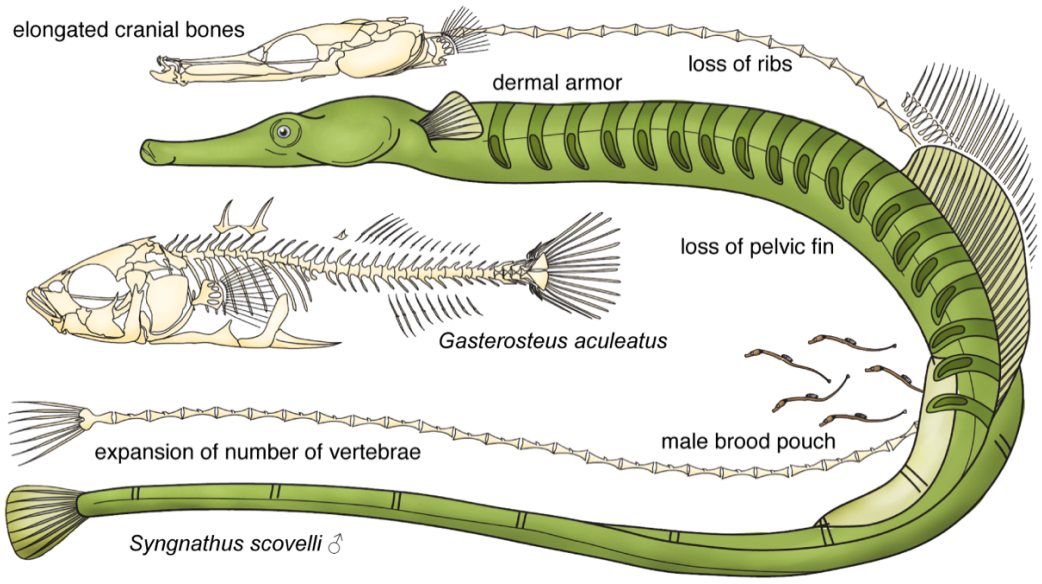


Figure 2.1: A cartoon representation of key derived traits in pipefishes and their relatives. Syngnathid fishes such as the Gulf pipefish have increased numbers of vertebrae and an elongated head, are missing pelvic fins and ribs, and have an evolutionarily novel structure, the male brood pouch. Shown for comparison is the axial skeleton of a percomorph with more typical morphology, a threespine stickleback. Note that not all derived syngnathid skeletal features are depicted in this cartoon. For detailed, anatomical illustrations of syngnathid skeleton attributes, please see other studies (Leysen et al. 2011, Leysen et al. 2010).

Comparative genomics and evolutionary developmental approaches to effectively study the evolution of new forms, such as the diversification of the syngnathid body plan, or the origin of male pregnancy, require advanced genomic tools. The centerpiece of each toolkit is a properly assembled, well annotated genome model, which can be directly compared at the sequence and structural levels to other species, and efficiently mined to design molecular tools for manipulative genetic studies. To this end we produced an annotated chromosome-level genome model (Braasch et al. 2016) for *S. scovelli* by integrating a 176X-coverage, short-read genome assembly with a linkage map constructed from RAD-seq markers. We used this tool to reveal features of chromosome structure evolution, to investigate pipefish lineage-specific losses of genes associated with morphological development, to infer the likely phylogenetic position of the

syngnathids in the tree of ray-finned fishes, and to describe a unique cluster of tandemly duplicated *patristacins* (Harlin-Cognato, Hoffman, and Jones 2006) that demonstrate conspicuous expression changes in the brood pouch during male pregnancy. Others have reviewed the approaches best suited to small-scale genome projects (Ekblom and Wolf 2014), but our intention here is to provide a biological case study and methodological template for success, motivated by the desire to better understand how novelties arise. We expect our experiences to be of interest to similarly sized research groups ready to reap the benefits of a reference genome in their own pursuits of biological discovery.

METHODS

Genome sequencing libraries and genome sequence assembly

We isolated genomic DNA from a single adult male pipefish purchased from Gulf Specimen Marine Laboratories, Inc. (Panacea, FL, U.S.A.) in 2010 using standard organic extraction. We generated four different 100 nt paired-end Illumina libraries for whole genome shotgun assembly: 1. a short (~180 bp) insert length library, 2. a 2.5-5 kb insert length jumping library, 3. a 5-10 kb insert length jumping library, and 4. a 11-15 kb insert length jumping library. To construct the 180 bp library we sheared 1 µg of genomic DNA to less than 500 bp using sonication in a Bioruptor (Diagenode), and size selected fragments by agarose gel electrophoresis, followed by end repair of the fragments, addition of adenosine overhangs, ligation of Illumina sequencing adapters, and 12 cycles of PCR amplification with Phusion polymerase (NEB). We used the Illumina Nextera Matepair Sample Preparation Kit (Illumina, cat. #FC-132-1001) to generate the three jumping libraries. Briefly, we performed a single tagmentation reaction using 5 ng of genomic DNA, selected the three aforementioned fragment size ranges using agarose gel

electrophoresis, and performed the remaining library preparation steps in parallel, including circularization, shearing by Bioruptor (30 sec. on, 60 sec. off, for 15 min.), streptavidin bead pull-down, end repair, addition of adenosine overhangs, Illumina indexed adapter ligation, and 15 cycles of PCR amplification. We sequenced the short-insert library (two lanes) and three jumping libraries (all in one lane) on an Illumina HiSeq2000 at the University of Oregon Genomics Core Facility (UOGCF).

To minimize the inclusion of sequencing adaptors, sequencing errors and repetitive DNA sequences in the assembly process, we used tools from the Stacks software suite (Catchen et al. 2013, Catchen et al. 2011) to adaptor-trim and discard low quality read pairs (*process_shortreads*) and filter pairs containing abundant k-mers (*kmer_filter*). Remaining were 238.6 million overlap pairs, 3.5 million 11-15 kb mate-pairs, 21.6 M 5-10 kb mate-pairs, and 44.4 M 2.5-5 kb mate-pairs, which we used for assembly with ALLPATHS-LG (Gnerre et al. 2011). Because initial k-mer spectrum analyses suggested a highly polymorphic genome, we ran ALLPATHS-LG with HAPLOIDIFY=TRUE. To assess completeness of the assembly with respect to core eukaryotic genes, we used CEGMA (Parra et al. 2009). For a summary of all Illumina sequencing data used in the assembly, see Additional File 3 at https://doi.org/10.1186/s13059-016-1126-6.

We confirmed several apparent pipefish gene losses via comparison among preliminary genome assemblies derived from independently constructed molecular libraries and generated using SGA (Simpson and Durbin 2012) and Velvet (Zerbino and Birney 2008), and via targeted Sanger sequencing. Briefly, SGA and Velvet assemblies incorporated a shotgun genomic DNA library with an insert length of 470 nt, sequenced independently with 120 nt, 100 nt, and 80 nt paired-end Illumina reads. For the SGA

assembly, the overlap value was optimized to 70 during the contig construction phase. Scaffolding was performed using SSPACE (Boetzer et al. 2011), with the three mate-pair libraries mentioned above and an additional 2-8 kb mate-pair library. These analyses filled 7 small gaps ranging from 51 to 1753 nt in the *HoxBa*, *HoxBb*, *HoxCa*, and *HoxDa* clusters. The degraded nature of *hoxa7a* was also confirmed by Sanger sequencing.

RNA-seq libraries and transcriptome assemblies

Embryo and fry transcriptome

Embryos, flushed from the pouch of lab-reared pregnant males, and fry were euthanized in Tricaine-S and stored in RNA-Later (Ambion). Tissue including the head to just posterior to the pectoral fin was dissected and pooled from 17 embryos (including 15 at 8 days post fertilization (dpf) and 2 at 10 dpf) and from 18 fry (including 2 at 16 dpf and 16 at 17 dpf). Double stranded cDNA was produced from these tissues via standard methods including RiboPure Kit (Ambion) for total RNA isolation, MicroPoly(A)Purist Kit (Ambion) for mRNA enrichment, mostly hexameric Random Primers (ThermoFisher, #48190-011) and Superscript III reverse transcriptase (Invitrogen) for first strand synthesis, and Random Primers with Kleno exo- DNA polymerase (Epicentre). Paired-end Illumina sequencing libraries were created using standard methods including mechanical shearing of the cDNA and TA ligation of adaptors (top, 5'ACACTCTTTCCCTACACGACGCTCTTCCGATC*T3'; bottom, 5'Phos-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG3'), slab gel size fractionation to isolate fragments in 200-500 bp range, and amplification using Illumina-compatible primers (5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT

CCGATCT3' and P2 reverse primer, 5'CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT3'). The library was sequenced on an Illumina GAIIx platform to produce 60 nt paired-end reads and on an Illumina HiSeq2000 platform to produce 100 nt paired-end reads (see Additional File 3 for details at https://doi.org/10.1186/s13059-016-1126-6).

Male brood pouch

Six non-pregnant and six early-stage pregnant adult males were captured from Redfish Bay, TX (Lat: 27.86795057508745, Long: -97.08869218576297), transported to the laboratory, and euthanized as described above approximately 24 hours after capture. We carefully dissected all brooding tissues, including the pouch "flaps" and epithelium, but excluding all embryonic tissue in the case of pregnant males. We fixed tissues in RNA-Later (Ambion) before freezing, homogenized by pestle upon thawing, and isolated total RNA using Trizol Reagent (Invitrogen) and RNeasy MinElute columns (Qiagen). A unique RNA-seq library was generated for each individual from 1 ug of total RNA using the TruSeq RNA v2 Kit (Illumina), and the 12 mRNA-seq libraries were sequenced across two lanes of Ilumina HiSeq 2000 100 nt paired-end reads.

De novo transcriptome assemblies

We removed low-quality and adaptor sequences from RNA-seq reads using *process_shortreads* from Stacks (Catchen et al. 2013, Catchen et al. 2011), overlapped paired-end reads using FLASH (Magoc and Salzberg 2011), and performed rare k-mer filtering and digital normalization using *kmer_filter* from Stacks. We then generated two

separate *de novo* transcriptome assemblies (one for each tissue type) from the cleaned, filtered RNA-seq data using Trinity (Grabherr et al. 2011) with --*min_kmer_cov* set to 3.

Genome annotation

Prior to genome annotation, the assembly was soft-masked for repetitive elements and areas of low complexity with RepeatMasker (Smit, Hubley, and Green 2013-2015) using a custom Gulf pipefish library created by RepeatModeler (Smit and Hubley 2008-2015), Repbase repeat libraries (Jurka et al. 2005), and a list of known transposable elements provided by MAKER (Holt and Yandell 2011). In total 15.36% of the genome assembly was masked by RepeatMasker. Repetitive elements were annotated with RepeatModeler. Hidden Markov Models for gene prediction were generated by SNAP (Korf 2004) and Augustus (Stanke and Waack 2003) and were iteratively trained for the assembly using MAKER as described by Cantarel et al. (2008). Training was performed on the five largest scaffolds and two additional scaffolds that were UTR rich; totaling 25 Mb. Evidence used by MAKER for annotation included Gulf pipefish mRNA-seq transcriptomes from embryonic head tissue and brood pouch tissue (assembled with Trinity—see below), protein sequences from threespine stickleback (*Gasterosteus aculeatus*), zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), and tilapia (*Oreochromis niloticus*) (downloaded from Ensembl: Broad S1, GRCz10, HdrR, Orenil1.0, respectively), and all Uniprot/swissprot proteins (Cunningham et al. 2015).

We filtered the annotations by MAKER to include evidence-based annotations with assembled transcriptome or protein support and those *ab initio* gene predictions that contained protein family domains as detected with InterProScan (Quevillon et al. 2005). Gene annotations were manually refined for *Hox*, astacin-like metalloprotease, and *pitx*

genes. For each annotated amino acid sequence we queried the NCBI nr database using BLASTP and compiled the results for the top BLASTP hit per gene in Additional File 2.2, SH6 at https://doi.org/10.1186/s13059-016-1126-6.

Linkage map and map integration

Mapping cross

For the genetic cross, wild male and female *S. scovelli* were captured from Redfish Bay and maintained in the lab. A total of 6 sequential broods from a single mated pair, totaling 108 G1 progeny including fry from the brood pouch plus 15 collected just prior to emergence, were gathered and flash frozen over a span of four months. Genomic DNA was isolated from individual progeny and from their parents via the Qiagen DNeasy Kit. RAD-seq libraries were made using the restriction enzyme SbfI as in Baird et al. (2008), Hohenlohe et al. (2010), and Etter et al. (2011) with the Illumina-compatible, barcoded P1 adapters and primer types used in Hohenlohe et al. (2012) and the P2 adapter type used in Hohenlohe et al. (2010). Single-end reads of 100 nt were produced from two lanes on an Illumina HiSeq2000 (see Additional File 3 for details at https://doi.org/10.1186/s13059-016-1126-6). The parents were sequenced to greater depth than progeny (see below) to make an accurate catalog of diploid genotypes possible in the cross.

Marker genotyping

The two lanes of Illumina data resulted in 367,085,475 raw reads which were analyzed using the software, Stacks (Catchen et al. 2013, Catchen et al. 2011). Using the *process_radtags* program, reads were demultiplexed according to barcode and discarded

if the barcode could not be determined after correcting for sequencing error, if the restriction enzyme cut site was not intact, or if the sequencing quality was too degraded. The 218,309,324 remaining reads were analyzed by the Stacks *de novo* pipeline to assemble and genotype the RAD loci. A minimum of three identical reads (-m 3) was required to form a "stack" or putative allele in each individual, up to five differences were allowed when merging stacks into putative loci (-M 5) and up to 3 differences were allowed when merging loci from different individuals into the catalog (-n 3) to accommodate fixed differences between the cross parents. The *genotypes* program from Stacks was used to export data in a CP cross format for use in JoinMap, and the genotypes were uploaded to the Stacks web interface. Genotype data with markers present in at least 75 of the 108 individual progeny was exported from the web interface for linkage analysis.

Map construction

Linkage analysis was performed with JoinMap 4.1 (Van Ooijen 2006) using only markers that were present in at least 75 of the 108 individual progeny. Markers were initially grouped in JoinMap 4.1 using the "independence LOD" parameter under "population grouping" at a minimum LOD value of 15.0, and markers that remained unlinked at LOD<15 were excluded. Marker sets were partitioned into paternal and maternal markers to enable the construction of sex-specific linkage maps. Marker ordering was performed using the Maximum Likelihood (ML) algorithm in JoinMap 4.1 with default parameters. Supposed double recombinants were identified using the "genotype probabilities" feature in JoinMap4.1 and by visual inspection of the colorized graphical genotypes in the male, female and consensus maps. After visual inspection of

the individual sequences in the web interface of Stacks, markers were manually corrected as needed in the web interface and re-exported. For example, if a double recombinant was a homozygote with a small number of sequences, the genotype was eliminated because it might represent a heterozygote with no sequences for the second allele. Conversely, if the double recombinant was a heterozygote with only one sequence for the second allele, the genotype was eliminated because the second sequence could be sequencing error. The new dataset with corrected genotypes was loaded again into JoinMap 4.1 and the process was repeated until no suspect genotypes were identified. The "expected recombination count" feature in JoinMap4.1 was used to identify individuals with higher than expected recombination events; marker order was visually inspected and, when necessary, optimized by moving a marker or sets of markers to a new map position that reduced the number of recombination events. When a marker or sets of markers could be in multiple map positions, the markers were moved to a position congruent with their physically aligned scaffold location if there was no cost to the map.

Integrating the assembly and the linkage map

The 4,375 markers from the linkage analysis were integrated with the assembled pipefish scaffolds to create a chromonome using the software, Chromonomer (http://catchenlab.life.illinois.edu/chromonomer/). Markers were aligned to the set of assembled pipefish scaffolds using GSnap (Wu and Nacu 2010), requiring unique alignments, allowing up to five mismatches (-m 5), counting gaps as four mismatches (-i 4), and requiring 99% of the RAD locus to align (--min-coverage=0.99). The AGP file produced by ALLPATHS-LG that describes the assembly, the linkage group and map position of the markers in the map, the alignments of the markers to the scaffolds, and the

FASTA file containing the sequence from the assembly are all fed into Chromonomer, which integrates them in the following way. First, markers are arrayed along the scaffolds they are aligned to and scaffolds that have markers from more than one linkage group are identified (no scaffolds were split between linkage groups). A coherent ordering of markers must be found for each scaffold so that physical base pair and map position are consistent among all markers for that scaffold. Markers that are out of order with respect to the map or scaffold are discarded (unless it is the last marker holding a scaffold into the map). Of the 4,375 markers, 649 were excluded in this phase, leaving 3,726 markers in the final "chromonome". If a scaffold spans more than one map position, and physical order is the same as map order, the orientation of the scaffold is positive. If physical and map order are inverted, the scaffold is considered in negative orientation and the sequence is reverse complemented. Otherwise orientation is unknown and the scaffold remains in positive orientation by default. Scaffolds are then hung from the linkage group they occur on, according to map position. Ordered markers may place the scaffold in more than one place within the linkage group, that is, one or more scaffolds occur within the focal scaffold according to the linkage map. This can be due to an incorrect assembly join, or because a smaller scaffold is filling a gap in a larger scaffold. In these cases, the scaffold is split at the largest gap that can be found between the markers in the map that indicate where the split must occur. Starting with 553 scaffolds, 5 scaffolds were split one time each for a total of 558 scaffolds in the chromonome. Sequence from the scaffolds is then concatenated into chromosomes according to the orientation and integrated order with standard 100bp gaps placed in between each join resulting in a chromonome of 266,330,253bp (53.6Kb scaffold join gaps) with 40,734,039bp of sequence remaining in unintegrated scaffolds. Finally, the genome annotation is translated to the new

chromonome providing a genome-level ordering of genes for use in conserved synteny analysis and new AGP, FASTA, and GFF files are generated to describe the chromonome.

Conserved synteny analysis

In order to visualized evolutionarily conserved gene neighborhoods—i.e. conserved synteny, we used the Synolog software (Catchen, unpublished). We used Synolog to identify orthologs between the Gulf pipefish, threespine stickleback, medaka, green spotted pufferfish (*Tetraodon nigroviridis*), zebrafish, spotted gar, and southern platyfish and to identify conserved gene neighborhoods pairwise between the different species. Genome-wide images of conserved synteny were drawn by Synolog by combining the conserved synteny blocks across the genome and incorporating the integrated linkage map/assembly output by Chromonomer where appropriate (Figure 2.2c). Protein gene models for each non-pipefish species were downloaded from Ensembl. While Synolog is a new and independent implementation, the algorithm to identify conserved synteny and the biological inferences stemming from its application are as described in Catchen, et al. (2009).

Phylogenomic analysis using ultraconserved elements

We added ultraconserved elements (UCEs) from Gulf pipefish, Pacific bluefin tuna, and southern platyfish genomes to an existing UCE dataset containing sequences for 27 actinopterygiian fishes and published by Faircloth et al. (2013). To retrieve each of the 491 UCEs from the three genomes above we generated a consensus sequence of each alignment from Faircloth et al. (2013) using *em_cons* from EMBOSS (Rice, Longden,

and Bleasby 2000), searched for each consensus sequence in each genome using LASTZ (Harris 2007), and extracted unique search hits from each genome using BEDTools (Quinlan 2014). For this we used the tuna reference genome available from http://nrifs.fra.affrc.go.jp/ResearchCenter/5_AG/genomes/Tuna_DNAmicroarray/index.html and the platyfish genome from Ensembl. We obtained 457, 453, and 479 single-copy UCEs for Gulf pipefish, tuna and platyfish, respectively. A multiple sequence alignment for each UCE was generated using MAFFT v7 (Katoh and Standley 2013) with options —localpair and —maxiterate 1000, and minor manual adjustments were made when necessary.

We performed substitution model selection for each UCE alignment using the corrected Akaike Information Criterion (AICc), as implemented in jModeltest-2.1.10 (Guindon and Gascuel 2003, Darriba et al. 2012). The GTR+gamma model was selected for the largest percentage of the total aligned sequence data. We concatenated UCE alignments, ordering them so that the loci having the same best-fitting substitution model were grouped together. We proceeded with a partitioned phylogenetic analysis using the concatenated alignment (153,032 nt total), and the GTR+gamma model for all partitions. Maximum likelihood (ML) phylogenetic inferences were conducted with RAxML version 8.2.4 (Stamatakis 2014) using default settings. We produced a consensus ML tree using the rapid bootstrap search algorithm described in Stamatakis et al. (2008). Briefly, 1000 rapid bootstrap searches were conducted, followed by fast ML searches on 200 of these, followed by a slow ML search on the 10 best fast ML trees. Clade confidence was assessed with SH-aLRT support values and bootstrap replicate frequencies. We specified *Polypterus senegalus* as the outgroup for tree rooting.

Characterization of *Hox* clusters

Hox gene content

Teleost *Hox* gene sequences acquired from Ensembl were used as queries for BLAST searches of the final Gulf pipefish genome assembly using Geneious (version 8.0.5). Exon boundaries were annotated by hand using alignments with the query *Hox* genes. The *Hox* genes annotated in the Gulf pipefish assembly were then BLAST-searched against the NCBI NR sequence database to confirm gene identity using Geneious (version 8.0.5). Additionally, *Hox* genes were identified, following the method outlined above, in the Pacific bluefin tuna genome (see genome source above) (Nakamura et al. 2013).

Hox cluster microRNAs and long-noncoding RNAs within the *Hox* cluster were identified using VISTA analyses based on conserved noncoding elements (CNE) within *Hox* clusters across Gulf pipefish, threespine stickleback, mouse (*Mus musculus*), spotted gar, zebrafish, Pacific bluefin tuna, medaka, and fugu (*Takifugu rubripes*) (Frazer et al. 2004, Mayor et al. 2000, Brudno, Do, et al. 2003, Brudno, Malde, et al. 2003). We aligned primary miRBase (Kozomara and Griffiths-Jones 2011) microRNA sequences from stickleback, zebrafish, medaka, and fugu to *S. scovelli Hox* regions using MUSCLE (Edgar 2004) to supplement annotations. The hairpin loops of the annotated microRNAs were confirmed using RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi). When known *Hox* cluster microRNAs were not detected in the Gulf pipefish genome, we further confirmed absence of the conserved seed sequence, which was the case for *mir196b* between *hoxb13a* and *hoxb9a* and *mir10a* between *hoxb5b* and *hoxb3b*. All conserved noncoding sequences annotated within the Gulf pipefish *Hox* cluster were queried against miRBase Sequence Databases (Release 21) for mature miRNA chordate

sequences and miRNA chordate hairpins (downloaded from miRBase) using BBMapSkimmer (Bushnell) for further identification of microRNAs. Kmer index size was set to 7, max indel set to 0, approximate minimum alignment identity set to 0.50, secondary site score ratio set to 0.25, behavior on ambiguously-mapped reads set to retain all top-scoring sites, and maximum number of total alignments to print per read set to 4 million. See Additional File 2.2, SH7 for scaffold locations and sequences of microRNAs and long non-coding genes at https://doi.org/10.1186/s13059-016-1126-6.

Characterization of *dlx* CNEs

CNEs between *dlx1* and *dlx2*, between *dlx3* and *dlx4*, and between *dlx5* and *dlx6* were identified using mVISTA analyses based on levels of sequence conservation within *dlx* clusters across Gulf pipefish, Atlantic cod, threespine stickleback, zebrafish, human, Pacific bluefin tuna, medaka, and fugu (Frazer et al. 2004, Mayor et al. 2000, Brudno, Do, et al. 2003, Brudno, Malde, et al. 2003). Sequences were downloaded from Ensembl for cod, stickleback, zebrafish, human, medaka, and fugu. Tuna sequences were downloaded from the reference genome source cited above. Medaka was set as the reference sequence for the *dlx1/2* and *dlx5/6* comparisons and stickleback was the reference for the *dlx3/4* comparisons. Alignment of each sequence from these species were aligned using the shuffle-LAGAN algorithm through the mVISTA website under default parameters. See Additional File 2.2, SH7 for scaffold locations of CNEs at https://doi.org/10.1186/s13059-016-1126-6.

Characterization of pelvic fin development candidates

Pitx1, Pitx2, and Pitx3 protein sequences were obtained from our pipefish annotation, Ensembl, and Genbank (in the case of stickleback Pitx1) for human, coelacanth (*Latimeria chalumnae*), spotted gar, zebrafish, blind cavefish (*Astyanax mexicanus*), medaka, tilapia, green spotted pufferfish, and threespine stickleback, and aligned using MAFFT (with default settings). To isolate DNA fragments for Sanger sequencing of *pitx1* from the messmate pipefish (*Corythoichthys haematopterus)* and the robust ghost pipefish (*Solenostomus cyanopterus)* genomic DNA, we designed degenerate PCR primers (in IUPAC notation, forward 5'-CGGAGCGCAACCAGCARATGGA-3' and reverse 5'-GGACGACGACATGSCSCWGTTGAT-3') for amplification using Phusion DNA polymerase (New England Biolabs) in Phusion HF buffer, and an annealing temperature of 55ºC.

Because *tbx4* was not represented in the pipefish genome annotation, we attempted to determine its location in the genome assembly manually by using a targeted profile Hidden Markov Model (HMM) generated from several aligned teleost Tbx4 protein sequences. HMM-based approaches are more sensitive than BLAST-based approaches when searching for divergent homologs (Karplus, Barrett, and Hughey 1998), a possible scenario when a gene has evolved rapidly or has degenerated. Briefly, we used an alignment of Ensembl Tbx4 sequences from spotted gar, zebrafish, medaka, southern platyfish, threespine stickleback, green spotted pufferfish, and tilapia to generate a profile Hidden Markov Model (HMM) with hmmer2 (Johnson, Eddy, and Portugaly 2010), then searched for sequences in the Gulf pipefish genome with this model using the

genewisedb program of wise2 (http://www.ebi.ac.uk/~birney/wise2/) with default search settings.

Differential expression analysis

We aligned adaptor- and low-quality-trimmed, forward reads from the 12 brood pouch RNA-seq libraries to the annotated Gulf pipefish genome using GSNAP (Wu and Nacu 2010). We counted the number of uniquely-mapped reads per exonic region of each annotated gene using HTSeq-count (Anders, Pyl, and Huber 2015), and used the counts to test for differential gene expression between pregnant and non-pregnant males using the negative binomial exact test (Robinson and Smyth 2008), after TMM normalization, implemented by the R/Bioconductor package edgeR (Robinson, McCarthy, and Smyth 2010). We limited differential expression analysis to those genes with at least one read per million counted (cpm) in at least four of the 12 fish, which reduced the data set to 15,253 genes.

To connect genes annotated in the pipefish genome with putative functional information, we mapped the pipefish amino acid sequences to KEGG Orthology (KO) entries (Kanehisa et al. 2016) using the KEGG Automatic Annotation Server (Moriya et al. 2007). We then identified KEGG PATHWAYS enriched for pipefish KOs with extreme $\log_2$ fold change values from the pregnancy differential expression analysis using the R/Bioconductor package GAGE (Luo et al. 2009). To visualize individual members of KEGG PATHWAYS enriched for pregnancy-sensitive genes we used the R/Bioconductor package Pathview (Luo and Brouwer 2013). We also used ENSEMBL IDs for putative *D. rerio* orthologs of Gulf pipefish genes to test for overrepresentation of PANTHER GO-slim Biological Process terms among pregnancy-enriched and

pregnancy-depressed genes using binomial tests implemented by the online resource PANTHER (pantherdb.org), (Mi et al. 2013, Mi et al. 2016). For the overrepresentation tests we used all genes tested for differential expression (see above) and matched with a zebrafish ortholog as the comparison set. To interpret the results of overrepresentation tests for pregnancy-enriched and -depressed sets we only considered GO-Slim terms represented in the comparison set by at least five genes, and we controlled the False Discovery Rate at 0.1 as in Benjamini & Hochberg (Benjamini and Hochberg 1995). Results for these overrepresentation tests are in Additional File 2.2, SH4 (PregUp GOs Overrepresented) and Additional File 2.2, SH5 (PregDown GOs Overrepresented) at https://doi.org/10.1186/s13059-016-1126-6.

To visualize and quantify multivariate differences among individual brooding tissue samples in transcript space, we calculated Bray-Curtis dissimilarity based on TMM-normalized cpm values, performed non-metric multidimensional scaling (nMDS), and conducted permutation-based multivariate analysis of variance (perMANOVA) to test for a global transcriptional effect of pregnancy status, all using the R package vegan (Oksanen et al. 2015). Similarly, to visualize clustering of genes and pouch libraries via co-expression patterns, we generated heat maps for all pouch-expressed genes and several immune system related KEGG pathways. Ward clustering was used, based on Euclidean distance calculated from scaled, $\log_2$-transformed cpm values, implemented by the R function hclust. Unless noted otherwise, all additional analyses related to the gene expression were conducted using core packages within the statistical programming language R (Team 2015).

Characterization of *patristacins*

Previous work identified members of the astacin-like metalloprotease gene family as candidates for playing a functional role in male pregnancy (Harlin-Cognato, Hoffman, and Jones 2006, Small, Harlin-Cognato, and Jones 2013). We confirmed extreme transcriptional differences for two of these *patristacins* between brood pouch tissue of pregnant and non-pregnant males (see differential expression section) and set out to characterize the distribution of this gene family in the Gulf pipefish and other teleost genomes. We compared protein sequences from pipefish gene annotations bearing similarity to *patristacins* against the Ensembl zebrafish GRCz10 protein set using BLAST and discovered that all similar zebrafish homologs belong to Ensembl protein family ENSFM00500000270265 (choriolytic enzymes). We used all actinopteryigiian fish sequences from this Ensembl protein family alignment to generate a Hidden Markov Model (HMM) profile using hmmer2 (Johnson, Eddy, and Portugaly 2010), then searched for similar sequences in the Gulf pipefish genome using the genewisedb program of wise2 (http://www.ebi.ac.uk/~birney/wise2/) with default search settings. These protein family-specific annotations allowed us to both correct and supplement initial MAKER annotations as necessary. Most of the *S. scovelli* astacin-like metalloproteases annotated in this manner, including at least 4 tandemly arrayed patristacins on scaffold 62, shared high sequence similarity with zebrafish homologs from Ensembl protein family ENSFM00500000270265. Six of the *S. scovelli* astacin-like metalloproteases were most similar to three additional Ensembl protein families, including ENSFM00500000282854 (Metalloendopeptidases), ENSFM00570000851071 (Bone morphogenetic 1/Tolloid-like proteins), and ENSFM00500000270104 (Meprins).

To identify potential *patristacin* orthologs and/or close paralogs in several teleost genomes, we repeated the HMM search using a hmmer2 profile generated from an alignment of the four pipefish *patristacins*, but included the Gulf pipefish assembly, and the Ensembl genomes of spotted gar, zebrafish, platyfish, and green spotted pufferfish as targets. Hits from these searches were used to understand the evolution of patristacins in the syngnathid lineage. Excluding hits that corresponded to the more distantly paralogous Bmp1/Tolloid-like and Merprin proteins (Mohrlen et al. 2006), with the exception of Meprin1b as an outgroup (see Figure 2.7), we aligned all unique astacin-like amino acid sequences from the aforementioned actinopterygii genomes with MAFFT v7 (Katoh and Standley 2013) using options —localpair and —maxiterate 1000. We then made manual adjustments to the alignment by removing non-conserved residues at the ends, yielding a final alignment of 55 sequences, covering 269 amino acids. We used the PhyML 3.0 web server (Guindon et al. 2010) for Akaike Information Criterion (AIC) model selection and ML phylogenetic inference. The WAG+G+I+F model was selected, and we proceeded with two separate evaluations of ML tree clade support: PhyML's fast SH-like aLRT, and 500 bootstrap replicates.

RESULTS

The pipefish genome assembly is of high quality and completeness

The only published estimate of Gulf pipefish genome size is based on Feulgen staining, (Hardie and Hebert 2004), from which a haploid genome size of 523.23 Mb was calculated for the species. We obtained a short read k-mer-based genome length estimate of 351.44 Mb using ALLPATHS-LG (Gnerre et al. 2011). Using the RAD markers from our genetic map to estimate the number of RAD sites per scaffold and infer the amount of

sequence missing from the assembly by estimating the number of missing RAD sites, we obtained an estimated genome size of 334 Mb. These data suggest that, consistent with the k-mer-based estimate, no more than approximately 27 Mb, or 8% of sequence is missing from the assembly (not including repetitive sequence), and that the Feulgen estimate is likely too large.

We assembled overlapping and mate-pair Illumina paired-end 100 nt reads (176X total coverage of 351 Mb) into 2,123 scaffolds, yielding an assembly length of 307.02 Mb with 6.58% gaps. Contig and scaffold N50 were 32.24 kb and 640.41 kb, respectively, and the maximum scaffold size was 6.71 Mb. An analysis of core eukaryotic genes (CEGs) using CEGMA (Parra et al. 2009) revealed that our assembly contained complete information for 245 of 248 CEGs and "partial" information for the remaining 3 CEGs. These assembly quality metrics are comparable to other recently published, high-quality scaffold-level genomes for fishes. Table 2.1 presents a side-by-side comparison of the Gulf pipefish assembly with several other published ray-finned fish assemblies.

Table 2.1: Scaffold-level assembly statistics for the Gulf pipefish genome.

Genome	# of Scaffolds	Longest Scaffold	Scaffold N50	Contig N50	Assembly Length	% Gaps	% CEGs Complete
Gulf pipefish (*Syngnathus scovelli*)	2,104	6.7 Mb	640.4 kb	32.2 kb	307.0 Mb	6.6%	98.8%
African turquoise killifish (*Nothobranchius furzeri*)	29,054	0.7 Mb	119.7 kb	8.7 kb	1010.9 Mb	7.7%	94.8%
blind cave fish (*Astyanax mexicanus*)	10,542	9.8 Mb	1775.3 kb	14.7 kb	1191.1 Mb	19.1%	87.9%
spotted gar (*Lepisosteus oculatus*)	2,105	21.3 Mb	6928.1 kb	68.3 kb	945.8 Mb	8.1%	90.7%

The genome assembly of *S. scovelli* is comparable in quality to three recently published fish reference genomes. Shown are assembly statistics calculated from scaffold-level genome assemblies, considering scaffolds 1000 nt and longer, except for the 248-gene CEGMA analysis, which was applied to all scaffolds. Assembly versions are *N. furzeri* GCA_000878545.1 (Valenzano et al. 2015), *A. mexicanus* GCA_000372685.1 (McGaugh et al. 2014), and *L. oculatus* GCF_000242695.1 (Braasch et al. 2016).

Using MAKER, (Holt and Yandell 2011), we initially generated 37,696 total protein-coding gene annotations, but we retained only 20,834 of these based on biological evidence from protein databases, RNA-seq data, or protein domain detection. After manual annotation correction for several genes of interest, the final annotation included 20,841 protein-coding genes. Mean and median protein sequence length were 539.55 and 386.00 amino acids, respectively.

A genetic map integrates 87% of the genome assembly into chromosomes

To order and orient scaffolds and to unite them into chromosomes, we generated an G1 pseudo-test cross genetic linkage map from a cross of wild *S. scovelli* with 108 progeny. Of 21,680 RAD tags, 4,779 polymorphic tags were informative and met our criteria for inclusion in the genetic map (see methods). The genetic map readily coalesced into 22 distinct linkage groups (see Additional File 1, Fig. S1 for schematics of the consensus genetic map at https://doi.org/10.1186/s13059-016-1126-6). Markers could be aligned to 553 scaffolds, thereby tying nearly 266.3Mb—87%—to chromosome models (see Additional File 1, Fig. S2 for plotted lengths and gene densities of the scaffolds at https://doi.org/10.1186/s13059-016-1126-6). 271 scaffolds (49%) were anchored at more

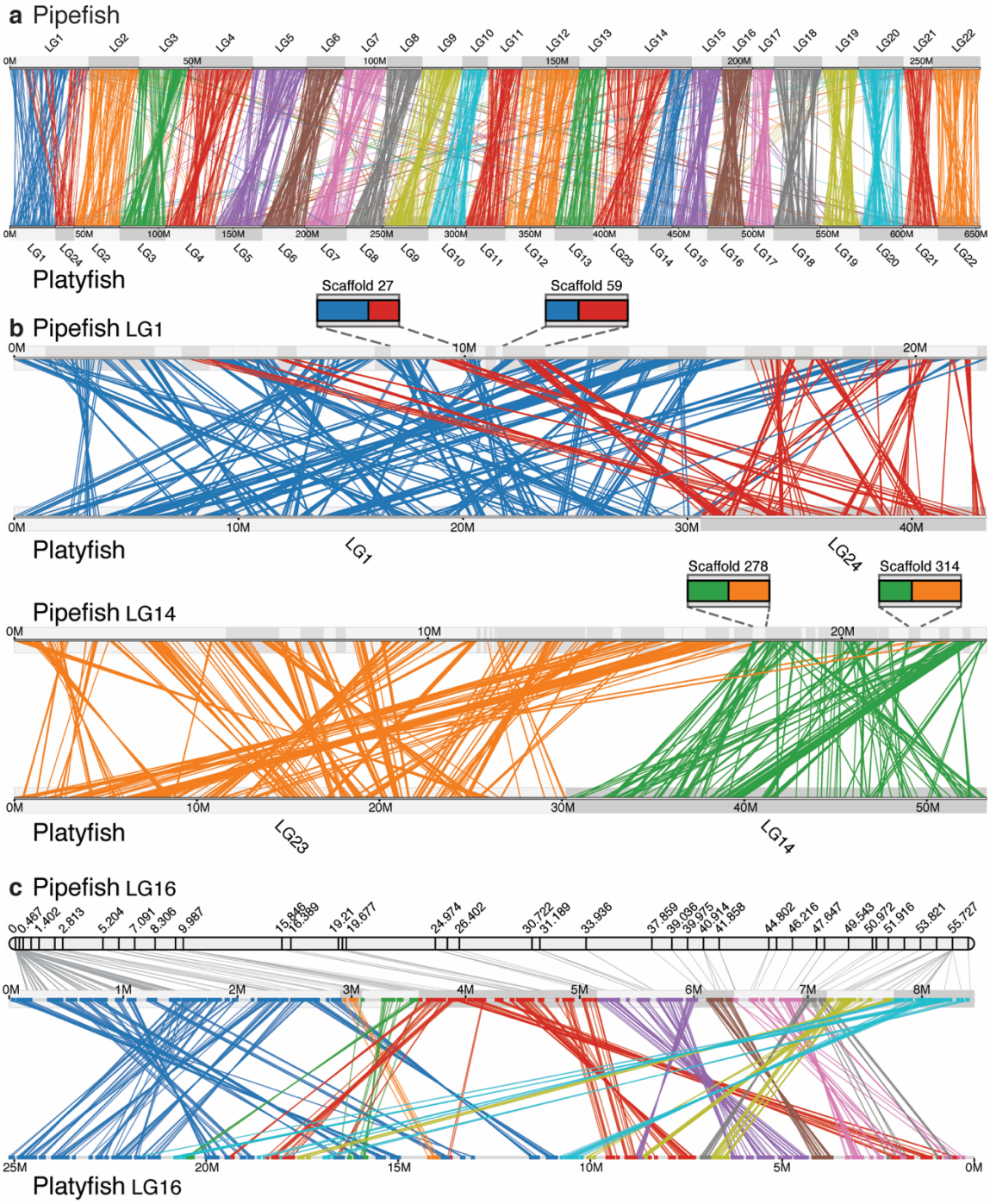
than one map position with two or more markers, which allowed us to assign an orientation. Unplaced scaffolds tended to be shorter and more depauperate of annotated genes, on average, than scaffolds incorporated into chromosomes (see Additional File 1, Fig. S2 for plotted lengths and gene densities of the scaffolds at https://doi.org/10.1186/s13059-016-1126-6). Possibly the same sequence characteristics that make assembly difficult—a higher occurrence of repetitive DNA—could help explain the lower gene density of these smaller scaffolds. There were few initial conflicts between the genome assembly and the linkage map, and none that could not be ruled out as artefactual due to poor support. For instance, three scaffolds were initially tied to more than one linkage group; in all three cases, however, only a single marker, with equivalent alignments to multiple locations, created this conflict and could be reasonably ruled incorrect, particularly when patterns of conserved synteny were taken into account. There were also apparent within-linkage group conflicts, which in most cases could be resolved by movement of markers without any cost to the linkage map. In total, five scaffolds where conflicts remained were split by our software Chromonomer (see methods) to reconcile the map and the assembly; in each of these cases, a small scaffold (1.2 to 3.1 kb) was inserted into a gap in a larger scaffold. Only the largest of these small scaffolds contained an annotated gene, and in that case, its insertion into the larger scaffold agreed with the relative position of its ortholog in other teleost genomes.

Chromosome evolution is revealed by patterns of conserved synteny

Evidence based on ancestral state reconstruction supports an ancestral chromosome number of 24 in the teleosts (Mank and Avise 2006). Though chromosome number has been shown to vary across the broad group of Syngnathidae, the 22 linkage

groups that coalesced in this linkage map in *S. scovelli* accords well with published karyotypes for two other species in *Syngnathus*, *S. abaster* and *S. typhle* (Vitturi et al. 1998). Using a genome-wide synteny analysis, we investigated how this change from the ancestral chromosome number likely occurred. Genes are called syntenic when they lie on the same chromosome or chromosomal segment, and a pair of compared genomes show "conserved synteny" when orthologous genes that are syntenic in one genome also lie together, though not necessarily in the same gene order, in the comparator genome. The pattern of conserved synteny between Gulf pipefish and other teleosts, such as southern platyfish (*Xiphophorus maculatus*), which has the ancestral number of chromosomes (Figure 2.2A), suggests that the reduced chromosome number in *Syngnathus* resulted simply from two chromosomal fusions (Figure 2.2B). Two large blocks covering the length of one linkage group in *S. scovelli* have strong conserved synteny of orthologs along both platyfish LG 1 and 24, respectively, and another pair of blocks covering all of a second pipefish linkage group are orthologous to platyfish LG 14

Figure 2.2: (next page) Chromosomal rearrangements inferred from a conserved synteny comparison. a) Pipefish and platyfish chromosomes are broadly congruent. Strings connecting orthologous genes between the species' genomes are colored by pipefish chromosome. b) Pipefish LG 1 and 14 are each orthologous to two platyfish chromosomes, likely because chromosome fusions occurred in the syngnathid lineage. Several scaffolds from fused chromosomes 1 and from 14, including those shown in the insets, show blocks of conserved synteny to both "ancestral" chromosomes in platyfish (LG 1 and 24 or LG 14 and 23). This pattern indicates that some number of intra-chromosomal rearrangements blended segments across the chromosomal junction after the chromosomes fused. Strings connecting orthologs are color-coded by platyfish chromosome. Pipefish scaffolds are shown in alternately shaded rectangles along the chromosome. c) On LG 16, differences in the orientation and location of orthologous gene blocks suggest inversions and transpositions have occurred since the last common ancestor of pipefish and platyfish. Strings connecting orthologous genes are colored according to the pipefish scaffold each gene resides on. Support for scaffold order and orientation can be seen in the linkage map for pipefish LG 16, shown above.


a Pipefish
Platyfish
b Pipefish LG1
Scaffold 27
Scaffold 59
Platyfish
LG1
LG24
Pipefish LG14
Scaffold 278
Scaffold 314
Platyfish
LG23
LG14
c Pipefish LG16
Platyfish LG16

and 23 (Figure 2.2B). The resulting pipefish chromosomes, which we here name LG 1 and 14 to reflect this orthology, are the largest in the genome. Several scaffolds linked to pipefish LG1 and LG14 contain genes orthologous to the two ancestral chromosomes that constitute each of them (Figure 2.2B), suggesting that intra-chromosomal rearrangements have blended the original margins of the chromosomes since they became fused.

Other within-chromosome rearrangements relative to various teleost reference genomes can be confidently inferred using the pipefish assembly and linkage map, where they provide mutual support. It is beyond the scope of this paper to catalogue such chromosomal differences and is the subject of other studies. As an example, however, pipefish LG 16 can be used to illustrate a subset of these rearrangements because all scaffolds that map to this linkage group are ordered and all but two very small scaffolds are oriented, with strong map support. Here, likely inversions and transpositions can be discerned in a comparison between pipefish and platyfish, based on stretches of conserved synteny of protein coding genes (Figure 2.2C).

Phylogenomic analysis supports an alternative hypothesis for the position of syngnathiform fishes among the Percomorpha

Knowing the phylogenetic placement of syngnathid fishes relative to other teleosts with sequenced genomes is critical for using comparative genomic approaches to polarize the evolution of traits in the Syngnathidae. Conflicting hypotheses regarding the origin of syngnathid fishes and their relatives are a barrier to this understanding, and resolving phylogenetic relationships for the crown clade of teleosts (Superorder Percomorpha) in general has been a problem (Nelson 1989, Betancur-R et al. 2013, Sanciangco, Carpenter, and Betancur 2016).

Ultraconserved elements (UCEs) offer a genome-wide alternative to small panels of nuclear and mitochondrial phylogenetic markers because they exist by the hundreds or thousands in vertebrate genomes, are often easily identifiable as well-conserved, single-copy orthologs that contain divergent regions, and can be used to address hypotheses over a broad range of phylogenetic scales (Faircloth et al. 2012). Faircloth et al. (2013) used UCEs to produce a well-supported phylogeny at both deep and shallow time scales for ray-finned fishes. We added to this dataset UCEs from Gulf pipefish, Pacific bluefin tuna (*Thunnus orientalis*), and southern platyfish, and performed phylogenetic analysis. Interestingly, our phylogenomic analysis provides an alternative hypothesis regarding the relationships among Scombriformes (tunas and their relatives) and Syngnathiformes (Syngnathid fishes and their relatives). Briefly, the two orders would not be interpreted as a monophyletic clade from our topology, in contrast to conclusions based on trees inferred by others (Betancur-R et al. 2013, Sanciangco, Carpenter, and Betancur 2016, Near et al. 2013). Statistical support for clades bracketing this region of the topology was high (Figure 2.3), but should be interpreted with caution given evidence that phylogenetic discordance across different regions of the genome can limit the accuracy of species-level inferences based on concatenated sequence data (Edwards, Liu, and Pearl 2007, Kubatko and Degnan 2007). We recovered all relationships reported by Faircloth et al. (2013) and found, consistent with previous studies (Betancur-R et al. 2013, Sanciangco, Carpenter, and Betancur 2016, Near et al. 2013), that the Syngnathiformes are not nested within the clade containing species commonly used in genetic and genomic studies (i.e., medaka, platyfish, stickleback, and pufferfish). Given this phylogenetic hypothesis for the origin of syngnathids, the Gulf pipefish genome fills a useful outgroup role in comparative genomics studies using these model species. The currently understood relationships also

highlight a need for phylogenetic analyses including fish lineages that diverged just prior to origin of the syngnathids, in order to help understand the unusual derived traits in the Syngnathidae.

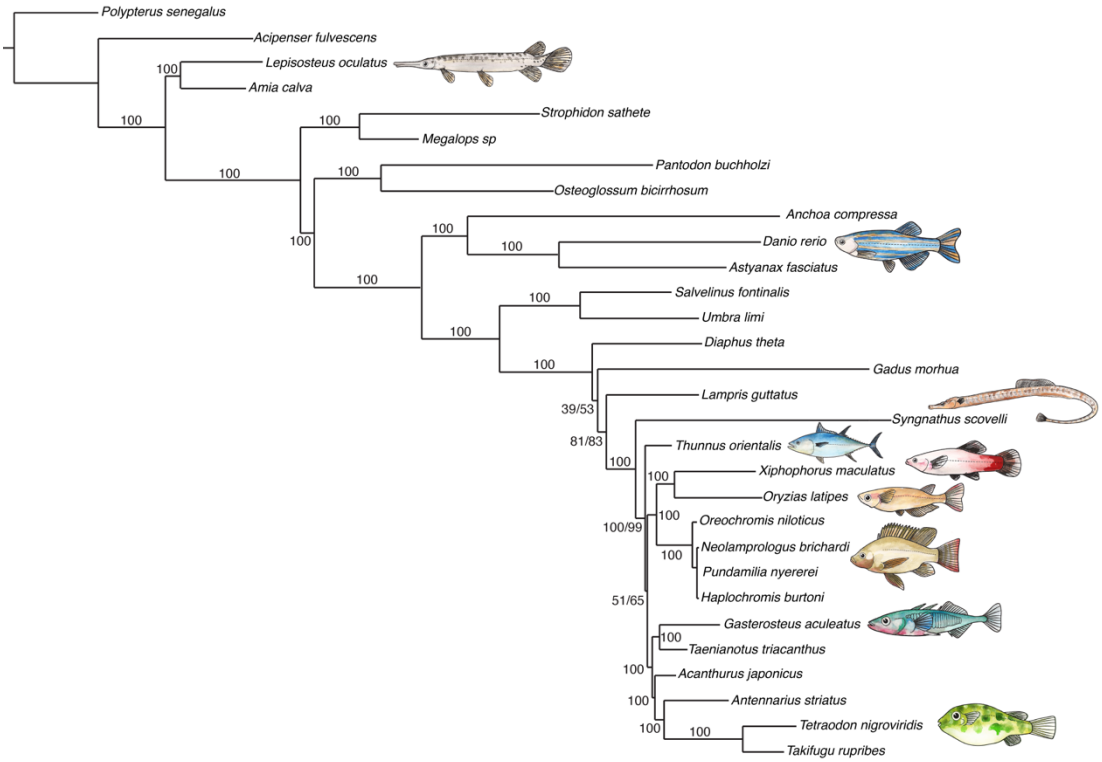


Figure 2.3: Phylogenomic inference supports a syngnathiform clade distinct from the clade containing commonly studied fish models. A well-supported maximum likelihood tree of ultraconserved elements places Syngnathiformes as an outgroup relative to fellow percomorph species used as genetic models, consistent with previous work regarding the molecular systematics of Percomorpha [29, 30, 33]. Note, however, that our topology is not consistent with a monophyletic group including Syngnathiformes and Scombriformes, as previously reported. Bootstrap and SH-aLRT support is listed for each node; a single number is listed where both values agree.

Convergent and unique gene losses have occurred in the pipefish *Hox* clusters

The *Hox* clusters, which include tandem arrays of homeobox genes interspersed with non-coding RNAs that regulate *Hox* and other genes, are critical for patterning the body axis and paired appendages (reviewed in (Zakany and Duboule 2007, Mallo,

Wellik, and Deschamps 2010, Mallo and Alonso 2013)). Pipefish have elongated bodies, including more trunk and especially more caudal vertebrae than relatives like medaka and threespine stickleback, and they lack pelvic fins, key examples of derived traits depicted in cartoon form in Figure 2.1. We therefore scrutinized the gene content of the *Hox* clusters for differences from pipefish's percomorph relatives (including pufferfish, medaka, stickleback, and tuna). Just as in many other gene families, differential loss of *Hox* genes among lineages followed the whole genome duplication that occurred near the base of the teleost lineage (e.g., (Amores et al. 1998)). Gulf pipefish appears to share some of these losses with other percomorph fishes, to the exclusion of the outgroup lineage zebrafish (Figure 2.4). A parsimonious interpretation of the pattern of losses suggests that *hoxb10a*, *hoxb8b*, *hoxd13a*, the entire *HoxCb* cluster and *mir196c* were absent in the common ancestor of pipefish and other percomorphs. Several other *Hox* cluster genes have been lost in pipefish as well as in some but not all model percomorphs; based on the topology of the phylogenetic tree in Figure 2.3 and those inferred by others (Betancur-R et al. 2013, Sanciangco, Carpenter, and Betancur 2016, Near et al. 2013), we conclude that these losses are likely to be convergent (Figure 2.4). These include *hoxa7a*, *hoxb7a*, *hoxc3a*, *hoxc1a*, *mir196b* in the *HoxBa* cluster and *mir10a* in the *HoxBb* cluster. For example, *hoxb7a* was likely lost independently at least three times (in pufferfish, medaka and pipefish), but it is still present in stickleback and tuna. *hoxa7a* was lost independently in both pipefish and pufferfish, leaving both lineages with no *hox7* paralog in any cluster. By contrast, zebrafish and all of the other percomorphs surveyed here retain either *hoxa7a* or *hoxb7a* or they have both of these genes. There is a remnant of the pipefish *hoxa7a* sequence, found between *hoxa5a* and *hoxa9a*; it is likely a pseudogene, as there is no trace of the sequence for the homeobox-containing second exon and an

early stop codon in the first exon is predicted also to eliminate the hexapeptide. In addition to these losses, the pipefish *HoxBa* cluster remarkably no longer has *evenskipped* gene *eve1*, a gene that is present in zebrafish and all other percomorphs compared here (Figure 2.4). We detected pipefish sequences for orthologs of long non-coding RNA genes *hotairm1* between *hoxa1a* and *hoxa2a,* and *hottip* between *evx1* and *hoxa13a* (not shown). *hotairm1* is missing in zebrafish and so far, unreported in any teleost (though annotated in the Ensembl reference genome for spotted gar, an actinopterygiian basal to the teleosts).

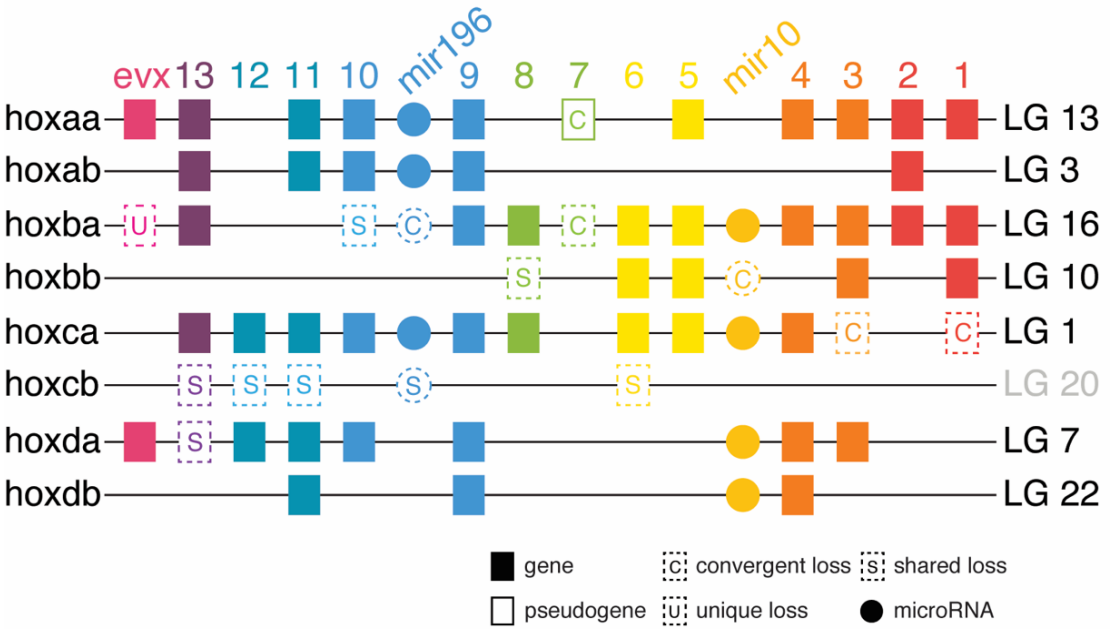


Figure 2.4: The pipefish *Hox* clusters have experienced convergent and unique gene losses. A cartoon of the *Hox* clusters in *S. scovelli*, with boxes representing genes arranged along chromosome segments of different linkage groups, summarizes gene content changes relative to other teleosts. Seven gene losses, of both coding and non-coding genes, are here labeled shared losses among the compared percomorph lineages because these genes are retained by the non-percomorph outgroup, zebrafish. Six other pipefish gene losses are inferred to be convergent losses with respect to some members of Percomorpha because other species that are not pipefish sister lineages have also lost these genes. *Hox* cluster-associated *evenskipped* gene *eve1* (a member of the *evx* paralogy group) is missing in pipefish, a loss that hasn't been reported in other teleosts. Though percomorphs likely share the loss of the *HoxCb* cluster, comparison via conserved synteny with zebrafish shows that the orthologous region is on pipefish LG 20.

Syngnathus scovelli dlx gene clusters are missing deeply conserved noncoding elements

The vertebrate *dlx* genes, a family of homeobox transcription factors important for patterning the central nervous system, head skeleton and limbs, are arranged in tandem pairs associated with specific *Hox* clusters. Some percomorphs, like stickleback and pufferfish retain *dlx1/2a*, *dlx3/4a*, *dlx3/4b and dlx5/6a* clusters, while medaka appears to lack a *dlx3/4a* cluster, and zebrafish (a non-percomorph) has lost *dlx3a* but has retained an unpaired *dlx2b* not found in percomorphs (Renz et al. 2011). We found the four typical percomorph clusters, totaling eight genes, in the Gulf pipefish genome and performed a search via mVISTA (Frazer et al. 2004, Mayor et al. 2000) for conserved non-coding elements (CNEs) within the *dlx* clusters by comparing sequences from mammals and other teleosts. We found that pipefish retains some non-coding elements conserved between mammals and teleosts, as well as other CNEs shared only among teleosts (Renz et al. 2011, Ghanem et al. 2003) (Figure 2.5; see Additional File 1, Fig. S3 for VISTA comparisons of the *dlx3/4a*, *dlx3/4b* and *dlx5/6a* clusters at https://doi.org/10.1186/s13059-016-1126-6). For example, we identified pipefish orthologs of two inter-dlx CNEs (Figure 2.5) that were found previously to be conserved between mouse, zebrafish and pufferfish and that were shown to direct reporter gene expression in subsets of *dlx* domains (Ghanem et al. 2003). A third CNE that was not functionally tested but was conserved in both zebrafish and pufferfish (Ghanem et al. 2003) is not preserved in pipefish. We identified two other notable losses in this pipefish cluster: *S. scovelli* has lost an inter-*dlx1/2a* CNE that we find conserved in the other percomorphs, and it also lacks an element in the intron between coding exon 1 and exon 2 of *dlx1a,* a CNE that is conserved in both mammals and other teleosts. There are no

gaps in the assembly in these regions of the pipefish genome. Several other CNEs are missing from other clusters, including two elements on either side of the last exon of dlx4a that are, notably, conserved between other percomorphs such as pufferfish and stickleback and cod, a non-percomorph (Additional File 1, Fig. S3 at https://doi.org/10.1186/s13059-016-1126-6).

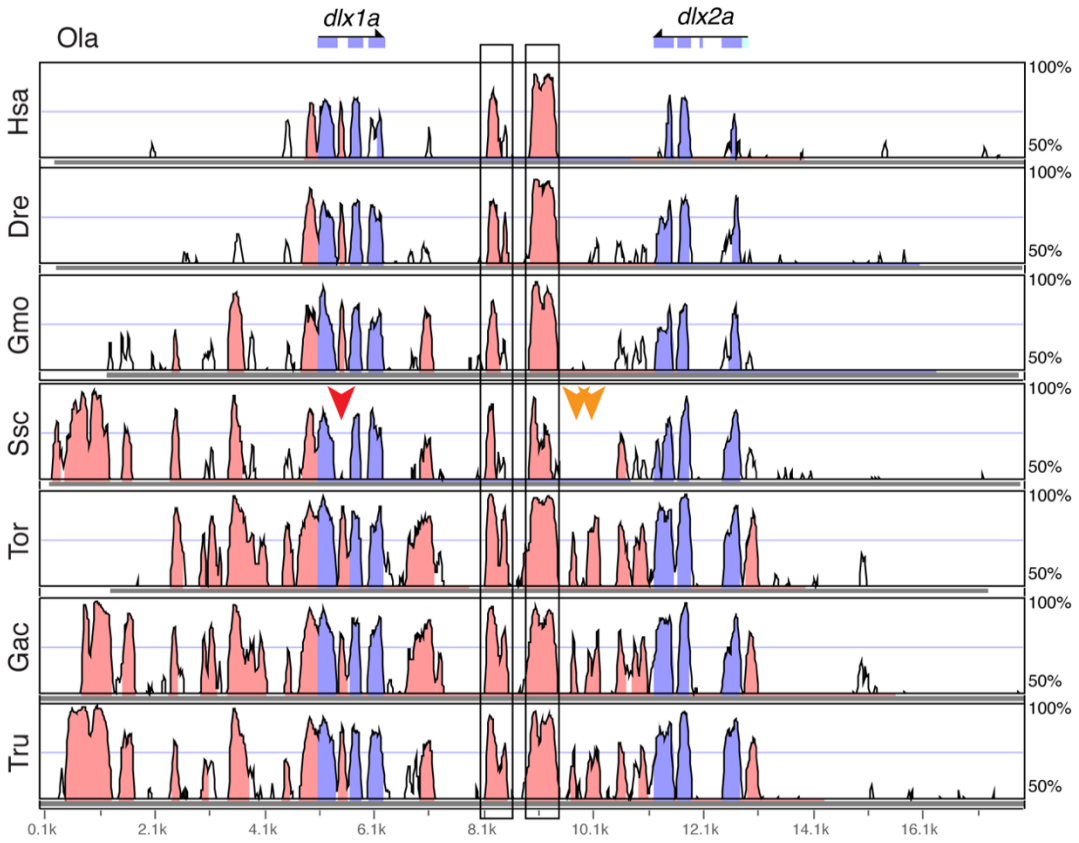


Figure 2.5: Three conserved non-coding elements are not detectable in the pipefish dlx1a-dlx2a cluster. One CNE present in other teleosts and mammals is missing from a gapless region between exon1 and -2 in the *S. scovelli* assembly (red arrow). Two other CNEs in the *dlx* intergenic region that are conserved among percomorphs are also missing from this region in pipefish (orange arrows). Two CNEs previously shown to direct reporter gene expression in murine *Dlx* expression domains are boxed (Ghanem et al. 2003). Exons are highlighted in blue, CNEs in pink. The reference, Ola, is medaka; Hsa, human; Dre, zebrafish; Gmo, cod; Ssc, pipefish; Tor, tuna; Gac, stickleback; Tru, pufferfish.

Syngnathid hindlimb loss implicates modification of the *tbx4-pitx1* pathway

Pipefish, seahorses and seadragons all lack paired pelvic fins. *tbx4*, *pitx1*, *and pitx2* are genes at the top of the regulatory cascade described in vertebrate hindlimb development, including teleosts that have pelvic fins (Marcil 2003, Naiche 2003, Don et al. 2016). We found no trace of the protein-coding sequence for *tbx4* in the pipefish genome assembly. The genomic segments flanking *tbx4* were also not identified, as pipefish orthologs of genes adjacent to *tbx4* in other teleosts were either undetected, as in the case of *tbx2b*, or were on small scaffolds not anchored to the genetic map. TBLASTN also failed to identify *tbx4* among our *de novo* assembled gene transcripts generated from RNA-seq data. Gulf pipefish *pitx1* is present in the assembly but divergent. The predicted pipefish Pitx1 amino acid sequence, supported by transcriptome sequencing, contains homopolymeric expansions of alanine and proline, and an amino acid insertion in the conserved OAR domain not seen in orthologs from other fish lineages or from human (Figure6). A fragment amplified with degenerate PCR primers shows that a second syngnathid species, the messmate pipefish (*Corythoichthys haematopterus*), shares one of the alanine expansions (Figure 2.6). Both Gulf pipefish and human Pitx3, a protein associated more strongly with eye and neural development than limb development (Semina et al. 1998, Shi et al. 2005) also have polyalanine runs in different locations from those found in Pitx1. Pitx2 aligns well with other fish orthologs and apparently contains no homopolymeric expansions.

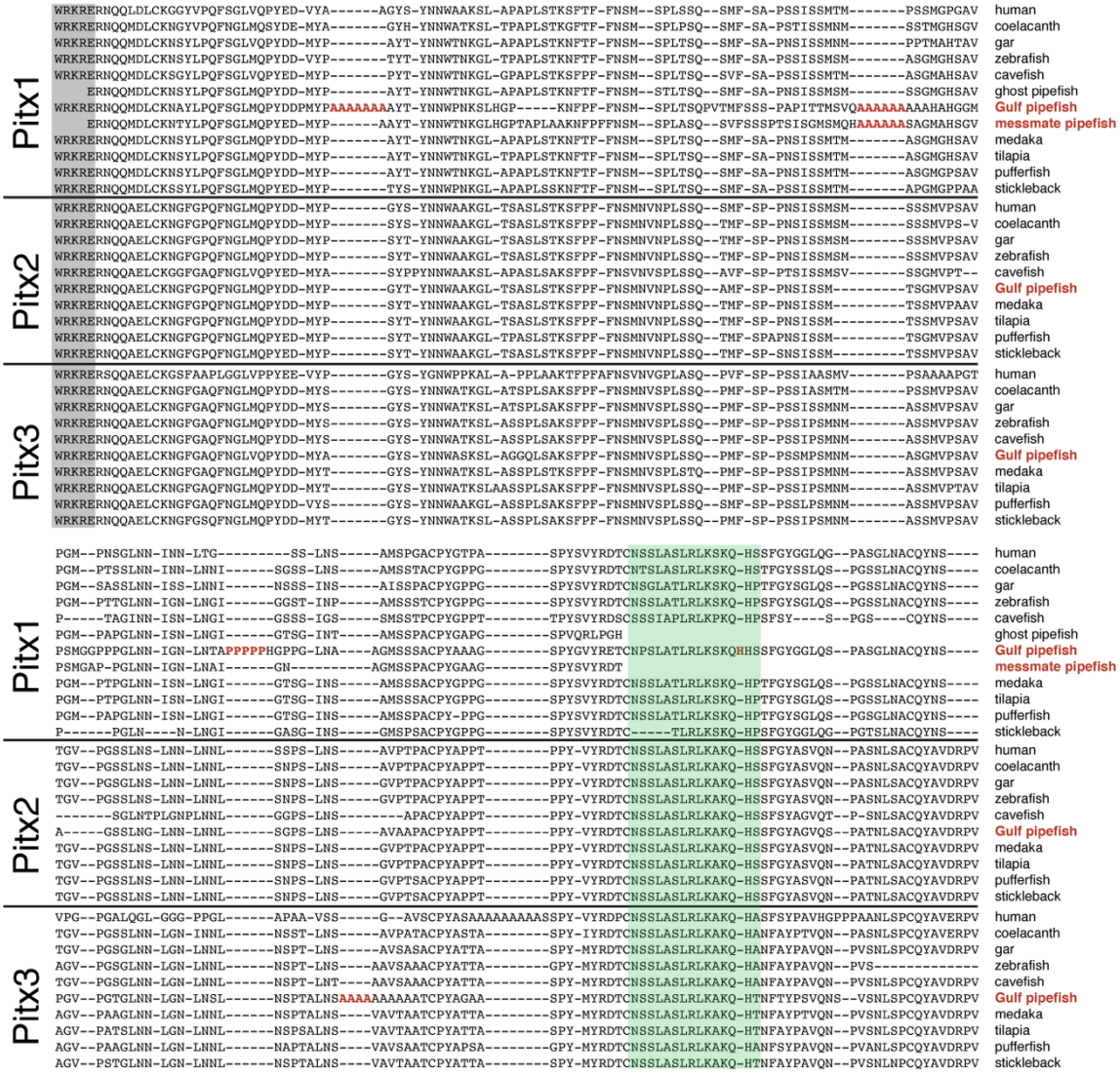


Figure 2.6: Pipefish Pitx1, a vertebrate protein important for hindlimb and tooth development, contains several homopolymeric expansions. Shown are well-aligned regions of Pitx proteins across several vertebrate species, starting from the last 5 amino acids of the homeodomain (shaded gray). Poly-alanine and poly-proline expansions (shown in red) in pipefish Pitx1 and Pitx3 between the homoedomain and the OAR domain (shaded turquoise) are not found in the Pitx proteins of other compared fish; however, there is a poly-alanine expansion at a different location in human Pitx3. One of the Pitx1 polyalanine expansions is shared with the messmate pipefish (*Corythoichthys haematopterus*), a distantly related syngnathid (Wilson et al. 2003) and none are present in the robust ghost pipefish (*Solenostomus cyanopterus*), a member of a close, pelvic-fin-bearing outgroup to the syngnathids (Kawahara et al. 2008, Hamilton et al. 2017). Gulf pipefish also has a single amino acid insertion (also shown in red) in the conserved OAR domain.

Pregnancy-specific gene expression in the brood pouch is widespread and reflects regulation of the innate immune system

We aligned to the annotated genome RNA-seq data from six pregnant male brood pouches (excluding embryonic tissue) and six non-pregnant male pouches. Based on these digital gene expression data, the transcriptional landscape of male brooding tissues differed substantially as a consequence of pregnancy, as 26.19% of the total multivariate dissimilarity among the 12 individual transcriptomes was explained by pregnancy status (Additional File 1, Fig. S4a; perMANOVA: $G1_{,11} = 3.55$, $p = 0.004$ at https://doi.org/10.1186/s13059-016-1126-6). Univariate tests of differential expression between pregnant and non-pregnant males revealed different transcript abundances for 1145 genes of 15,253 genes (FDR=0.1) expressed robustly across at least 4 of 12 individuals. 526 genes were pregnancy-enriched and 619 were pregnancy-depressed, demonstrating fold change differences as extreme as 215 (Tables 2.2 and 2.3; see Additional File 2.2, SH2 for a complete tabulation of differentially expressed genes at https://doi.org/10.1186/s13059-016-1126-6).

We identified several KEGG pathways enriched for genes subject to strong pregnancy-specific expression patterns, including “complement and coagulation cascades”, “cytokine-cytokine receptor interaction,” “calcium signaling” and “neuroactive ligand-receptor interaction” (See Additional File 2.2, SH3 for a full tabulation of KEGG pathways enriched for differentially expressed genes at https://doi.org/10.1186/s13059-016-1126-6). Many pipefish genes within the first two of these pathways, which include innate immune system cascades, were expressed at higher levels in pregnant, relative to non-pregnant pouch tissues. For example, members of the complement membrane attack complex (MAC), which are cell membrane pore-forming

Table 2.2: List of the top 15 pregnancy-enriched pouch tissue genes

Gene ID	Fold change	CPM	P-value	Gene Description	KO ID
SSCG00000006913	15.66	7.22	2.13E-24	WNT1-inducible-signaling pathway protein 2 isoform X2	K06827
SSCG00000005974	21.04	6869.88	1.87E-18	**patristacin, partial**	K08778
SSCG00000007802	4.15	93.44	7.69E-16	podocan	
SSCG00000014514	3.15	46.38	1.45E-15	fos-related antigen 2-like	
SSCG00000015977	12.38	229.24	1.39E-14	myocilin-like	
SSCG00000006209	6.53	4.72	4.91E-14	dickkopf-related protein 2	K02165
SSCG00000007875	2.93	188.72	8.81E-14	neuroepithelial cell-transforming gene 1 protein	
SSCG00000013720	5.13	233.89	3.85E-13	lipopolysaccharide-binding protein/bactericidal permeability-increasing protein	
SSCG00000011252	2.88	72.11	2.72E-12	beta-galactoside alpha-2,6-sialyltransferase 1-like isoform X1	K00778
SSCG00000004944	6.64	29.73	7.33E-12	collagen alpha-2(VI) chain-like	K06238
SSCG00000006480	3.10	18.93	1.81E-11	CTTNBP2 N-terminal-like protein	
SSCG00000013244	2.30	34.04	2.10E-11	LIM domain transcription factor LMO4-B-like	
SSCG00000004636	3.22	386.88	3.62E-11	NA	
SSCG00000002072	29.24	1.59	3.77E-11	potassium channel subfamily K member 2-like	K04913
SSCG00000007792	5.21	7.06	4.20E-11	excitatory amino acid transporter 5-like	K05618

Included are the fold change (pregnant/non-pregnant), average expression level across 12 pouch libraries in copies per million (cpm), edgeR negative binomial exact test p-value, gene description from top BLASTP hit, and the assigned KEGG orthology ID for each pipefish gene. See Supp. Spreadsheet Preg DE Genes for the full list.

Table 2.3: List of the top 15 pregnancy-depressed pouch tissue genes

Gene ID	Fold change	CPM	P-value	Gene Description	KO ID
SSCG00000006879	27.36	56.49	7.91E-43	Serine/threonine-protein kinase WNK2	K08867
SSCG00000018539	12.37	15.96	2.04E-26	FXYD domain-containing ion transport regulator 12	
SSCG00000007973	4.73	53.34	1.66E-24	A disintegrin and metalloproteinase with thrombospondin motifs 6, partial	K08621
SSCG00000013585	10.78	19.10	1.07E-23	Tetratricopeptide repeat protein 18	
SSCG00000005985	214.58	652.27	7.29E-23	**patristacin, partial**	K08076
SSCG00000008728	14.12	6.03	2.22E-22	Uridine-cytidine kinase-like 1	K00876
SSCG00000000969	4.32	19.82	1.25E-17	ras-like protein family member 11A	K07852
SSCG00000017729	6.14	359.52	1.71E-17	nidogen-2-like isoform X5	K06826
SSCG00000004506	6.00	12.98	4.08E-17	syntaxin-2-like isoform X1	K08486
SSCG00000010275	14.47	3.28	1.00E-16	acid-sensing ion channel 1	
SSCG00000016046	6.75	8.51	1.51E-16	leucine-rich repeat-containing protein 4-like	K16351
SSCG00000014649	10.15	7.67	1.77E-16	homeobox protein MSX-2-like	K09341
SSCG00000019217	66.66	3.26	1.82E-16	leucine-rich repeat-containing protein 3-like	
SSCG00000007661	5.19	24.20	2.23E-16	cytochrome P450 27C1-like	K17951
SSCG00000005388	19.81	1.44	5.60E-16	glutamate receptor ionotropic, delta-2 isoform X5	K05207

Included are the fold change (non-pregnant/pregnant), average expression level across 12 pouch libraries in copies per million (cpm), edgeR negative binomial exact test p-value, gene description from top BLASTP hit, and the assigned KEGG orthology ID for each pipefish gene. See Supp. Spreadsheet Preg DE Genes for the full list.

toxins (Humphrey and Dourmashkin 1969) (reviewed in (McCormack et al. 2013)), tended to be expressed at higher levels in pregnant males (Additional File 1, Fig. S5a, Additional File 1, Fig. S6a at https://doi.org/10.1186/s13059-016-1126-6). Pro-inflammatory chemokines IL8, CXCL9, CXCL10, and CXCL12 of the CXC subfamily were also expressed at higher levels in pregnant males, as were several members of the CC subfamily (Additional File 1, Fig. S5b at https://doi.org/10.1186/s13059-016-1126-6). Not all transcriptional signatures of the immune system reflected this pattern, however. A suite of genes belonging to the natural killer cell cytotoxicity response pathway, for example, was expressed at higher levels in non-pregnant males (Additional File 1, Fig. S4d at https://doi.org/10.1186/s13059-016-1126-6). Furthermore, genes in KEGG pathways associated with the adaptive immune system, including "antigen processing and presentation", "T cell receptor signaling pathway," and "B cell receptor signaling pathway," were transcriptionally less sensitive to pregnancy status than those in innate immunity KEGG pathways (Additional File 1, Fig. S6b at https://doi.org/10.1186/s13059-016-1126-6). Consistent with a characterization of the immune gene repertoire in *Syngnathus typhle* (Haase et al. 2013), we failed to detect MHC class II alpha and beta chain genes in the genome of *S. scovelli*, so the potential for some functionality of the adaptive immune system in this pipefish genus may be limited in general.

Gene Ontology terms overrepresented among pregnancy-enriched genes included those related to the complement system, coagulation, and immunity, consistent with the KEGG analysis, but we also identified terms related to hemopoiesis, homeostasis, proteolysis, and others (Additional File 2, SH5 at https://doi.org/10.1186/s13059-016-1126-6). GO terms overrepresented among pregnancy-depressed genes included those

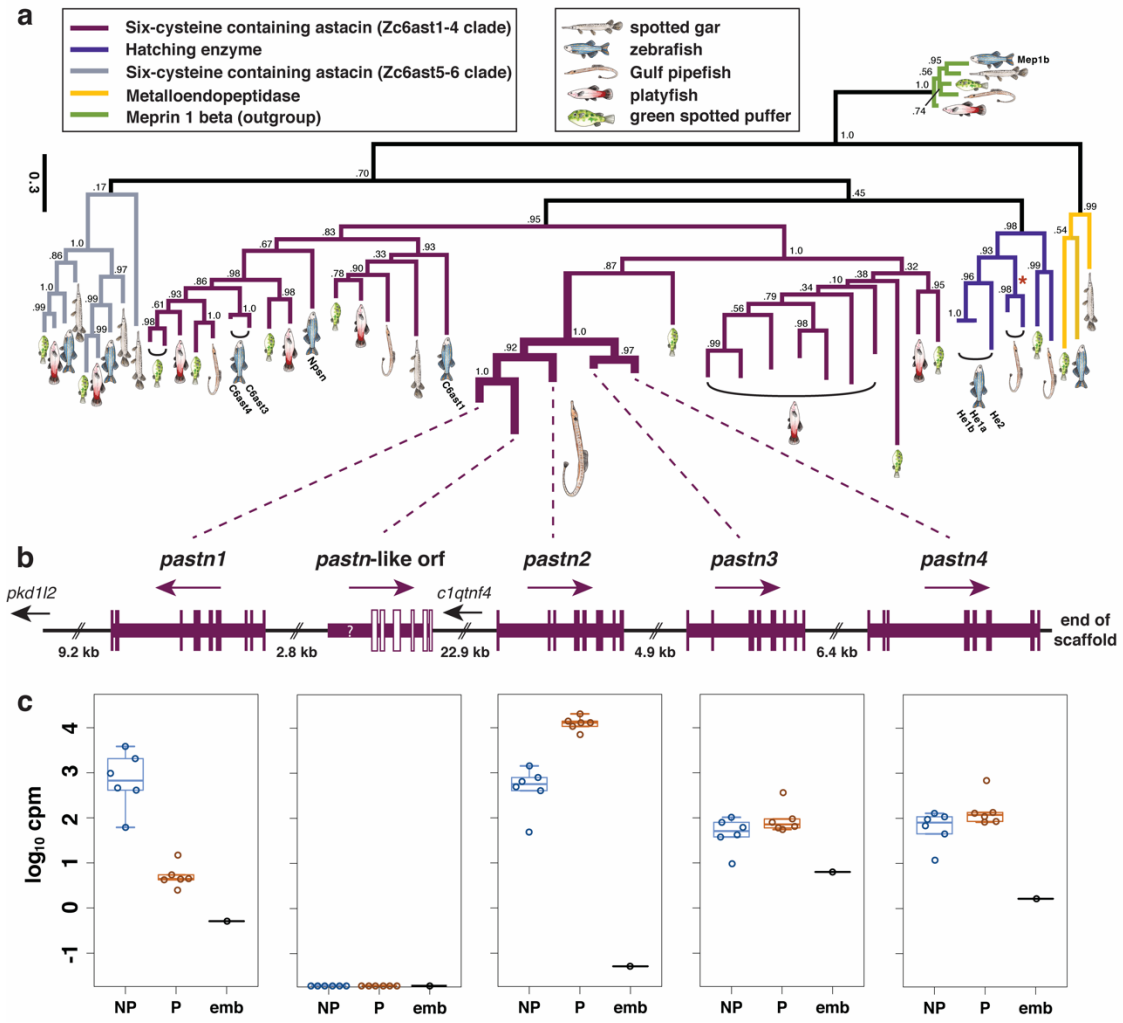
related to developmental processes, cell-to-extracellular matrix (ECM) adhesion, and protein glycosylation (Additional File 2, SH6 at https://doi.org/10.1186/s13059-016-1126-6).

Lineage-specific duplication of *patristacins* associated with male pregnancy

As documented previously in *S. scovelli* and *S. floridae* (Small, Harlin-Cognato, and Jones 2013), two similar astacin-like metalloproteases, demonstrated strikingly opposite patterns of gene expression: one markedly pregnancy-enriched and the other highly pregnancy-depressed (Table 2.2, Table 2.3, Figure 2.7B-C). We here find that

---

Figure 2.7: (next page) Gene duplication of *patristacins* preceded the evolution of diverse expression patterns related to male pregnancy. Patristacins are unique, tandemly arrayed C6 astacin-like metalloprotease genes presumably co-opted during the evolution of male pregnancy (Harlin-Cognato, Hoffman, and Jones 2006). a) A maximum likelihood gene tree inferred from astacin-like metalloprotease amino acid sequences, representing five fish genomes, is rooted assuming Meprin1b proteins as an outgroup. Different protein subfamily clades (colored by clade and including terminology from Kawaguchi et al. (2006)) mostly correspond to conserved syntenic regions. Clade support values are SH-aLRT, but see Fig. S8 (in Additional File 1 at https://doi.org/10.1186/s13059-016-1126-6) for bootstrap values and tip accession numbers. Zebrafish sequences with annotated Ensembl gene names are labeled for reference. Patristacins comprise a monophyletic group nested within the Zc6ast1-4 clade, suggesting pipefish or syngnathid lineage-specific duplication events. Note the absence of pipefish orthologs from the Zc6ast5-6 clade (colored gray). In medaka, orthologs from this group are expressed exclusively in the developing jaw (Kawaguchi et al. 2006). Also note the red asterisk in the hatching enzyme clade, which corresponds to intron loss in the pipefish lineage. b) The physical arrangement of *patristacins* in the Gulf pipefish genome, with two other genes in the region (small text). Arrows indicate the direction of the sense strand, and vertical bars reflect coding exons. Note that the status of *pastn*-like orf as a gene is uncertain, so it is depicted by open bars and a question mark where 3 missing exons would normally be. c) *Patristacin* expression levels from RNA-seq data for six non-pregnant male brood pouch samples (blue), six pregnant samples not including embryos (orange), and a pooled embryo library (black). Y-axis values are copies per million (cpm) on a log scale. Individual data points and boxplots are shown. Note the extreme expression differences between *pastn1* and *pastn2*.



these "*patristacins*" (Harlin-Cognato, Hoffman, and Jones 2006) are adjacent genes belonging to a small cluster of duplicates that includes two additional *patristacins* expressed at lower levels in the brooding tissues at the stages examined (Figure 2.7B-C). This cluster, located on scaffold 62 of pipefish LG4 also included a fifth, partial coding sequence for which we could identify neither a likely start methionine nor the first three typical *patristacin* exons. A phylogenetic analysis including astacin-like metalloprotease sequences from global searches of five ray-finned fish genomes suggests that the

patristacin cluster is a gene family expansion unique to the lineage leading to syngnathids (Figure 2.7A). We found protein-coding genes from platyfish and green spotted puffer genomes that share a recent common ancestor with *patristacins*, but these sequences were not nested within the *patristacin* subclade. Furthermore, *patristacins* and their closest homologs most likely diverged via gene duplication from the subfamily of 6-cysteine astacins that includes zebrafish *nephrosin*, given the topology of our current gene tree and that all paralogs share the same genomic region on pipefish LG4.

DISCUSSION

Despite the explosive teleost species radiation over the last 300 million years, these fishes have been conservative in karyotype evolution relative even to the much younger mammalian lineage, with the majority of teleost species having a haploid number of 24 or 25 (Naruse et al. 2004). Variations from the inferred ancestral number of 24 (Mank and Avise 2006) do exist across the teleost radiation, stemming from chromosome duplications, fissions, and fusions. We have shown that two chromosomal fusions in an ancestor of *Syngnathus scovelli* have likely led to a haploid karyotype of 22 (Figure 2.2A, B). Comparisons of sequenced genomes suggest that interchromosomal rearrangements (translocations) are relatively uncommon in teleosts (Naruse et al. 2004), and this is reflected in the striking one to one correspondence of chromosomes across most of the genome between Gulf pipefish and other percomorphs, such as southern platyfish (Figure 2.2A). The stability of teleost genomes simplifies comparisons, and increases confidence in correctly determining orthology of genes and chromosome segments based on observed patterns of conserved synteny. We have exploited the

exceptional conservation of synteny among sequenced teleosts to explore the evolution and behavior of genes that might play a role in syngnathid innovations.

The remarkable morphology of syngnathids was noted in "The History of Animals" by Aristotle, who construed the peculiar phenomenon of pipefish live birth as a splitting open of the body. Prior to our characterization of the Gulf pipefish genome, however, with the exception of a few transcriptomic resources (Haase et al. 2013, Small, Harlin-Cognato, and Jones 2013, Whittington et al. 2015), virtually no information existed for how key developmental genes and their modification might be responsible for derived syngnathid phenotypes. Now, with the availability of the genome of *Syngnathus scovelli*, and likely other related genomes soon to follow, we expect researchers interested in the developmental genetic underpinnings of novel vertebrate morphologies to make the critical experimental connections between genomic differences in syngnathids and their functional consequences. In anticipation of exciting functional genomics work enabled by the latest genome editing approaches (Boettcher and McManus 2015, Sternberg and Doudna 2015), here we highlight a few especially promising examples of molecular signatures with implications for hallmark traits of pipefishes, seahorses, and their relatives.

We explored the constitution of the syngnathid *Hox* genes because these Vertebrate *Hox* clusters are tandem arrays of transcription factor genes with many developmental roles, including segmental identity in the axis and in limb morphogenesis (reviewed in (Zakany and Duboule 2007, Alexander, Nolte, and Krumlauf 2009)). Our investigation of Gulf pipefish *Hox* cluster content revealed that the evolution of an elongated, ribless body was not accompanied by drastic reorganization of the *Hox* genes. While there are multiple losses of pipefish *Hox* genes and the *Hox*-regulating microRNA

genes that are interspersed among them, many of these same genes have been lost from other percomorphs that have less modified skeletons (Figure 2.4).

Two gene losses from the Gulf pipefish *Hox* clusters stand out, however. The loss of *eve1* is unique among described teleost *Hox* clusters. This gene belongs to the *evenskipped* (*evx*) gene family, whose members reside at the ends of particular clusters. In zebrafish embryogenesis, the *HoxBa* cluster-associated *eve1* gene is expressed during gastrulation and in the extending tail tip; its knockdown suppresses trunk and tail development, prompting the experimentalists to suggest *eve1* acts as a posterior organizer (Cruz et al. 2010) (but see (Seebald and Szeto 2011) for another interpretation). It is therefore remarkable that *eve1* is deleted in pipefish (Figure 2.4). It is possible that some of these early ontogenetic functions of *eve1* have been distributed to the remaining two pipefish *evx* genes or otherwise compensated for. However, syngnathids have neither oral nor pharyngeal teeth, consistent with evolutionary loss of *eve1*, the only reported *evx* gene that is expressed during teleost tooth development (Laurenti et al. 2004, Debiais-Thibaud et al. 2007). In addition, it appears that pufferfish and pipefish lineages have independently lost all copies of *hox7*, a paralogy group that when experimentally knocked out in mouse causes reduction and mispatterning of ribs (Chen, Greer, and Capecchi 1998); consistent with this biological role for *hox7*, both pufferfish and pipefish lack ribs.

A uniting trait of the Syngnathidae is an absence of pelvic fins. Two other percomorphs that have evolutionarily lost pelvic fins appear to have done so by alteration of a hindlimb-positioning *hoxd9a* expression boundary (pufferfish (Tanaka et al. 2005)) or by loss of *pitx1* expression in the developing hindlimb (freshwater threespine stickleback (Shapiro et al. 2004, Chan et al. 2010)). Pitx1, a transcription factor, directly

activates initial expression of *tbx4* in the hindlimb primordium (Logan and Tabin 1999), and *tbx4* is required for initial limb bud outgrowth (Naiche and Papaioannou 2007). We found that pipefish *pitx1* has an amino acid insertion in the OAR, a functional domain thought to modulate DNA binding (Brouwer et al. 2003), and unusual homopolymeric alanine and proline repeat expansions between the homeodomain and OAR (Figure 2.6). Homopolymers are known to cause several developmental diseases in humans (reviewed in (Brown and Brown 2004)) and to affect subcellular localization, protein-protein interaction and transcriptional regulation (Galant and Carroll 2002, Oma et al. 2004). In particular, expansions of alanine and proline homopolymers within transcription factors can modulate the proteins' ability to regulate transcription of gene targets. A distantly related pipefish species, the messmate pipefish, shares one of the homopolymeric repeats (Figure 2.6), suggesting that this divergence of *pitx1* began early in the syngnathid lineage. It is conceivable that changes in the amino acid sequence of syngnathid Pitx1 have had functional consequences for the protein's interaction with its gene targets (such as *tbx4*), affecting hindlimb development. We found no pipefish ortholog of *tbx4*. Failure to find pipefish *tbx4* in the genome assembly does not necessarily mean the gene has been evolutionarily lost; however, the possible loss of this gene with an apparently narrow developmental role in teleosts—in hindlimb development(Don et al. 2016)—is consistent with the evolutionary loss of the hindlimb itself in syngnathids. Loss of the pelvic fins in a syngnathid ancestor may have occurred shortly before or after the origin of the lineage, because the closest extant relatives—the ghost pipefishes (Family Solenostomidae) (Kawahara et al. 2008, Hamilton et al. 2017)—have large, clasping pelvic fins in which females brood the embryos (Playfair and Günther 1866).

Interestingly, Pitx1 in robust ghost pipefish (*Solenostomus cyanopterus*) lacks the homopolymeric repeats described above (Figure 2.6).

A family of homeodomain transcription factors important for limb, brain, and craniofacial development, the *Dlx* genes, are arranged in gene pairs associated with specific *Hox* clusters. Within and near the *Dlx* gene pairs are conserved non-coding elements (CNEs) recognizable by alignment among sequences from even distantly related vertebrates. Several teleost *Dlx* clusters, for example, have CNEs in common with mammals (Renz et al. 2011, MacDonald et al. 2010). Putatively these CNEs are preserved because they have a function, perhaps in regulating gene expression of the *dlx* genes themselves. For instance, two CNEs that fall between *dlx1* and *dlx2* and that are conserved between teleosts and mammals direct reporter gene expression in the developing forebrain and first and second pharyngeal arches in murine (Ghanem et al. 2003) and in zebrafish (MacDonald et al. 2010) embryos. We found that pipefish has retained these two ancient CNEs but has apparently lost a third element that is as deeply conserved (i.e., between mammals and teleosts), from within an intron of *dlx1a.* In addition, at least two more CNEs in the intergenic region of *dlx1/2a* that are conserved among other percomorphs are lost or diverged beyond recognition in pipefish (Figure 2.5). Experimental mutation of mouse *Dlx1/2* genes creates defects in the development of pharyngeal arch derivatives, such as the mandible and teeth (Qiu et al. 1995). Knockdown of these genes in zebrafish causes embryos with shortened faces and mispatterning of first and second arch cartilages and a reduced ethmoid (a cartilage of the ventral neurocranium) (Sperber et al. 2008). In addition, *dlx2* genes are expressed in developing teeth in cichlids, catfish, and cyprinids (Jackman, Draper, and Stock 2004, Stock, Jackman, and Trapani 2006, Fraser et al. 2009), and *dlx2a* is expressed in

migrating neural crest that will form the anterior pharyngeal arch cartilages (Sperber et al. 2008, Akimenko et al. 1994). Pipefish embryos show modified development of the anterior skull including cartilage derivatives of the first and second pharyngeal arches, particularly elongation of the hyosymplectic (a cartilage of the second arch), as well as unusual early curvature and later elongation of the ethmoid cartilage (see Additional File 1, Fig. S7, for a view of pipefish craniofacial development at https://doi.org/10.1186/s13059-016-1126-6), implicating changes in expression of early acting genes such as *dlx2a*, involved in cranial neural crest survival and patterning. Functional testing in other teleosts could reveal whether the CNEs here shown to be erased in pipefish are functional units that modulate expression of the *dlx1/2a* cluster genes and possibly affect pharyngeal arch or tooth development.

Male pregnancy in syngnathid fishes is a true example of evolutionary novelty. In many lineages, including *S. scovelli*, males gestate developing embryos in a tightly regulated environment defined by a complex brood pouch. Extensive cellular and developmental changes in the pouch occur leading up to and during pregnancy, including proliferation of epithelial cells, development of specialized secretory cells, and angiogenesis (Carcupino 2002, Watanabe, Kaneko, and Watanabe 1999, Laksanawimol, Damrongphol, and Kruatrachue 2006). These specializations are likely the consequence of adaptation, as they enable functions directly relevant to fitness, including solute, gas, and nutrient delivery to a male's brood (Ripley 2009, Ripley and Foran 2009, Goncalves, Ahnesjo, and Kvarnemo 2015), as well as immune priming of offspring (Roth et al. 2012). Consistent with this functional diversity, our genome-based analysis of male pregnancy in *S. scovelli* revealed a transcriptionally rich brood pouch in which over 73% of annotated genes were expressed robustly, and over 1000 were differentially expressed

as a consequence of pregnancy (Additional File 2.2, SH2 at https://doi.org/10.1186/s13059-016-1126-6). Previous studies, based on *de novo* transcriptome assemblies, characterized pregnancy-specific gene expression in pipefish species of *Syngnathus* (Small, Harlin-Cognato, and Jones 2013) and in the seahorse *Hippocampus abdominalis* (Whittington et al. 2015), but lack of a reference genome in those surveys limited insights into the transcriptional breadth of the pouch and single gene resolution for transcript abundance measurements. Our differential expression analysis comparing early-stage pregnant to non-pregnant male pouch tissue echoes many of the patterns described in the comprehensive seahorse study (Whittington et al. 2015), including evidence for positive regulation of developmental processes, lipid transport, homeostasis, and the immune system during pregnancy. Interestingly, we noted a more pronounced signature of pregnancy-specific gene expression for innate, relative to adaptive, immune pathways in Gulf pipefish (Additional File 1, Fig. S6 at https://doi.org/10.1186/s13059-016-1126-6). This observation is likely in part a consequence of pipefishes in *Syngnathus* having lost important genetic components of MHC class II mediated immunity (Haase et al. 2013), although MHC class I components remain intact. Syngnathid fathers face unique demands with respect to immunity and pregnancy, given that the brood pouch is a non-urogenital organ more directly exposed to the environment than internal uterine structures of other vertebrates. A seemingly difficult balance among pathogen control, maintenance of beneficial microbes, and mitigation of attack against non-self (embryonic) tissues must therefore be struck. Although future work regarding the details of this balance will be required to say so, perhaps a uniquely fine-tuned division of labor between innate and adaptive immunity

has been an evolutionary outcome of male pregnancy, a balance we hypothesize differs across syngnathid lineages with varying brood pouch complexity.

The significance of gene duplication to adaptation and biological diversification in general is continually of interest to evolutionary biologists (Ohno 1970, Force et al. 1999, Lan and Pritchard 2016). We identified at least four clustered members of the *patristacin* gene subfamily on a single scaffold of LG4 in the Gulf pipefish genome (Figure 2.7). Given the striking patterns of gene expression for *pastn1* and *pastn2* with respect to pregnancy, it is possible that gene duplication followed by neo- or subfunctionalization played a key role in the evolution of male pregnancy, although surveys of other syngnathid genomes and those of their closest relatives are needed to test this hypothesis. Our interpretation of the evolution of *patristacins* is distinct from that of Harlin-Cognato et al. (2006), who suggested that one *patristacin*, identified without the advantage of a complete *S. scovelli* genome, took on a novel role in male pregnancy by a spatiotemporal shift in gene expression, and not via gene duplication. Our genome-wide approach has provided additional information, however, by revealing the complete coding sequence for multiple *patristacin* paralogs in *S. scovelli*. Because the two *patristacins* with exceptional pregnancy-specific gene expression (*pastn1* and *pastn2*) likely diverged by gene duplication after pipefish separated from the other fish lineages in our comparison, we provide evidence for a role of relatively recent gene duplication in *patristacin* evolution. Our phylogenetic analysis highlights a second, large expansion of *patristacin*-like genes in the genome of *Xiphophorus maculatus*, suggestive of high duplicate retention in multiple live-bearing fish lineages.

The specific functional roles *patristacins* play in male pregnancy are currently unknown, but our current phylogenetic understanding of their place among teleost

Astacin-like metalloproteases suggests that they may be more functionally similar to Nephrosin-like proteins than hatching enzyme components (Fig. 7A, Additional File 1, Fig. S8 at https://doi.org/10.1186/s13059-016-1126-6). Kawaguchi et al. (2006) showed, for example, that medaka 6-cysteine astacin genes *mc6ast1* and *mc6ast2*, orthologs of zebrafish *c6ast1* and zebrafish *c6ast3/4*, respectively, were expressed in a wide range of tissues, in contrast to medaka hatching enzymes, which were expressed exclusively in pre-hatching embryos. Another member of this gene subclade, *cimp1*, is expressed epithelially in the developing cichlid jaw and may play a role in extracellular matrix (ECM) turnover during development (Kijimoto et al. 2005). We hypothesize that *patristacins* evolved from an already transcriptionally promiscuous ancestor and now, following subsequent duplication events, work in concert to regulate the remodeling of the pouch epithelium necessary for the sustenance of pregnancy. Our characterization here of their structural organization and expression patterns in the brood pouch will inform and facilitate future functional studies of these gene duplicates and their specific roles in male pregnancy.

CONCLUSIONS

We present the first annotated reference genome assembly, organized into chromosomes, for a syngnathid fish. Our comparisons of the Gulf pipefish genome to other fish genomes reveal two chromosomal fusions in the syngnathid lineage. We provide additional evidence suggesting that syngnathiform fishes are an outgroup relative to fellow percomorph fishes commonly used in comparative genomics studies. The Gulf pipefish genome will therefore serve as a useful comparator in studies that aim to understand rates of genome evolution among percomorphs for which there are existing

genomic resources. We show that losses of both genes and conserved non-coding elements have occurred in pipefish gene families important for vertebrate craniofacial, tooth, hind limb, and axial development, all features that are highly modified in syngnathids. In addition, we detail aspects of the molecular biology of male pregnancy, a unique and unifying feature of the pipefish, seahorses and seadragons; in particular, we exploited the annotated Gulf pipefish genome and transcriptional profiling to show how pregnancy is associated with clear changes in gene expression in the male brood pouch tissue, a broad example being regulation of the innate immune system, and a specific example being regulation of duplicated *patristacins*.

BRIDGE

Chapter II consists of the published Gulf pipefish (*Syngnathus* scovelli) genome paper, with the specific parts I contributed to outlined in detail in the abstract. I contributed significantly to the production of this annotated reference genome from the family of Syngnathidae. It was a crucial resource to develop in order to accomplish my subsequent dissertation research. A subset of my *Hox* gene dissertation research—restricted to only presenting the coding genes and microRNA contents of the *Hox* cluster—were included in that chapter. I described the genomic organization of *Hox* clusters in a species of syngnathid pipefish—the Gulf pipefish (*Syngnathus scovelli*). I assessed the phylogenetic placement of syngnathid fish relative to other representative fish taxa using ultraconserved elements and I compared the *Hox* cluster gene content of the Gulf pipefish against other teleost fish species. I found that the *Hox* gene content has remained largely conserved in the Gulf pipefish relative to other teleost fish with annotated *Hox* clusters with a few key losses.

In Chapter III, I document the outcome of functional genomic studies performed to determine possible effects on the evolution of the syngnathid body plan of the loss of *hox7* genes. In this chapter I describe creating mutations in these orthologous genes in the threespine stickleback fish (*Gasterosteus aculeatus*) using the CRISPR/Cas9 system. Similar genetic manipulations of syngnathids using CRISPR/Cas9 is not possible. Therefore, I decided to test the hypotheses that stickleback could lose one or the other copy and survive because of genetic redundancy, and that the loss of both copies would result in phenotypic effects in the axial skeleton that mirror syngnathids. I successfully established transgenic lines for the *hox7* gene knockouts and I describe some preliminary results that indicate the possible role for *hox7* genes in rib and vertebrae development.

CHAPTER III

A SURVEY OF AXIAL PHENOTYPIC EFFECTS INDICATES GENETIC REDUNDANCY IN TELEOST *HOX7* GENES

INTRODUCTION

Modifications to the axial skeleton pattern accounts for a significant amount of body plan diversity seen among vertebrates (Carroll 1988, Gadow 1933). Axial skeletal diversity can be achieved through global or regional addition or subtraction of elements and modification to the size and shape of these elements (Ward and Brainerd 2007). On one end of the spectrum, snakes often have over 300 vertebrate that results from an expansion of the rib-bearing thoracic vertebrae, and have lost their forelimbs and have completely lost or have highly reduced hindlimbs (Cohn and Tickle 1999). At the other end, frogs have evolved extreme truncation of their vertebral column, including a loss of their caudal vertebrae, and have elongated hindlimbs (reviewed by (Handrigan and Wassersug 2007)).

Hox genes—since their initial description in *Drosophilia*—have been known for their ability to cause homeotic transformations to the body plan (Lewis 1963, 1978). These core developmental genes code for homeodomain transcription factors that are responsible for helping to determine the body plan of an embryo by specifying positional information along the anterior-posterior axis. *Hox* genes are expressed early in development, and each comprise of two exons and one intron. The gene includes several protein and DNA binding domains, including a 183 base pair homeobox DNA sequence. The homeobox encodes a protein domain that is referred to as the homeodomain and binds to DNA sequences of regulatory elements of often a very large number of genes

(Gehring, Affolter, and Bürglin 1994, Gehring et al. 1990). For this reason, *Hox* genes were one of the first discovered and described of the so-called 'master regulatory genes.'

Many years of research on *Hox* gene sequence, function, and evolution has provided a much deeper understanding of metazoan developmental genetics. Several surprising findings have emerged from this work. First, *Hox* genes are very conserved in terms of sequence, genome organization and function throughout vertebrate evolution. That observation, combined with studies involving experimental perturbations of these genes in the lab, have demonstrated that these homeotic transcription factors have an important role in patterning the vertebrate axial body plan early in development. Although *Hox* mutations affect a variety of cell types including neural tissue, neural crest, endodermal derivatives and mesodermal derivatives, they are most notably documented to affect derivatives of the segmented paraxial mesoderm which lead to axial phenotypes (Krumlauf 1993, Mallo, Vinagre, and Carapuco 2009, Manley and Capecchi 1998, Trainor and Krumlauf 2000, Wellik 2009, Wellik, Hawkes, and Capecchi 2002, Iimura, Denans, and Pourquie 2009).

The spatial organization of *Hox* genes in genomes is also highly conserved, arguing that the correct spatial and temporal patterns of expression depend to some extent on regulatory elements in intervening DNA sequence between the genes. In vertebrates, *Hox* genes are organized into 13 paralogous groups that are arranged into gene clusters (Scott 1992). The ancestral set of *Hox* genes consisted of a single cluster of genes, resulting from tandem duplications of an ancestral proto-*Hox* gene (Garcia-Fernandez 2005). Due to subsequent rounds of whole genome duplications, vertebrates have duplicate copies of the *Hox* complex (Pascual-Anaya et al. 2013). In vertebrates, tetrapods have four *Hox* gene clusters (denoted as *Hox* clusters A, B, C, and D), while

teleost fish have eight clusters of *Hox* genes due to the whole teleost genome duplication (*Hox* clusters Aa, Ab, Ba, Bb, Ca, Cb, Da, Db) (Amores et al. 1998).

Earlier studies examining expression patterns of *Hox* genes noted that genes in the same paralogous groups have overlapping expression along the axis. From these early expression studies, the idea of *Hox* gene collinearity was established. This means that the order they appear in the genome reflects the order they are expressed along the anterior-posterior body axis (Gaunt 1988, Graham, Papalopulu, and Krumlauf 1989, Peterson et al. 1994, Dekker et al. 1993, Godsave et al. 1994, Duboule and Dollé 1989), with the *Hox3* to *Hox11* genes causing defects in the axial skeleton (reviewed in (Wellik 2009)). Subsequent experiments using gain-of- and loss-of-function experiments further demonstrated that *Hox* genes in the same paralogous groups have redundant functions—where knocking out all members of a single paralog groups would confer a stronger phenotype that knocking out a single member of a paralogous group (Chen and Capecchi 1997, 1999, Chen, Greer, and Capecchi 1998, Condie and Capecchi 1994, Fromental-Ramain et al. 1996, Gavalas et al. 1998, Horan et al. 1995, Manley and Capecchi 1998, McIntyre et al. 2007, Studer et al. 1998, van den Akker et al. 2001, Wahba, Hostikka, and Carpenter 2001, Wellik and Capecchi 2003, Wellik, Hawkes, and Capecchi 2002).

Although there is a general pattern of conservation of *Hox* genes across metazoans in general, and across vertebrates in particular, the content and function of *Hox* genes has been demonstrated to vary (i.e. (Cohn and Tickle 1999, Tanaka et al. 2005, Smith et al. 2016)). Similarly, vertebrates exhibit a wide range of morphological diversity, including in traits that are likely affected by *Hox* gene expression early in development. Both of these beg the question about whether—and if so, to what extent—variation in *Hox* gene cluster content can be linked to macroevolutionary patterns in

morphological evolution. We asked this question through developmental genetic study of the highly derived family of fish Syngnathidae, comprising seahorses, pipefish and seadragons.

We previously described that the Gulf pipefish (*Syngnathus scovelli*) has lost all copies of their *hox7* genes (Small et al. 2016). The nearly simultaneous publication of the genomes of two seahorse lineages—*Hippocampus erectus* and *H. comes*—allowed us to confirm that the loss of all *hox7* genes is common across a large proportion of the syngnathid family and is therefore likely a synapomorphy for this clade. This loss of *hox7* genes raised several more questions to address. What was the phenotypic consequence could have resulted from the loss of the *hox7* genes? Is there any aspect of the divergent syngnathid morphology that could be linked to this homeotic gene loss?

Findings from previous functional studies in model organisms provided a plausible link between evolution of *hox7* gene loss and morphological evolution in syngnathids. The functional role of *hox7* genes has only been tested in mice, indicating that *hox7* genes can affect several aspects of vertebral element identity, including rib development. A previous study found that various combinations of knockouts of *Hoxb7* and *Hoxa7* in mice caused defects in rib morphology including decreased sternebra number, decreased rib number, and rib fusion (Chen, Greer, and Capecchi 1998). To date, the complete loss of *hox7* genes have only been reported in syngnathid fish, pufferfish, and the dwarf minnow *Paedocypris* genus (Amores et al. 2004, Small et al. 2016, Malmstrom et al. 2018, Lin et al. 2016, Lin et al. 2017). Interestingly, consistent with this biological role for *hox7* in mice, both pufferfish and pipefish lack ribs (Figure 3.1). In addition, the dwarf minnow *Paedocypris* genus have reduced, poorly ossified ribs (Britz and Conway 2009). This correlation between *hox7* gene and rib loss is not perfect,

however. The ocean sunfish *Mola mola*, which is in the same order Tetraodontiformes as pufferfish, have lost their ribs but have retained their *hox7* genes—making their true teleost body plan function unclear (Pan et al. 2016). Together these data argue for direct tests of *hox7* function in teleosts, particularly testing the hypothesis that *hox7* genes have redundant functions in teleosts, and the phenotypic effects of *hox7* mutations are only seen when all copies are lost.

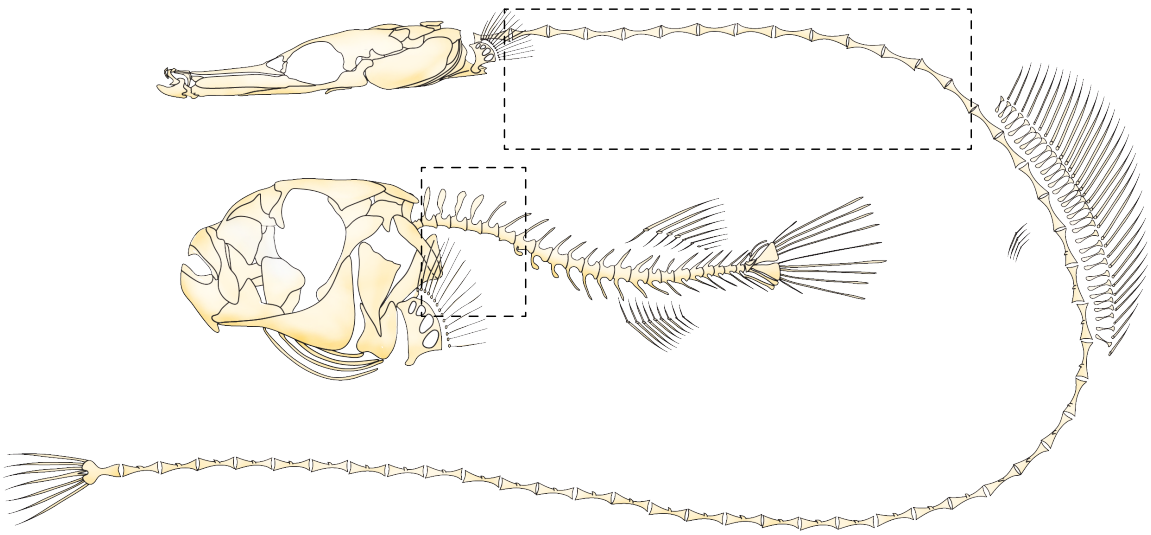


Figure 3.1: Fugu and pipefish have convergently lost their ribs. Cartoon illustrating loss of ribs in pipefish (outer skeleton) and pufferfish (inner skeleton). Skull of pufferfish redrawn based on (Tyler 1980). Boxes mark where ribs are missing.

Despite the seeming connection between *hox7* genes and rib loss, it remains unclear and untested what impact the loss of these *Hox* genes have on the teleost body plan. To date, only one study has manipulated *hox7* expression in fish. Morpholinos targeting the *hoxb7a* gene in zebrafish resulted in developmental delay with hypopigmentation and shortening and bending of the tail (Rochtus et al. 2015). Unfortunately, these fish were phenotyped at pre-skeletal development stages (24 hours

post fertilization), therefore the potential downstream skeletal phenotypes are still undescribed in fish. Because of the evolutionary divergence of mice from teleost, which encompasses an entire round of *Hox* cluster duplication (Figure 1.1), it is hard to confidently apply the phenotypic knockout results seen in mice to the evolutionary loss of these genes in teleost fish. Additionally, what is the phenotypic impact of losing the function of one as opposed to both of these *Hox* genes in fish?

In this study, we directly test for the effect of loss or modification to *hox7* genes have on the body plan of a teleost percomorph fish, the threespine stickleback (*Gasterosteus aculeatus*), in order to provide further insight into the role the loss of these genes played in the transition to the highly modified syngnathid body plan. To perform our work, we utilized the CRISPR/Cas9 system to induce indels in the *hox7* genes in stickleback fish. We tested the prediction that the loss of *hox7* genes in the stickleback will lead to phenotypes affecting the ribs. We chose stickleback because of their relatively closer evolutionary position to syngnathids than zebrafish. An additional benefit is the presence of naturally segregating genetic variation in stickleback that will permit future studies of epistatic interactions between induced mutations and natural modifier alleles. We successfully made CRISPR mutants in stickleback. An initial survey of phenotypic effects indicate that single mutations do not seem to affect phenotypes, but the abrogation of function in both paralogs does create effects in rib phenotypes in expected ways.

MATERIALS AND METHODS

Overview of experimental design

The generation of transgenic lines in stickleback began with injecting half of a clutch of fertilized eggs with the Cas9 mRNA and guide RNA. The uninjected siblings were raised to adulthood, then euthanized and a portion were phenotyped. The injected siblings were raised until breeding age (about nine to 12 months). They were then placed into individualized tanks, live fin-clipped, genotyped via examination of chromatograms from Sanger sequences, and reorganized into group tanks based on whether they were screen positive or negative for a CRISPR lesion at the targeted locus. The screen negative fish were euthanized.

An G1 generation was made using sperm and eggs from fish that screened positive for the lesion. These fish were raised to adulthood and then screened for CRISPR lesions via TOPO cloning. G1 families that contained frameshift alleles were kept for making G2 lines and for phenotyping. The remaining G1 fish were used in a preliminary screen of axial defects and further assessment of CRISPR induced lesions in the *hox7* genes (Figure 3.2).

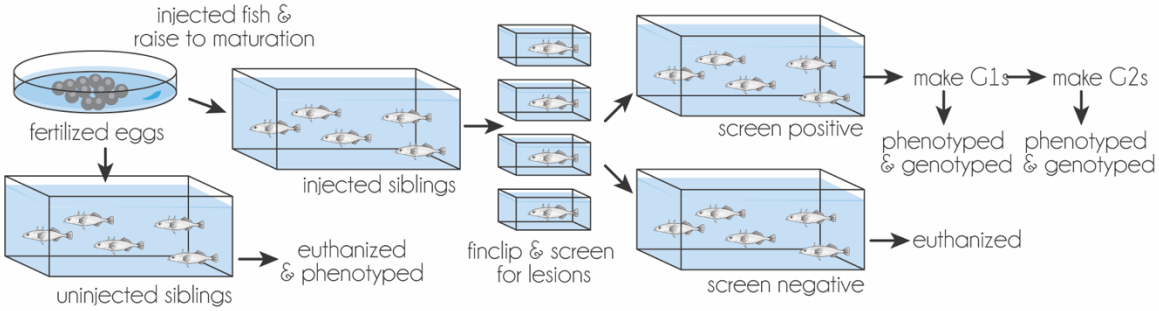


Figure 3.2: Overview of experimental design for CRISPR injection and screening. This design was repeated in stickleback where only the *hoxa7*a was targeted, only the *hoxb7a* was targeted, and both the *hoxa7a* and *hoxb7a* genes were simultaneously targeted.

CRISPR guide RNA (gRNA) design and injections

The CRISPR/Cas9 system was used to induce indels or larger deletions in *hox7* genes (*hoxa7a*, *hoxb7a)* in stickleback (Figure 3.3a). The overall technique is based on Hwang et al. (2013) and Jao et al. (2013) which provide methodological details on successful use of the CRISPR/Cas9 system in zebrafish embryos. Cas9 mRNA was used that is both optimized for zebrafish and stickleback codon usage. We modified and optimized the procedure as shown in Fig. 3.2.

Target sequences following the GG-(N)18-NGG or G-(N)19-NGG pattern to be used for the CRISPR gRNA for the *hoxa7a* and *hoxb7a* in the stickleback genome (Ensembl BROAD S1 assembly) was identified using CCTOP (https://crispr.cos.uni-heidelberg.de) (Table 3.1). Target sites were designed to target upstream of conserved homeodomain and the hexapeptide in the coding sequence of *hoxa7a* and *hoxb7a* (Figure 3.3).

Table 3.1: CRISPR recognition sites present in target genes using the GG-(N)18-NGG recognition site in stickleback genome. Gray shaded recognition sites were used for the CRISPR gRNA experiment.

Target Location	Sequence
hoxa7a, exon 1	GGGACCCCTCACCTTGCCGCCGG
hoxa7a, exon 1	GGCGGCAAGGTGAGGGGTCCCGG
hoxa7a, exon 1	GGCTGGGCGGTTCTGGTACACGG
hoxa7a, exon 1	GGCCGTATCCCGTGAAGGCTGGG
hoxa7a, exon 1	GGCCGCACAGTCCGAGCCGAGG
hoxb7a, exon 1	GGCGACGAGGAAGAATGGGAGGG
hoxb7a, exon 1	GGCAGAGCTGAGACCAATCGGGG
hoxb7a, exon 1	GGGCGACGAGGAAGAATGGGAGG
hoxb7a, exon 1	GGAAAGAGATGAAGAAATGGTGG

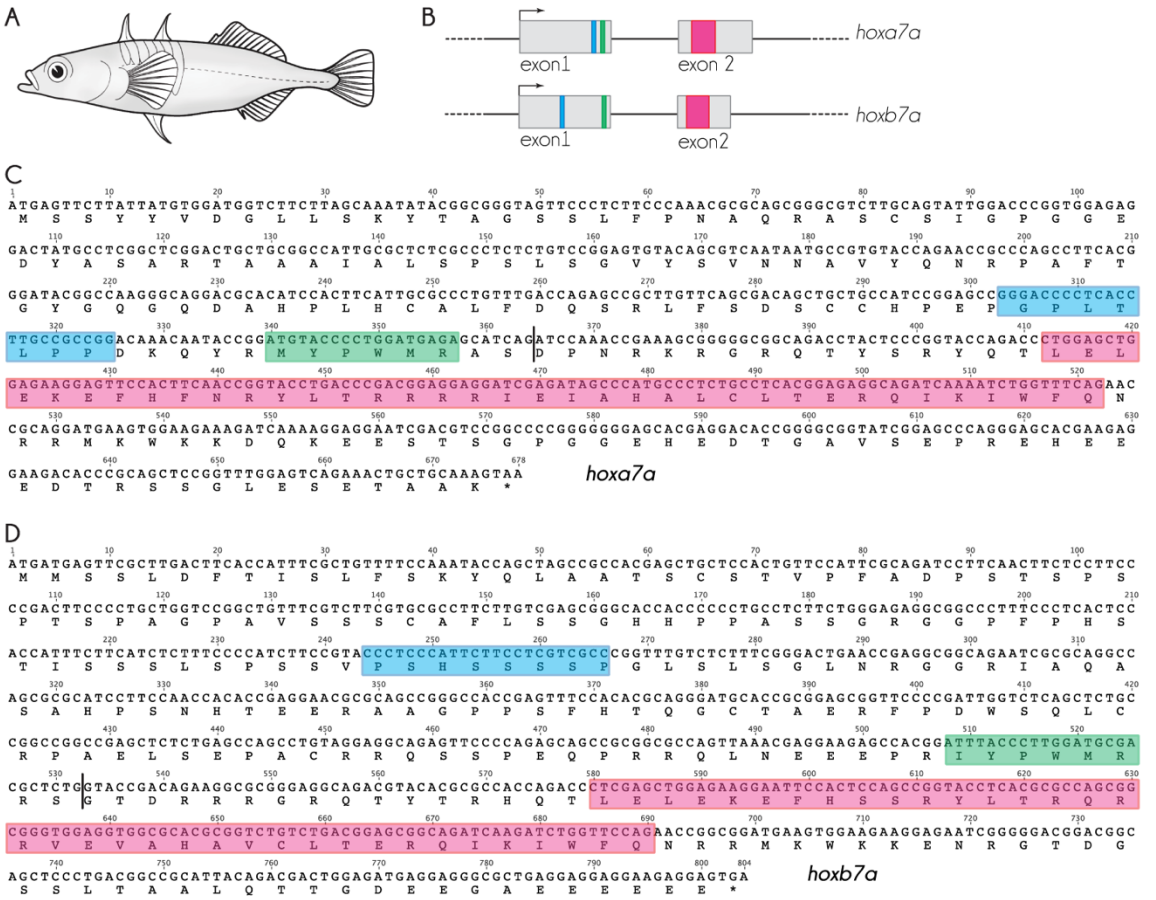


Figure 3.3: CRISPR/Cas9 system was used to induce indels in *hoxa7a* and *hoxb7a* genes in threespine stickleback. a) A cartoon representation of the threespine stickleback. b) CRISPR target site (blue bar) were designed early in the coding sequence of *hoxa7a* and *hoxb7a*, upstream of conserved homeobox sequence (pink bar) and the hexapeptide (green bar). c) *hoxa7a* coding sequence and d) *hoxb7a* coding sequence with CRISPR target site (blue box), hexapeptide (green box), conserved homeobox sequence (pink box), location of intron (black dash) marked. Intron length is 864 base pairs and 4226 base pairs, respectively.

Site-specific gRNA was transcribed from templates created by annealing two long oligonucleotides and using PCR to generate dsDNA using a custom designed gene specific oligo and a scaffold oligo. The MegaScript T7 kit was used to create the gRNA DNA Template. Custom, gene specific oligos were ordered from Eurofins Genomics with the following sequence organization: 5'—[T7 promoter]-[Target Sequence]-[start of

gRNA sequence]—3' (5`-aattaatacgactcactata-[20 bp Target Sequence]-gttttagagctagaaatagc-3'). gRNA scaffold oligo was as follows: 5`-gatccgcaccgactcggtgccacttttttcaagttgataacggactagccttatttttaacttgctatttctagctctaaaac-3`. RNA Clean and Concentrator-5 kit was used gRNA scaffold was ordered from Eurofin Genomics custom oligos.

Crosses and husbandry of stickleback fish

Crosses were made using the Rabbit Slough genetic line, and the offspring were grown using standard husbandry procedures developed in the Cresko Lab (Cresko et al. 2004). All protocols and procedures adhere to University of Oregon IACUC approved methods for the ethical care and use of animals.

Briefly, after embryos entered the two cell stage, about one hour after fertilization at 20°C, they were cleaned with embryo medium (EM), consisting of 4 ppt artificial sea water (Instant Ocean) dissolved in nanopure water. Groups of 20 embryos were placed in individual 26 Å~ 100 mm^2 Petri dishes filled with ~75 ml of EM, and raised in an incubator maintained constantly at 20°C. Any non-developing embryos were removed daily and 100% of EM was changed.

Rearing continued in this manner until 9 dpf, at which point the fry had hatched and their yolks had been absorbed. Fry were placed in a recirculating aquaculture system. Water temperature was maintained at 20°C, and a salinity of 4 ppt was maintained with Instant Ocean. Fish were fed ad libitum with live *Artemia nauplii* (brine shrimp) and dry food (Ziegler AP100 larval food) twice per day.

Injection of guide RNA and Cas9 mRNA into stickleback embryos

Embryos were made using the Rabbit Slough genetic line, and the offspring were grown using standard husbandry procedures developed in the Cresko Lab (Cresko et al. 2004) (see previous *Crosses and husbandry of stickleback fish.* section for more details). Eggs were fertilized in the lab via squeezing eggs from gravid female stickleback and using dissected testes from males. Once fertilized, each individual clutch was divided into two lots—one lot that would be injected with CRISPR and one lot that would remain as uninjected sibling controls.

Stickleback embryos were injected with Cas9 mRNA and target specific gRNA in a one cell stage (45 minutes post fertilization) with 1–2 nl of injection mixture that consisted of water with 1/10 to 1/20 volume phenol red, 50 ng/µl Cas9 RNA and 50 ng/µl gRNA per target site. Two clutches were injected with CRISPR targeting *hoxa7a*, two clutches were injected targeting *hoxb7a*, and three clutches were injected targeting both *hoxa7a* and *hoxb7a* (Figure 3.2).

Screening of injected stickleback for potential mutations

The number of dead embryos was recorded daily for the first nine days post fertilization of injected fish until the fry were moved to the open system tanks. Embryos were grown to maturation (eight to ten months). Injected fish were individualized and fin-clipped.

DNA extractions and Sanger sequencing was performed to screen for CRISPR indels. Qiagen DNEasy protocol and AMPure beads were used to extract DNA from fin clips. PCR primers were designed around the CRISPR target sites, PCR reactions were performed DNA extracts using Thermo Fisher Scientific PCR Master Mix (2x)

(Appendix A, Table S3.1). PCR was cleaned using AMPure beads or Zymo Research Clean and Concentrator columns. PCR product was sent to Genewiz for Sanger sequencing. Chromatogram files from the Genewiz Sanger sequencing were examined on Geneious 8.1.9 to look for indels at the target site (Figure 3.4).

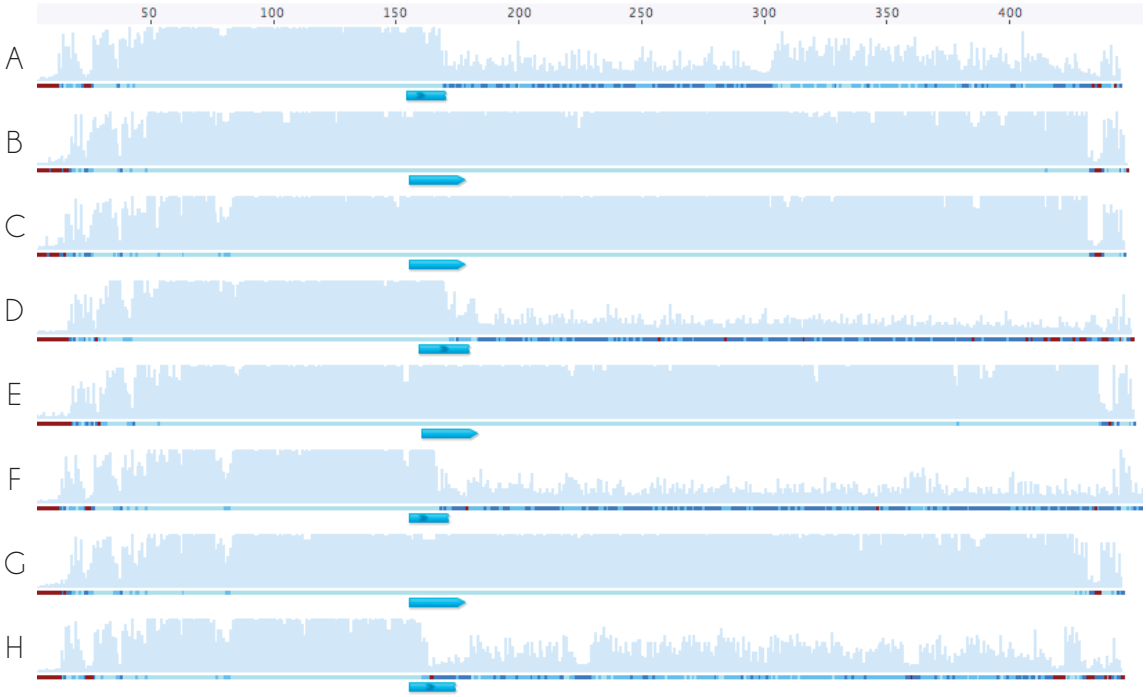


Figure 3.4: Chromatogram files were used to identify presence of CRISPR indels in injected stickleback. Above is a screenshot from the Geneious software program of eight example chromatogram files from Sanger sequences from the CRISPR target region location of the CRISPR recognition site is labeled with the blue rectangles underneath the chromatograms. A dip in height of chromatogram peaks starting at the CRISPR recognition site indicates successful introduction of a CRISPR indel (sequences A, D, F, H). Uniform, tall peaks for the length of the Sanger sequence chromatogram indicates failed introduction of a CRISPR indel (sequences B, C, E, G).

G1 crosses and screening

Once individual fish were identified as potential carriers for CRISPR indels based on Genewiz Sanger sequencing chromatogram files, they were labeled as screen positive

for a CRISPR indel and separated from fish from the same clutch that screened negative for indels. Fish that screened negative for indels were euthanized with MS-222. Only males and females that screened positive for indels were used to produce the next generation of crosses (referred to as the G1 generation).

Six G1 crosses were made for *hoxa7a*, 13 crosses for made for *hoxb7a*, and six crosses were made for *hoxa7a* and *hoxb7a*. In each cross, both the mother and father were screen positive for CRISPR indels. A portion of G1 crosses were eventually screened for CRISPR indels from fin clips. TOPO cloning and Sanger sequencing was performed to identify the type of CRISPR lesions present in each cross. Two crosses from each condition was kept alive in the fish colony and other crosses were euthanized and phenotyped.

Alcian and alizarin staining

Once collected, fish were euthanized with MS-222, fin-clipped for possible future genotyping , and individualized. These fish were then fixed in 2% paraformaldehyde PFA, washed and then stained for bone based on protocol from Walker and Kimmel (2007). In order to achieve the appropriate degree of clearing and staining for phenotyping the axial morphology of adult stickleback, the fixation with 2% PFA was limited only two to four hours, fish were bleached in 3% hydrogen peroxide until their body pigment turned white (about one hour), enzymatic clearing in 2% trypsin lasted until fish body were flexible, and fish were stained with 0.02% Alizarin/10% Glycerol/0.5% KOH for 24 hours in order to achieve a high degree of staining of the bones. After the alizarin staining step, fish were washed in 35% saturated Na-Borate for one to several days until the body of the fish cleared enough to visualize the vertebral

column and ribs. The length of each step of the protocol was modified according to the size of the fish and visual assessment. Most of the samples were stained only with Alizarin Red. Specimens were stored in 80% glycerol.

Phenotyping of rib morphology

After the alcian and alizarin staining, standard length of each fish was measured. Fish were dissected under a Leica MZ6 stereomicroscope in a solution of 50% glycercol in a 26 Å~ 100 mm^2 Petri dish with an agarose bottom. The lateral plates, along with the pelvic structure and jaw elements were carefully removed without damaging the ribs or vertebrae using tweezers and scissors. With the specimen laying laterally, the number of caudal vertebrae were counted. A specimen was then pinned ventral side up in the Petri dish. Using scissors and tweezers, the specimen was given a superficial midsagittal incision that pinning and visualization of the precaudal vertebrae and left and right ribs.

Number of precaudal vertebrae, number of caudal vertebrae were recorded, along with position and number of left and right pleural ribs (Figure 3.5). Epipleural ribs were unable to be counted without extreme damage to the specimen and therefore was not recorded. The first caudal vertebra was considered the one directly anterior to the first anal fin ray as defined in (Bowne 1994). An alternative count was also taken where the first caudal vertebra was considered the first vertebra with a well-defined haemal spine, as both definitions of what is considered the first caudal vertebrae is used. This method of defining the first caudal vertebra moved the position of the first caudal vertebrae to one position more posterior in 113 of the 63 fish. Any observed deformities were also recorded. Statistical analysis was conducted in R on the count data. Data was normalized to a Poisson distribution and a generalized linear model was performed to test for any

significant differences in the different groups of fish. Family effect was incorporated as a random effect in the models.

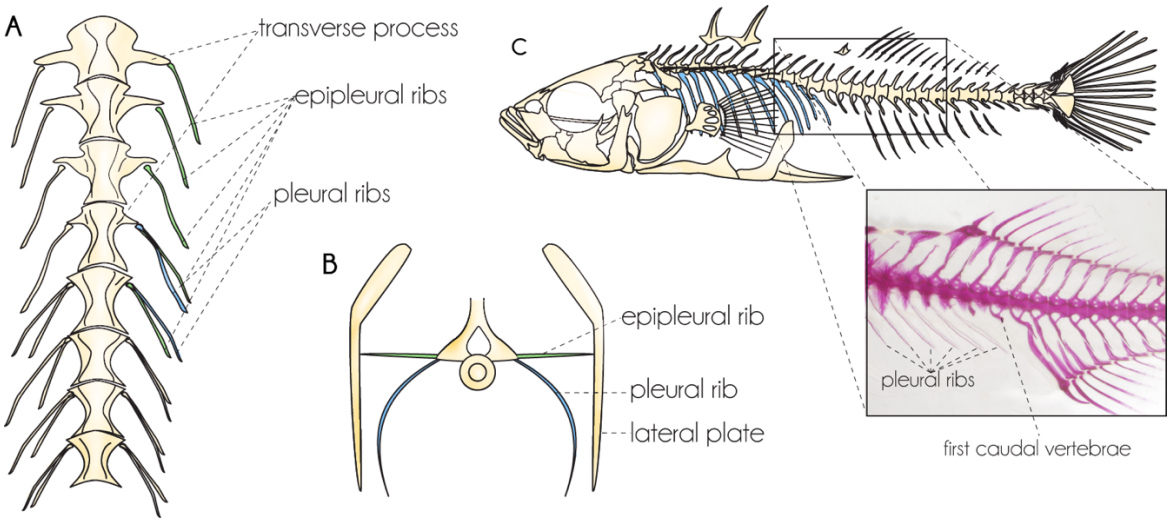


Figure 3.5: Rib morphology of the threespine stickleback. a) ventral view illustration of the first eight vertebrae on a representative wild type stickleback with examples of epipleural ribs (green), pleural ribs (blue), and transverse processes labeled. b) cross section of a rib-bearing stickleback vertebra illustrating the difference between epipleural (green) and pleural ribs (blue). Epipleural often articulate with lateral plates when present in stickleback. Illustration of cross section of rib redrawn from (Nelson 1971). c) illustration of threespine stickleback without lateral plates. Boxed area shows example of what was counted as the first caudal vertebra and pleural ribs.

RESULTS

Significant number of injected fish screened positive for lesions

All injected stickleback were screened for CRISPR indels. The percent of injected embryos from a single clutch that screened positive for an indel through chromatogram examination ranged from 20% to 70% (Table 3.2). There were no noticeable differences in survivorship or success rate of inducement of indels across the groups.

Table 3.2: Percentage of injected fish that screened positive for a CRISPR induced indel per clutch.

CRISPR target	Family Number	% screen positive P0
hoxa7a	3131	38%
hoxa7a	3135	65%
hoxb7a	3129	56%
hoxb7a	3133	70%
hoxa7a & *hoxb7a*	3127	57%
hoxa7a & *hoxb7a*	3141	40%
hoxa7a & *hoxb7a*	3143	20%

Germline transformation was efficient and created a range of lesions in both genes

Six G1 crosses were generated from *hoxa7a* P0 fish, 13 G1 crosses were generated from *hoxb7a* P0 fish, and six G1 crosses were generated from *hoxa7a* and *hoxb7a* P0 fish. Several individuals were genotyped from nine of the 25 G1 clutches. CRISPR alleles were detected in eight out of nine of the genotyped lines, giving us an estimated 89% success rate in the CRISPR alleles transferring to the germline. In many of the G1 lines, up to three different types of lesions for the CRISPR target were identified. This can only be explained by compound heterozygotes being generated in the PO generation at the injection stage and these alleles being present in their germline (Appendix A, Table S3.2). The lines that carried frameshift alleles were selected for generation of G2 lines.

From the 21 unique CRISPR alleles detected, ten generated an early stop codon, two added an extra 11 amino acids past the wild type peptide, and nine did not create an early stop codon. Deletions ranging from two to 21 nucleotides was the most common, with one 55 nucleotide deletion detected. Of these deletion alleles, seven of the 13 were in multiples of three and did not cause frameshifts. Three of the alleles were insertions

that were one, nine, and 14 nucleotides in length. Five of the alleles were complex indels that were made up of two, three, or four tandem insertion and deletions. The individual indels that made up these complex lesions ranged from deletions that were one to 14 nucleotides in length and insertions that were one to 18 nucleotides in length (Figure 3.5; Appendix A, Table S3.3).

No significant difference in number of axial elements in G1 fish

A total of 41 stickleback fish from three *hoxa7a* G1 families, 49 stickleback fish from four *hoxb7a* G1 families, 30 stickleback fish from two G1 families where both the *hoxa7a* and *hoxb7a* genes were simultaneously targeted (denoted as "*hoxa7a;hoxb7a*"), and 35 fish from wild type families were used for alcian and alizarin skeletal preparations and phenotyped. The 35 fish from wild type families were used as controls. Individual genotypes were not available for these fish, but these data can be collected subsequently because each fish has had a portion of its tissue sampled for DNA extraction. The goal of this round of phenotyping was to conduct a preliminary survey of axial skeletal variation present in these fish and to make note of any recurring deformities or difference in any of subset of these G1 fish compared to the control fish.

There was variation in number of vertebrae, number of caudal vertebrae, number of precaudal vertebrae and number of pleural ribs across all the fish examined, including in the controls (Table 3.3). Interestingly, the fish who came from clutches where both parents screened positive for an indel in the *hoxa7a* and the *hoxb7a* had a larger range of total vertebrae present, with specimens having the lowest total number of vertebrae (27), lowest number of precaudal vertebrae (10), and lowest numbers of left pleural ribs (eight) and right pleural ribs (seven).

Table 3.3: Variation in number of axial elements across the different categories of G1 families. The "CRISPR Target" column indicates at what locus both parents of that G1 family screened positive for indels. Families with the same CRISPR target were pooled together.

CRISPR Target	Number of Fish	Total vert.	Precaudal	Caudal	Left ribs	Right ribs
hoxa7a	41	30 to 32	14 to 15	16 to 18	9 to 12	10 to 12
hoxb7a	49	30 to 33	13 to 16	15 to 18	10 to 13	10 to 13
hoxa7a;hoxb7a	30	27 to 34	10 to 17	14 to 19	8 to 13	7 to 13
control	46	28 to 34	14 to 15	13 to 19	9 to 11	9 to 11

Despite this trend, the differences were not statistically significant. The total number of vertebrae was not significantly different between fish from the *hoxa7a* G1 families, from the *hoxb7a* G1 families, and from the *hoxa7a;hoxb7a* G1 families (χ^2=0.4729; d.f. =3, p=0.9248). The number of precaudal and caudal vertebrate were also not significantly different among the G1 family types (precaudal: χ^2=0.6126; d.f.=3, p-value=0.8935; caudal: χ^2=0.0821; d.f.=3, p-value=0.9939). Additionally, the precaudal ribs were not significantly different between the G1 family types (left pleural ribs: χ^2=0.9393; d.f.=3, p-value=0.8159; right pleural ribs: χ^2=1.436; d.f.=3, p-value=0.6971). The first vertebrae to carry a pleural rib also varies and sometimes the appearance is asymmetrical. For example, a fish might have their pleural ribs on the left side of the second vertebrae of its vertebral column, but then it might not appear on the right side until the third vertebrae. Therefore, the total number of anterior precaudal vertebrae that do not bear pleural ribs was counted, but there was no significant difference among the

Figure 3.6: CRISPR mutant alleles identified in G1 fish stocks. The top sequence in each column show wild type with gRNA sequence in red as reference. Blue indicate insertional mutations, dashes indicate deletion mutations. Indels that cause a frameshift are highlighted in green. See Appendix A, Table S3.3 for more information regarding the position of the early stop codons for the individual alleles.

G1 family types (left: χ^2=0.4604; d.f.=3, p-value=0.9275; right: χ^2=1.4365; d.f.=3, p-value=0.697) (Appendix A, Figures S3.1 and S3.2).

Hoxa7a G1 fish have few axial abnormalities

A total of 41 stickleback fish from three *hoxa7a* G1 families were used for alcian and alizarin skeletal preparations and were then phenotyped. The goal of that round of phenotyping was to conduct a preliminary survey of axial skeletal variation present in these fish and to make note of any recurring deformities or difference in any of subset of these G1 fish compared to the control fish. We can say that the parents of these G1 fish were either heterozygous or compound heterozygous for CRISPR-induced indels at the *hoxa7a* locus. Therefore, a certain percentage of these G1 fish from any given family is either wild type, heterozygous, or compound heterozygous for a CRISPR-induced indel.

A single fish from the *hoxa7a* G1 families exhibited an apparently mutant phenotype where an extra pair of epipleural and pleural is present on the right side of the third vertebra. We dubbed this phenotype the "doublet deformity." This type of deformity never appeared in the control fish examined (Figure 3.6). This same specimen also exhibited deformities on the first vertebra with an extra process developing on the left side of the vertebra. A second specimen also carried deformities on the first and second vertebra. A third specimen had deformed two caudal vertebrae and fourth specimen had an abnormal bump on one of their ribs (Table 3.4).

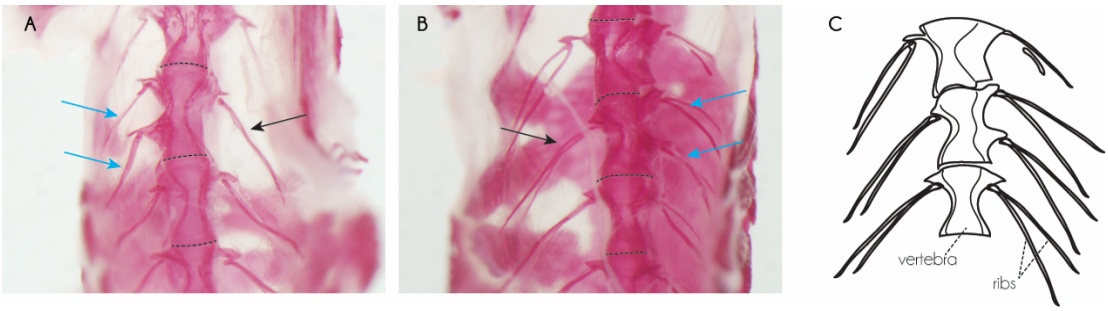


Figure 3.7: Doublet deformity appeared repeatedly in G1 fish. a) and b) Ventral views of alizarin stained rib cages as two examples of the doublet deformity in two different fish. Boundaries between individual vertebra marked with dashed line for clarity. Arrows indicate where the double deformity appears, with blue arrows pointing to two sets of epipleural and pleural ribs appearing on one side of a single vertebra and black arrows pointing to the single set of epipleural and pleural ribs appearing on the opposite side of that same vertebra. c) drawing illustrating the doublet deformity for clarity. Middle vertebra displays deformity.

Table 3.4: Percentage of phenotyped specimens with axial deformities. Each row is an individual family. The "CRISPR Target" column indicates at what locus both parents of that G1 family screened positive for indels. Sample size lists the number of fish from each family that was phenotyped. The percent of specimens with various axial deformities are listed in the last four columns.

CRISPR Target	Family	Sample Size	Doublet Deformity	Rib Deformity	Precaudal Deformity	Caudal Deformity
hoxa7a	3189	16	6.25%	0.00%	12.50%	6.25%
hoxa7a	3210	15	0.00%	6.67%	0.00%	0.00%
hoxa7a	3228	10	0.00%	0.00%	0.00%	0.00%
hoxb7a	3195	10	0.00%	0.00%	0.00%	0.00%
hoxb7a	3204	6	0.00%	50.00%	16.67%	16.67%
hoxb7a	3216	13	0.00%	0.00%	0.00%	0.00%
hoxb7a	3217	10	0.00%	0.00%	0.00%	0.00%
hoxb7a	3218	10	0.00%	10.00%	0.00%	0.00%
hoxa7a;hoxb7a	3240	15	33.33%	26.67%	20.00%	13.33%
hoxa7a;hoxb7a	3241	15	26.67%	26.67%	20.00%	40.00%
control	3126	15	0.00%	0.00%	0.00%	13.33%
control	3130	10	0.00%	0.00%	0.00%	0.00%
control	3142	10	0.00%	0.00%	0.00%	10.00%

Hoxb7a G1 fish have few axial abnormalities

A total of 49 stickleback fish from four *hoxb7a* G1 families were used for alcian and alizarin skeletal preparations and phenotyped. None of these fish exhibited the doublet deformity. Four of the 49 fish examined carried deformities that included mis-shaped ribs, bifurcating ribs, and fused ribs were two adjacent ribs joined together on their posterior ends. One fish had multiple deformed caudal vertebrae. This same specimen also had several precaudal deformities as well (but not the double deformity).

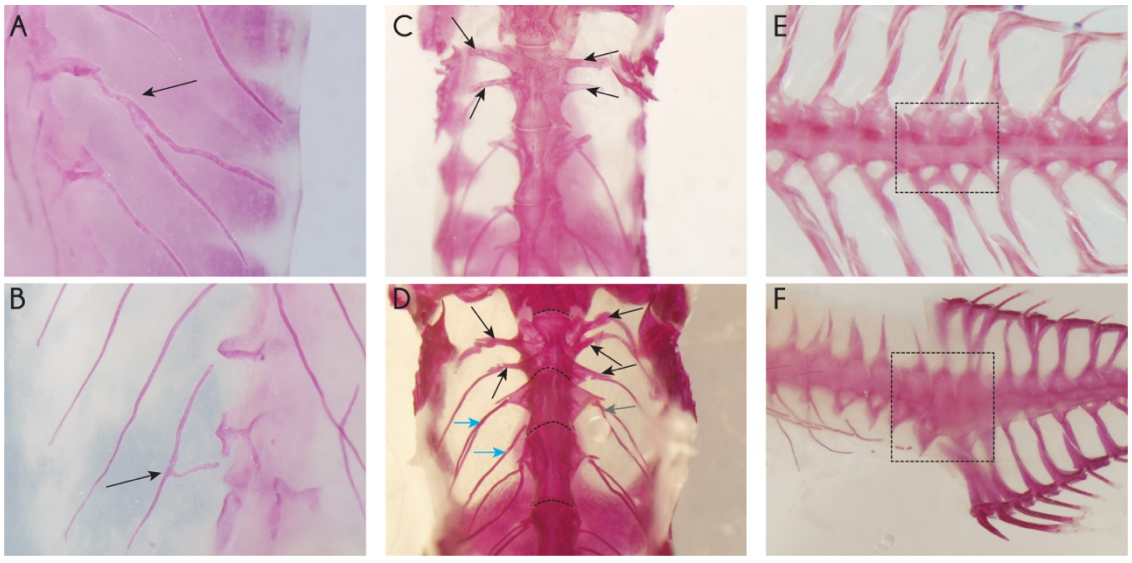


Figure 3.8: Representative pictures of axial deformities observed in G1 fish. a) black arrow points to start of bifurcation of a single pleural rib. b) black arrow indicates point of fusion for two adjacent pleural ribs. c) black arrows point to four processes located on the first vertebrae when normally only two processes develop. d) black arrows point to five processes located on the first vertebrae when normally only two processes develop, blue arrow and gray arrows point out a double deformity present on this specimen as well. Boundaries of individual vertebral elements drawn with dashed line for clarity. e) dashed box highlights location of fused caudal vertebrae. f) box highlights location of mis-shaped caudal vertebrae.

Hoxa7a;hoxb7a G1 fish have the highest occurrence of axial abnormalities

A total of 30 stickleback fish from two G1 families where both the *hoxa7a* and *hoxb7a* genes were simultaneously targeted were used for alcian and alizarin skeletal preparations and phenotyped. Nine of these fish had the doublet deformity (Figure 3.6). Four of the nine fish showed this deformity on two of their vertebrae and one of these fish had three vertebrae that had the doublet deformity. Seven of the 30 fish had rib deformities that extremely mis-shaped and bent ribs. The same type of fusion as described in the *hoxb7a* fish appeared in two of these fish. Three fish had both a doublet deformity and were one of the seven fish with rib deformities. Nine of the 30 fish also had caudal deformities. This included fused caudal vertebrae, bent caudal vertebrae, and caudal vertebrae with mis-shaped processes. Six of the 30 fish have other deformities on their precaudal vertebrae other than the doublet deformity. These deformities included mis-shaped ribs and extra processes developing on one side or both sides of a vertebral element.

In summary, the CRISPR/Cas9 approach to editing stickleback genomes was very efficient. On average, half of the injected fish screened positive for a CRISPR indel. In addition, a high percentage of these lesions caused germline transformations. In the first generation (made with injected fish as the parents), eight out of nine of the genotyped lines contained CRISPR lesions, giving us an estimated 89% success rate in incorporating CRISPR indels into the germline. Many of these G1 lines have more than two CRISPR mutated alleles, indicating a high rate of compound heterozygotes being produced in the parental, injected generation.

Multiple occurrences of axial deformities were present in the CRISPR *hoxa7a*, *hoxb7a*, and double-targeted *hoxa7a*;*hoxb7a* lines that never appeared in the control

groups (Figure 3.9). The highest rate of precaudal, caudal, and rib deformities appeared in fish whose parents both screened positive for indels in the *hoxa7a* and the hox*b7a* loci (Table 3.4).

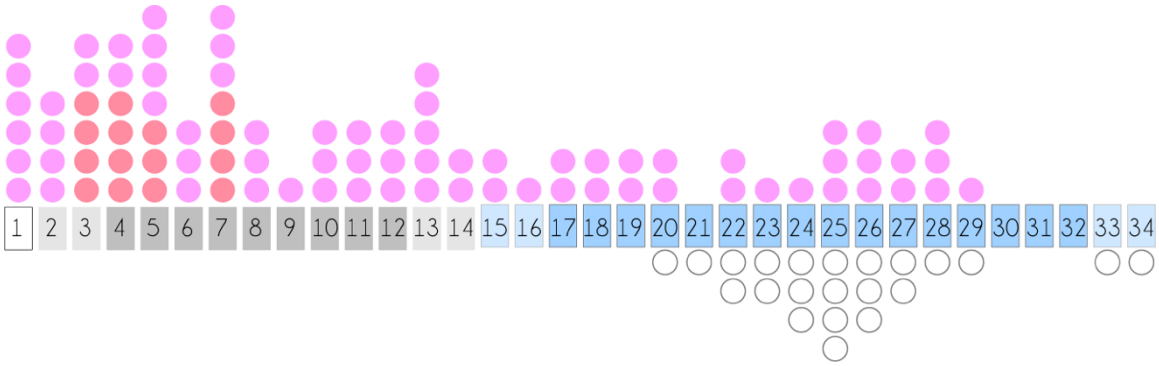


Figure 3.9: Distribution of deformities across all 155 phenotyped fish. Boxes represent individual vertebra numbered 1 to 34. Darker gray boxes indicate the average location of the pleural ribs, with lighter gray boxes indicate full range of pleural ribs. Darker blue boxes indicate the average extent of the caudal vertebrae with the lighter blue boxes indicating the full range of caudal vertebrae. Individual pink and red circles above the boxes represent an observed axial deformity and on what vertebra that deformity was found among the G1 fish examined. Red boxes represent doublet deformities specifically. Individual white circles below the boxes represent an observed axial deformity and on what vertebra that deformity was found among the control fish examined.

DISCUSSION

Creation of mutant stickleback by CRISPR is highly efficient

We showed previously that syngnathid fish have convergently lost all copies of their *hox7* genes, but the morphological impact of this loss of these developmental genes was unclear. To address this problem, we investigated the function of the *Hox* candidate genes to begin to characterize the developmental genetic underpinnings of the striking evolution of derived characters present in the pipefish family.

To study the function of *hox7* genes in teleost fish, we successfully disrupted the *hox7* paralogs from the threespine stickleback genome using the CRISPR/Cas9 system.

We were able to not only make mutations but do it very efficiently. From the 21 unique CRISPR alleles detected, ten generated an early stop codon. All deletions detected ranged from 2 to 21 base pairs, with one exception of a 55 nucleotide deletion detected. Three of the detected alleles were insertions that were one, nine, and 14 nucleotides in length. Five the alleles were complex indels that were made up of two, three, or four tandem insertion and deletions. The individual indels that made up these complex lesions ranged from deletions ranging from one to 14 nucleotides in length and insertions ranging from one to 18 nucleotides in length. This type of efficiency in gene editing allows one to reasonable create an allelic series of numerous different types of lesions from synonymous, to slight hypomorphs, to loss of function knockouts, to complete removal of the gene from the genome. To do so would just involve rounds of parallel injections and screening

Our findings show that CRISPR is a much more promising transgenic approach than other methods that have been used previously with little success in stickleback. The production of modified loci was so efficient that injecting few individuals with Cas9 mRNA and the guide RNA allowed for the parallel creation of numerous different single mutations. In fact, the transformation was so efficient, that many cases of compound heterozygotes were found in the injected fish. In fish where two separate genes were targeted, both genes were efficiently mutated, which was a key resource for our identification of phenotypic effects of mutated *hox7* genes. Transgenic lines for the *hox7* gene knockouts are now established in stickleback that can be used in future research.

Phenotypic effects are most prevalent in double target G1 families

We found variation in the total number of vertebral elements in the examined stickleback, although is not surprising as this has been previously reported the number

vertebrae ranges from 29 to 34, with precaudal ranging from 13 to 14 in *Gasterosteus aculeatus* (Ahn and Gibson 1999, Aguirre et al. 2016, Bowne 1994). Intriguingly, in our phenotypic survey, axial deformities affecting the ribs and precaudal were only found in the G1 fish and never found in the controls.

The prevalence of mutant phenotypes was higher in the G1 families whose parents screened positive for lesions at both the *hoxa7a* and *hoxb7a* genes. Redundancy in *Hox* genes from the same paralogous group has been documented (Chen and Capecchi 1997, 1999, Chen, Greer, and Capecchi 1998, Condie and Capecchi 1994, Fromental-Ramain et al. 1996, Gavalas et al. 1998, Horan et al. 1995, Manley and Capecchi 1998, McIntyre et al. 2007, Studer et al. 1998, van den Akker et al. 2001, Wahba, Hostikka, and Carpenter 2001, Wellik and Capecchi 2003, Wellik, Hawkes, and Capecchi 2002). *Hox7* genes were previously shown to be redundant in mice—where only mutations simultaneously targeting both the *Hoxa7* and *Hoxb7* genes led to most severe phenotypes (Chen, Greer, and Capecchi 1998). This trend we find in stickleback G1 fish argues for an overall redundancy of *hox7* genes in vertebrates.

It is possible that some degree of redundancy of function is also shared with the surrounding *Hox* genes, as it has been shown that *hox5*, *hox6*, *hox9*, *hox10*, and *hox11* genes are all important in rib cage development (McIntyre et al. 2007). Still, vertebrates have kept conservation of at least one of their *hox7* genes with the only known exceptions being pufferfish, Gulf pipefish, seahorses, and the dwarf cyprinids (Amores et al. 2004, Small et al. 2016, Malmstrom et al. 2018, Lin et al. 2016, Lin et al. 2017).

At this point, our interpretation of these results are limited because the individual genotypes are currently unavailable for these fish and will need to be collected for more refined testing. We can say that the parents of these G1 fish were either heterozygous or

compound heterozygous for CRISPR-induced indels at the target locus. Therefore, a certain percentage of the G1 fish from any given family is either wild type, heterozygous, or compound heterozygous. A follow up analysis with individual genotypes will provide much more insight into whether disrupting these genes are the causative factor for these recorded axial deformities. Lines should also be made to test the effects in different genetic backgrounds (e.g. ocean vs. freshwater stickleback populations) to order to see if some of the effects have epistatic contributions coming from natural host genetic variation interacting with the mutant allele. Nevertheless, this is a promising early result as these findings mimic to some degree what was seen in mice knockouts for *hox7* genes (Chen, Greer, and Capecchi 1998).

To a broader extent, these deformities also mimicked the axial morphologies of syngnathid fish. A skeletal synapomorphy for this family includes fusion of the first three vertebrae (Ward and Brainerd 2007, Johnson and Patterson 1993). The total number of vertebrae ranges from 31 to 94 in this elongated family. Modifications to the axial body plan such as loss of all ribs is ubiquitous and curved vertebral columns are prevalent in many of the syngnathid lineages (Dawson 1985). If the loss of all *hox7* paralogs in syngnathids was a key evolutionary transition to the loss of ribs in syngnathids, our results also motivate a hypothesis that subsequent modifier mutations would have occurred to stabilize the phenotype. Specifically, we see several additional axial deformities in fish that are likely mutated for both the *hoxa7a* and *hoxb7a* genes that would likely be maladaptive. If so, then the loss of *hox7* paralogs would have led to positive selection on modifier mutations that mitigated or abrogated the negative effects on the axial skeleton in syngnathids.

CONCLUSION

The striking morphology of syngnathid fish have captured the interests of scientists for many years, yet the developmental genetics underlying this unique evolutionary lineage of fish has remained unknown. Our results provide intriguing evidence that the loss of *hox7* genes in teleost fish like syngnathid fish could have led to a modification to their axial development. Although it is debatable whether examples of regressive evolution—the loss of useless characters over time such as ribs in syngnathid fish—are evolutionarily neutral or adaptive, either way, this research is a novel example of a loss of a gene being associated with the evolution of a new divergent body plan. Mechanisms of evolution by gene duplication such as neofunctionalization and subfunctionalization has been emphasized in the past, but with more and more genomes being sequenced, the concept of gene loss as a mechanism for evolution is now being highlighted (recently reviewed by (Albalat and Canestro 2016)). The results of this experiment provide the first insights into the developmental genetic regulation of these syngnathid skeletal modifications.

BRIDGE

Both Chapters II and III are focused on exploring the *Hox* gene content and the phenotypic impact of the evolutionary loss of some of these *Hox* genes. This included using a comparative genomics approach to compare *Hox* cluster gene content in the Gulf pipefish (*Syngnathus scovelli*) against other teleost genomes. I found several key gene losses in the Gulf pipefish *Hox* clusters. One of these losses—the loss of *hox7* genes—was further investigated in Chapter III the approach of using CRISPR/Cas9 system to induce indels in all *hox7* genes (*hoxa7a*, *hoxb7a)* in the threespine stickleback (*Gasterosteus aculeatus*). As discussed in Chapter I, it is thought that the *Hox* genes have stayed organized into genomic clusters due to selective pressure to maintain the numerous conserved noncoding elements found within the boundaries of these gene clusters. It is thought that perhaps modifications to these putative *cis*-regulatory elements have manipulated gene expression which allowed for the diversity of body plans to evolve while managing to maintain the high level of conservation in the *Hox* genes that we see today. Therefore, I was interested to see if there were any changes to putative regulatory elements in the *Hox* clusters of syngnathids that could possibly be contributing to their highly derived body plan.

For Chapter IV, I explore the conserved noncoding elements within the boundaries of the syngnathid *Hox* clusters. I used *Hippocampus erectus, H. comes* and *S. scovelli* as the syngnathid representatives and compared their CNEs to four percomorph teleosts fish, two non-percomorph teleost fish, one non-teleost fish, and two non-fish vertebrate.

CHAPTER IV

LOSS OF IMPORTANT AXIAL AND CRANIAL CONSERVED NONCODING ELEMENTS WITHIN THE SYNGNATHID *HOX* CLUSTERS

INTRODUCTION

Significant portions of genomes contain conserved elements that are not in coding regions of genes (Bejerano et al. 2004, Sandelin et al. 2004, Woolfe et al. 2004). These so-called conserved non-coding elements (CNEs) can be identified by comparing genomic regions among evolutionarily divergent species. These elements can sometimes show a higher level of conservation than the protein coding genes and are potential regulators for genes (reviewed by (Polychronopoulos et al. 2017)). Several studies have shown that these CNEs tend to be overrepresented near developmental genes and genes involved in transcriptional regulation (Sandelin et al. 2004, Shin et al. 2005, Woolfe et al. 2004, Venkatesh et al. 2006, Bejerano et al. 2004). These CNEs have consistently been shown to function as developmental gene *cis*-regulatory elements through functional assays (Shin et al. 2005, Woolfe et al. 2004, Pennacchio et al. 2006, Navratilova et al. 2009).

Despite progress in identifying CNEs near well studied genes, and in model organisms, we are still largely ignorant of the tempo of evolutionary changes in CNEs, particularly across closely related species within a family (Harmston, Baresic, and Lenhard 2013). This gap in our understanding exists because we now are only generating the whole genome sequence data necessary to examine the evolution of CNEs. In addition, it is difficult to identify the 'sweet spot' of lineages for comparative genomics that are divergent enough from one another to allow conservation of functional elements

to emerge from background conservation, but not so divergent that you cannot infer causal connection with phenotypic changes (Harmston, Baresic, and Lenhard 2013). What one needs is a defined set of previously described CNEs for well-studied genes, and to examine their evolution in a family of highly phenotypically diverse organisms.

One of the best studied sets of developmental regulatory genes are the *Hox* genes. These genes reside in clusters, and, therefore, significant work has focused on the identity and functions of CNEs that reside in and around *Hox* clusters and regulate those genes. Therefore, *Hox* clusters provide an excellent model for studies of CNE evolution. In vertebrates, *Hox* genes are a set of highly conserved developmental transcription factors. They code for homeodomain transcription factors that are responsible for determining the body plan of an embryo along the anterior-posterior axis. They are organized into 13 paralogous groups that are arranged into gene clusters (Scott 1992). Often, *evenskipped* (*evx*) genes are included as a member of the *Hox* clusters, as they are closely related homeodomain transcription factors found immediately upstream of the *hox13* genes.

The ancestral set of *Hox* genes consisted of a single cluster of genes, resulting from tandem duplications of an ancestral proto-*Hox* gene (Garcia-Fernandez 2005). Invertebrates, for the most part, still maintain just a single *Hox* complex. Due to subsequent rounds of whole genome duplications, vertebrates have duplicate copies of the *Hox* complex (Pascual-Anaya et al. 2013). In vertebrates, tetrapods have four *Hox* gene clusters (denoted as *Hox* clusters A, B, C, and D), while teleost fish have eight clusters of *Hox* genes due to the whole teleost genome duplication (*Hox* clusters Aa, Ab, Ba, Bb, Ca, Cb, Da, Db) (Amores et al. 1998) (Figure 1.1). The majority of teleost fish have lost their *HoxCb* cluster, while a smaller subset have lost their *HoxDb* cluster. It is thought that the *Hox* genes have stayed organized into these genomic clusters due to

selective pressure to maintain the numerous conserved noncoding elements found within the boundaries of these gene clusters.

One type of CNE identified among the *Hox* cluster are microRNAs—a class of noncoding RNA gene—that also serve as important post transcriptional regulators for expression of surrounding *Hox* genes. The *mir196* microRNAs are located between a subset of the *hox10* and *hox9* genes and *mir10* microRNAs are located between a subset of the *hox5* and *hox6* genes (Tanzer et al. 2005) (Figure 1.2). Numerous other studies have examined noncoding elements and annotated putative *cis*-regulatory elements within the vertebrate *Hox* clusters. This includes studies where CNEs were examined for binding motifs (Chiu et al. 2002, Matsunami, Sumiyama, and Saitou 2010, Kurosawa et al. 2006, Mainguy et al. 2003, Lee et al. 2006). Other studies have annotated the *Hox* clusters to identify long noncoding genes (Yu et al. 2012, De Kumar and Krumlauf 2016). Some research has presented more detailed examination of *cis*-regulatory elements surrounding particular *Hox* genes (Ferretti et al. 2005, McEllin et al. 2016, Tumpel et al. 2007, Tumpel et al. 2006, Tumpel, Wiedemann, and Krumlauf 2009, Knoepfler, Lu, and Kamps 1996, Parker, Bronner, and Krumlauf 2014, Maconochie et al. 1997). It is perhaps modifications to these microRNAs and putative *cis*-regulatory elements that have allowed for the diversity of body plans to evolve and manipulate the expression of these key developmental genes while managing to maintain the high level of conservation in the *Hox* genes that we see today.

Teleost fish make ideal models for studying the *Hox* gene evolution for several reasons. In general, teleost fish are recognized as important models for vertebrate evo-devo in the genomics era (Braasch et al. 2015). As a whole, this class of fish make up around 40% of all vertebrate diversity with over 27,000 described species (Hoegg et al.

2007, Nelson 2006). Additionally, because of the teleost whole genome duplication, fish have more copies and combinations of *Hox* genes and microRNAs than tetrapods. This makes teleost fish a robust comparative, evolutionary framework to study the significance the *Hox* genes play in morphological evolution (Amores et al. 2004, Hoegg et al. 2007). Finally, the duplication of the *Hox* clusters via the teleost whole genome duplication allowed for the possible partitioning of subfunctions among preserved duplicates which may be reflected in differential preservation of CNEs near each duplicate.

A great clade of fishes in which to examine *Hox* CNE evolution that can potentially be linked to morphological evolution are syngnathids. This family includes species of pipefish, seahorses, pipehorses, and seadragons. These charming teleosts display a remarkable level of morphological diversity and phenotypic novelties such as a highly derived head and body plan, elongated body, prehensile tail, and the presence of male pregnancy (Small, Harlin-Cognato, and Jones 2013, Neutens et al. 2014, Bruner and Bartolino 2008). Connections between the highly divergent body plan seen in this family of fish and modification to the *Hox* gene fish has remained an open question for curious biologists.

A key limiting factor in the ability to study the evolution of syngnathid CNEs, and those in the *Hox* clusters in particular, had not only been the lack of genome sequences, but the existence of very few DNA sequence data for this family in general. A watershed point was the production of not only one, but three, whole genome sequences for syngnathid fish from across the phylogeny of this family in late 2016 and early 2017 (Lin et al. 2016, Lin et al. 2017, Small et al. 2016). These are all high quality, gene annotated genomes, with the Gulf pipefish genome providing a chromosomal level assembly. The completeness of these genomes allows for confident annotation of gene and CNEs.

The *Hox* genes for these fish were, for the first time, reported in the Gulf pipefish (*Syngnathus scovelli*), tiger tail seahorse (*Hippocampus comes*) and lined seahorse (*H. erectus*) genome papers (Small et al. 2016, Lin et al. 2016, Lin et al. 2017). Overall, the *Hox* genes were conserved in syngnathids with a few exceptions of interesting gene losses (Small et al. 2016). In a subsequent paper Fuiten et al. (this volume) showed intriguing evidence that the loss of *hox7* genes in syngnathid fish could be related to axial modifications such as fused anterior vertebrae and loss of ribs that this family had evolved over time. However, the regulatory elements within the syngnathid *Hox* clusters remain to be described.

In this study, we asked how conserved the *Hox* noncoding elements are for syngnathids relative to other vertebrates and how many noncoding element gains or losses were specific to syngnathids. In addition, we addressed the question of whether the CNEs are more variable than the *Hox* cluster coding gene content among syngnathids because changes in CNEs are predicted to be less negatively pleiotropic than those in coding regions. If there are any differences to CNE content, can any of these changes be linked to morphological evolution?

We examined the regulatory elements within the *Hox* clusters of three syngnathid genomes—the Gulf pipefish (*Syngnathus scovelli*), the *tiger tail*ed seahorse (*Hippocampus comes*), and the lined seahorse (*Hippocampus erectus*). We found that the conserved noncoding microRNAs in the seahorse genomes match the microRNAs previously described in the Gulf pipefish, the conserved noncoding have remained largely conserved in syngnathid with various levels of phylogenetic conservation relative to other vertebrates, and there is a single key loss of an enhancer.

MATERIALS AND METHODS

Noncoding identification

Genomes used for comparison

CNEs were identified using mVISTA analyses based on levels of sequence conservation within *Hox* clusters across *Gasterosteus aculeatus*, *Takifugu rubripes*, *Oryzias latipes*, *Thunnus orientalis*, *Hippocampus erectus*, *Hippocampus comes*, *Syngnathus scovelli*, *Boleophthalmus pectinirostris*, *Gadus morhua*, *Danio rerio*, *Lepisosteus oculatus*, *Mus musculus*, and *Homo sapiens* (Frazer et al. 2004, Mayor et al. 2000, Brudno, Do, et al. 2003, Brudno, Malde, et al. 2003). Sequences for *D. rerio*, *L. oculatus*, *M. musculus*, and *H. sapiens* were downloaded from Ensembl. *T. orientalis* sequence was extracted from the *T. orientalis* genome ((Yasuike et al. 2016); http://nrifs.fra.affrc.go.jp/ResearchCenter/5_AG/genomes/Tuna_DNAmicroarray/index.html). *G. morhua* sequence was extracted from the *G. morhua* genome ((Torresen et al. 2017); https://figshare.com/articles/Transcript_and_genome_assemblies_of_Atlantic_cod/3408247). *S. scovelli* sequence was extracted from the *S. scovelli* genome ((Small et al. 2016); https://creskolab.uoregon.edu/pipefish/). The *H. erectus* sequence was extracted from the *H. erectus* genome ((Lin et al. 2017); NCBI with the project accession PRJNA347499). The *H. comes* sequence was extracted from the *H. comes* genome ((Lin et al. 2016); NCBI with the project accession PRJNA314292). The *B. pectinirostris* sequence was extracted from the *B. pectinirostris* genome ((You et al. 2014); NCBI with the project accession PRJNA232434). The *T. rubripes* sequences were retrieved from Genbank ((Lee et al. 2006); Genbank accessions DQ481663–9). The *O. latipes* sequences were retrieved from Genbank ((Kurosawa et al. 2006); AB232918–24). The *G. aculeatus*

sequences were from BAC clones, which were make available by Angel Amores. Sequences were softmasked using RepeatMasker.

Noncoding VISTA analysis

G. aculeatus and *S. scovelli* was set as the reference sequence for the VISTA analysis. Alignment of each sequence from these species were aligned using the shuffle-LAGAN algorithm and the LAGAN algorithm through the mVISTA website with Minimum conservation identity set to 65% and Minimum length for a CNS set to 50.

All conserved noncoding sequences annotated within the *S. scovelli Hox* clusters were queried against the NCBI NR database to identify coding exons, against RFAM, refseq_rna, and the miRBase Sequence Databases (Release 21) for mature miRNA chordate sequences and miRNA chordate hairpins (downloaded from miRBase). BBMapSkimmer was used to query against the miRBase Sequence Databases in order to identify RNA genes. Kmer index size was set to 7, max indel set to 0, approximate minimum alignment identity set to 0.50, secondary site score ratio set to 0.25, behavior on ambiguously-mapped reads set to retain all top-scoring sites, and maximum number of total alignments to print per read set to 4 million.

Annotation of microRNAs

Putative seahorse microRNA sequences were first identified using the mVISTA analyses described in the previous section. We aligned primary miRBase (Kozomara and Griffiths-Jones 2011) microRNA sequences from zebrafish and Gulf pipefish to *H. comes* and *H. erectus Hox* regions using MUSCLE (Edgar 2004) to supplement annotations. The hairpin loops of the annotated microRNAs were confirmed using RNAfold

(http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi). When known *Hox* cluster microRNAs were not detected in the seahorse genomes, we further confirmed absence of the conserved seed sequence, which was the case for *mir196b* between *hoxb13a* and *hoxb9a* and *mir10a* between *hoxb5b* and *hoxb3b*.

RESULTS

Seahorses have the same set of microRNAs as the Gulf pipefish

The microRNA content of the two seahorse genomes was not annotated (Lin et al. 2016, Lin et al. 2017). Therefore, we searched for and annotated the *Hox* microRNAs in these seahorse species (Figures 4.1 and 4.2). We found that the two seahorse genomes share the same microRNAs as the Gulf pipefish (Figure 2.4). This included the four *mir10*s microRNAs of *Hox* clusters Ba, Ca, Da and Db and the three *mir196* microRNAs of *Hox* clusters Aa, Ab and Ca that are identified in the Gulf pipefish (Small et al. 2016).

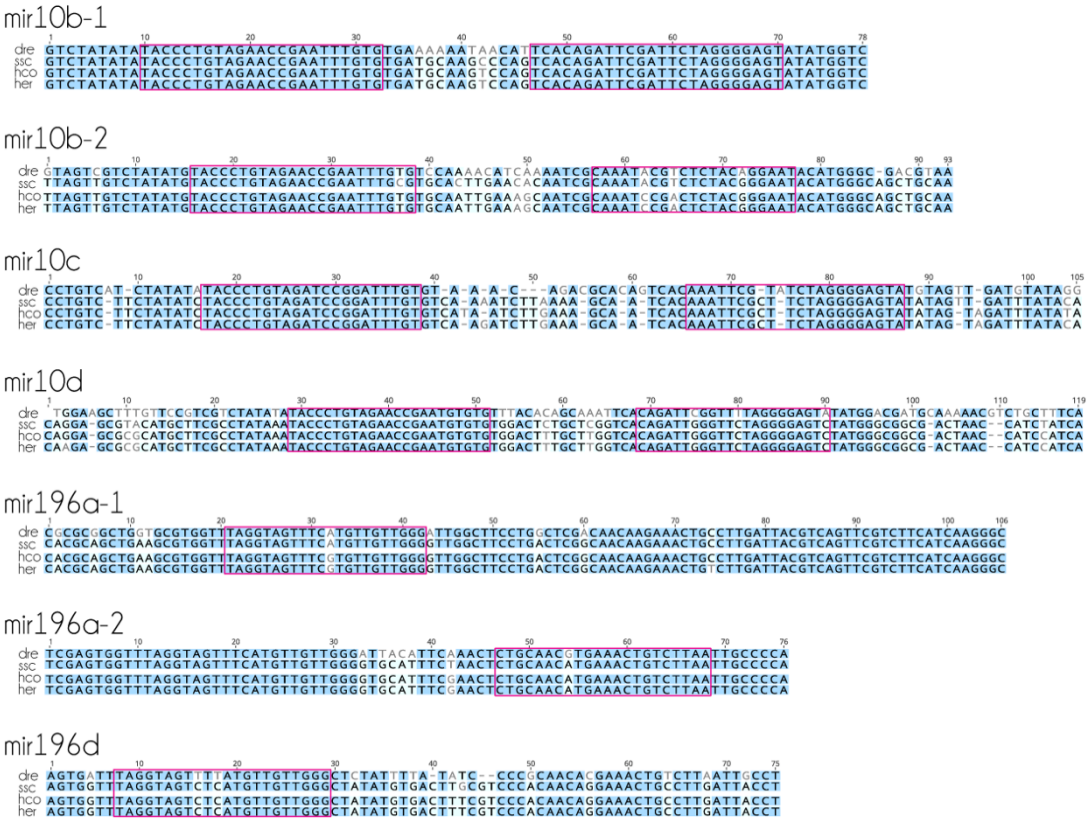


Figure 4.1: MicroRNAs sequences are conserved between seahorses and pipefish. Alignment of the Gulf pipefish (ssc), *tiger tail* seahorse (hco), lined seahorse (her) and zebrafish (dre) mir10 sequences of *Hox* clusters Ba, Ca, Da and Db and *mir196* sequences of *Hox* clusters Aa, Ab and Ca. Mature microRNA sequences in pink boxes.

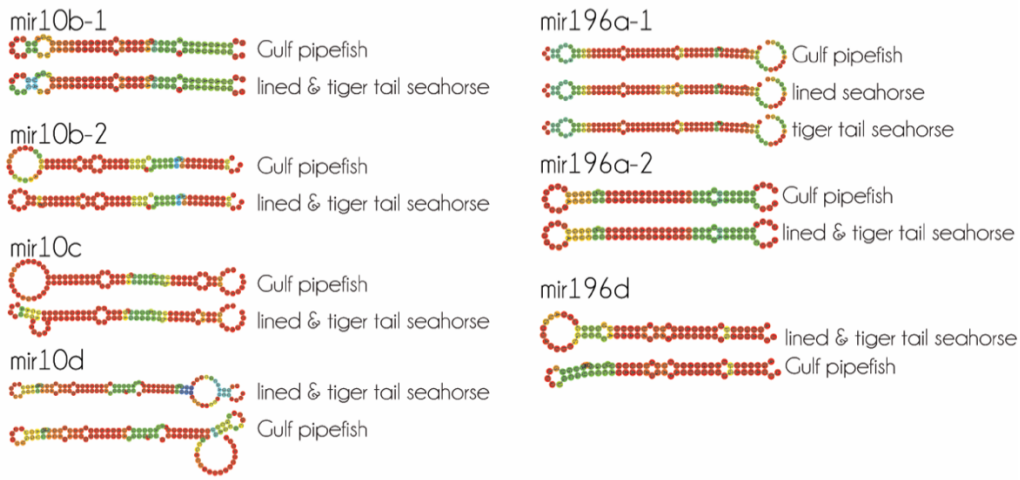


Figure 4.2: MicroRNA foldings are conserved between seahorses and pipefish. Hairpin structures for Gulf pipefish, *tiger tail* seahorse, and lined seahorse microRNAs. Lined and *tiger tail* seahorses have identical sequences for all microRNAs except *mir196*a-1.

Hox Cluster CNEs show various levels of phylogenetic conservation

We cataloged 718 putative conserved noncoding elements within the boundaries of the *Hox* clusters. Each of these elements were a minimal length of 50 bp and were at least 65% conserved with the reference genome. We used *Hippocampus erectus, H. comes* and *Syngnathus scovelli* as the syngnathid representatives and compared their putative CNE content to percomorph teleost fish (*Gasterosteus aculeatus, Takifugu rubripes*, *Oryzias latipes*, *Thunnus orientalis*), non-percomorph teleost fish (*Boleophthalmus pectinirostris*, *Gadus morhua*, *Danio rerio*), non-teleost fish (*Lepisosteus oculatus)*, and two non-fish vertebrates (*Mus musculus* and *Homo sapiens*). 330 of the 718 noncoding elements shared various levels of conservation with *B. pectinirostris*, *G. morhua*, *D. rerio*, *L. oculatus*, *M. musculus*, and *H. sapiens* (see Appendix B, Figures S4.8–S4.14; Table S4.1) (Table 4.1, Figure 4.3B). Additionally, there was a high degree of syngnathid specific noncoding sequence shared between the seahorses and Gulf pipefish. We found 388 of the 718 distinguishable elements (putative CNES) specific to *H. erectus, H. comes* and *S. scovelli* (Table 4.2, Figure 4.3A).

Table 4.1: Number of CNEs described within the seven syngnathid *Hox* clusters and the degree of conservation with other vertebrates. First column lists the *Hox* clusters, second column lists the total number of shared CNEs found in the syngnathid *Hox* clusters, third column lists the vertebrate CNEs, fourth column lists the actinopterygiian CNEs, fifth column lists the teleost CNEs, sixth column lists acanthomorpha CNEs, seventh column lists percomorph CNEs ("perc. 1"), eighth column lists percomorph CNEs that exclude mudskipper ("perc. 2").

Cluster	Total	vertebrate	actinop.	teleost	acantho.	perc. 1	perc. 2
HoxAa	87	32	5	1	30	5	14
HoxAb	23	7	3	1	1	8	3
HoxBa	60	19	9	5	19	3	5
HoxBb	26	6	3	4	11	0	2
HoxCa	70	21	16	8	17	5	3
HoxDa	44	26	11	3	3	1	0
HoxDb	20	4	1	3	11	1	0

Table 4.2: Number of CNEs annotated with the seven syngnathid *Hox* clusters. The number of syngnathid specific CNEs listed in the second column and the number of CNEs that are shared listed in the first column.

	Shared CNEs	Syngnathid CNEs
HoxAa cluster	87	42
HoxAb cluster	23	56
HoxBa cluster	60	50
HoxBb cluster	26	24
HoxCa cluster	70	88
HoxDa cluster	44	56
HoxDb cluster	20	72

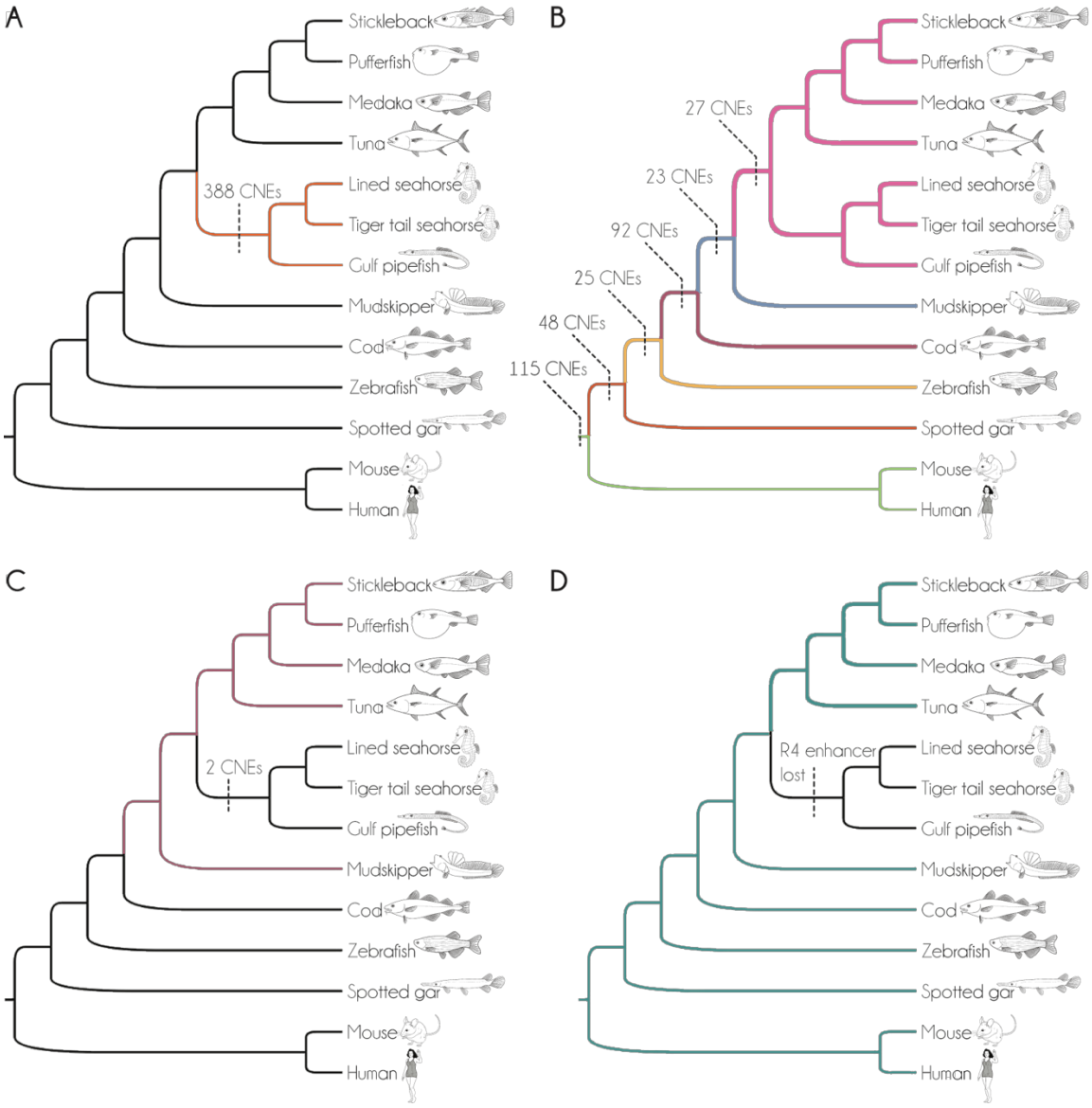


Figure 4.3: Distribution of CNEs cataloged within the syngnathid *Hox* clusters. Cladograms show evolutionary relationships between vertebrates included in the CNE analysis. a) there are 388 syngnathid specific CNEs, b) 330 CNEs shared at various levels of conservation with other species included in the analysis, c) 2 acanthomorph CNEs were uniquely lost in the syngnathid clade, d) 1 vertebrate CNE was uniquely lost in the syngnathid fish.

Syngnathids have relatively few losses of CNEs compared to other teleosts

From examining the VISTA plots, there were two CNEs that are shared between threespine stickleback, pufferfish, medaka and tuna, but are not present in the other lineages. One resided between CNE54 and CNE55—between *hoxa4a* and *hoxa3a* (Appendix B, S4.8D). The other was located between CNE27 and CNE29—between *hoxc11a* and *hoxc10a* (Appendix B, S4.12B). At this level of phylogenetic sampling, it is not possible to say that these two CNEs were uniquely or independently lost in the syngnathid clade or whether these CNEs arose after the syngnathid clade split from these percomorph fish.

There were five instances of syngnathid-specific losses of CNEs among the species examined. Of the five losses, these included two independent losses of a *Hox* cluster microRNA—*mir19b* and *mir10a*—that are reported to be lost convergently in other teleost. *Mir196b* was first described as an independent loss in the Gulf pipefish in (Small et al. 2016). There was an independent loss of *mir196b* previously reported in medaka (Hoegg et al. 2007). *Mir196b* was also missing the two seahorse species examined (Appendix B, Figure S4.10c). *Mir10a* was originally described as an independent loss in Gulf pipefish (Small et al. 2016). With the inclusion of cod and mudskipper in this analysis, *mir10a* also appeared to be missing independently in these lineages as well (Appendix B, Figure S4.11b).

There were also two syngnathid specific CNE losses in *HoxCa*—one between *hoxc8a* and *hoxc6a* and another between *hoxc4a* and *hoxc3a* (Appendix B, Figure S4.12c and S4.12d). Both of these CNEs were only found among the acanthomorph fish examined (cod, mudskipper, pufferfish, medaka, tuna, and threespine stickleback), and it is unknown whether these CNEs serve a functional role or are merely the result of neutral

sequence conservation (Figure 4.3c).

The fifth syngnathid specific missing element was located in the intron of *hoxa2b* in the *HoxAb* cluster of *Hox* genes. It was highly conserved in that it is present in all other species included in the VISTA analysis. This element is a known enhancer element for *hoxa2b* and will be further examined in the next chapter of this thesis (Figure 4.3d).

DISCUSSION

We investigated the changes to the *Hox* noncoding elements to characterize the developmental-genetic underpinnings of the striking evolution of derived characters present in the pipefish family. Changes to *cis*-regulatory elements are thought to be an important mechanism for evolutionary change (reviewed by (Carroll 2008)). Yet, we still only have a rudimentary understanding of the tempo of evolutionary changes in CNEs, particularly across closely related species within a family, because we are now only generating the whole genome sequence data necessary to examine the evolution of CNEs (Harmston, Baresic, and Lenhard 2013).

As a part of this study of *Hox* cluster noncoding elements, the *Hox* microRNAs were for the first time described in seahorses. *Hox* microRNAs for syngnathid fish were first described in the Gulf pipefish, but they were unannotated in the seahorse genomes (Small et al. 2016, Lin et al. 2016, Lin et al. 2017). The seahorse microRNA content matches that of the Gulf pipefish. This includes two convergent losses of *mir196b* and *mir10a*. The loss of *mir196b* is particularly intriguing. When the ortholog of this microRNA was targeted in a knockout experiment in mice, it led to extra rib-bearing vertebrae to develop. The loss of *mir196* led to a late activation of caudal *Hox* genes, and, as a result of this, led to axial extension (Wong et al. 2015). Similar phenotypes were

reported with knockdown experiments with *mir196b* using morpholinos in zebrafish, which included extra precaudal vertebrae to develop (He et al. 2011).

The syngnathid fish lineage has undergone an expansion of the vertebral column with the total number of vertebrae ranging from 31 to 94 depending on the lineage (Hoffman, Mobley, and Jones 2006). The position of where males carry their embryos is thought to be a selective pressure that results in a shift in relative proportion of tail and trunk vertebrae (Hoffman, Mobley, and Jones 2006). Perhaps the loss of *mir196b* was a factor in the evolutionary expansion of the vertebral column of these elongated syngnathid fish by leading to similar delayed activation of the caudal *Hox* genes.

In addition, we examined the evolution of the complete set of putative CNEs in the *Hox* Clusters of syngnathid fishes in comparison to other teleosts and vertebrates. Here we show that, similar to previous findings of teleost *Hox* gene content, the CNE elements have largely been conserved at the sequence level (Lee et al. 2010, Lee et al. 2006, Santini, Boore, and Meyer 2003, Chiu et al. 2002). I cataloged 718 putative CNEs with 388 of these elements that were specific to the Gulf pipefish, tiger tail and lined seahorse genomes. These units of conserved intergenic DNA should be considered putative CNEs because it is unknown whether these sequences serve a functional role or are merely the result of neutral sequence conservation. Subsequent studies will need to be done to examine whether these have been conserved functionally as well.

We found a few examples of unique losses. Two of these unique losses involve CNEs that are only found among the acanthomorph fish examined (cod, mudskipper, pufferfish, medaka, tuna, and threespine stickleback), and it is unknown whether these CNEs serve a functional role or are only the result of neutral sequence conservation

between the acanthomorph fish and subsequent studies will need to be done to examine whether these have a regulatory function.

In contrast to our general finding of conservation, we found one particularly interesting change in a *hoxa2b* regulatory element. This element has been well studied in other vertebrates and it increases expression of *hoxa2* in rhombomere 4 during development. This is the first reported loss of this element among fish, although it has been noted to have been lost in frogs (Tumpel et al. 2007). The knockout of this enhancer in *hoxa2b* in fugu led to differential expression of *hoxa2b* in rhombomere 4 (Tumpel et al. 2006).

Hoxa2 genes are known to send important patterning signals to pharyngeal arch 2 through rhombomere 4 during development (Minoux and Rijli 2010, Santagati and Rijli 2003, Parker, Bronner, and Krumlauf 2014). In fact, *hoxa2* has been previously described as a "master regulator of craniofacial programs and jaw formations" (McEllin et al. 2016). Inactivation of the *hoxa2* gene has led to various craniofacial phenotypes. Loss-of-function experiments of *Hoxa2* in mice, in *hoxa2a* zebrafish, and *hoxa2a* and *hoxa2b* in Nile tilapia led to duplications of jaw elements (Gendron-Maguire et al. 1993, Rijli et al. 1993, Santagati et al. 2005, Hunter and Prince 2002, Le Pabic, Scemama, and Stellwag 2010). Intriguingly, syngnathids have numerous modifications to their skulls (Leysen et al. 2010, Brown 2010, Kimmel, Small, and Knope 2017). Potentially, the loss of the *hoxa2b* enhancer element is tied to the highly modified skull in syngnathid fish. Detailed studies of the elements and expression of this the *hoxa2b* would further inform us of its role in syngnathid evolution.

CONCLUSION

We present the first examination of the *Hox* cluster CNEs for syngnathid fish. Among the three syngnathid species, there are many conserved noncoding sequences. These elements should be the subject of future investigations in order to distinguish whether any of these stretches of conserved intergenic sequence serve a regulatory function unique or novel to the syngnathid genomes. Additionally, we find the noncoding contents of the syngnathid *Hox* clusters broadly conserved with other vertebrates. We describe the loss of noncoding elements including a microRNA and an enhancer element in pipefish and seahorses. These elements are important for axial and cranial development.

BRIDGE

In the previous chapter I described the macroevolutionary patterns of conserved noncoding elements within the *Hox* cluster. Among these conserved noncoding sequences are putative *cis*-regulatory elements that regulate the expression of neighboring *Hox* genes. I completed a search for all conserved noncoding sequences present within the *Hox* clusters of the Gulf pipefish using a VISTA analysis by comparing the levels of intergenic sequence conservation between human, mouse, spotted gar, zebrafish, takifugu, threespine stickleback, two seahorse species, tuna, and medaka with Gulf pipefish using shuffle-LAGAN alignments.

In addition to the hundreds of putative CNEs present, I identified five Gulf pipefish and seahorse CNE losses. Two of these five losses were *mir10a* and *mir196b*, which I previously described as lost in Gulf pipefish. Another two of these unique losses involve putative CNEs that are only found among the acanthomorph fish examined (cod, mudskipper, pufferfish, medaka, tuna, and threespine stickleback), and it was unknown whether these CNEs serve a functional role or are only the result of neutral sequence conservation between the acanthomorph fish. One loss is identified as the rhombomere 4 enhancer for *hoxa2b*.

For my final experimental chapter of my thesis, I further researched the surprising loss of the *hoxa2b* enhancer element. We find that the binding element sequence motifs and spacing between the binding elements have been modified for this enhancer in syngnathid fish. Subsequently, we show expression of this gene in rhombomere 4 is lower relative to the surrounding rhombomeres in developing Gulf pipefish embryos, reflecting previously published functional tests for this enhancer.

CHAPTER V

EVOLUTIONARY LOSS OF A HINDBRAIN ENHANCER ELEMENT FOR *HOXA2B* IN SYNGNATHIDS MIMICS RESULTS OF FUNCTIONAL ASSAYS

INTRODUCTION

Despite expectations to the contrary by evolutionary biologists in the early 20th century, many developmental genetic pathways have remained surprisingly conserved across the different animal lineages over the course of metazoan evolution in terms of both sequence and function (Carroll, Grenier, and Weatherbee 2013, Duboule and Dollé 1989, McGinnis et al. 1984, Graham, Papalopulu, and Krumlauf 1989, Quiring et al. 1994, King and Wilson 1975). For example, *Hox* genes are a group of core developmental genes present in all animals that code for homeodomain transcription factors that are responsible for determining the body plan of an embryo along the anterior-posterior axis (Carroll 1995, Krumlauf 1994, McGinnis and Krumlauf 1992).

Following the initial description of *Hox* genes in *Drosophila melanogaster* in 1978, researchers discovered that *Hox* genes could be found in all animals examined (Lewis 1978, McGinnis and Krumlauf 1992, Duboule and Dollé 1989, Scott and Weiner 1984, McGinnis et al. 1984, Graham, Papalopulu, and Krumlauf 1989). The ancestral set of *Hox* genes consisted of a single cluster of genes, resulting from tandem duplications of an ancestral proto-*Hox* gene (Garcia-Fernandez 2005). Due to subsequent rounds of whole genome duplications, vertebrates have duplicate copies of the *Hox* complex (Pascual-Anaya et al. 2013). In vertebrates, tetrapods have four *Hox* gene clusters (denoted as *Hox* clusters A, B, C, and D), while teleost fish have eight clusters of *Hox*

genes due to the whole teleost genome duplication (*Hox* clusters Aa, Ab, Ba, Bb, Ca, Cb, Da, Db) (Amores et al. 1998) (Figure 1.1).

The vertebrate *Hox* genes are organized into 13 paralogous groups that span the gene clusters mentioned above (Scott 1992). *Hox* genes exhibit collinearity of expression along the body axis to confer positional identity information. This means that the order they appear in the genome reflects the order they are expressed along the anterior-posterior body axis (Gaunt 1988, Graham, Papalopulu, and Krumlauf 1989, Peterson et al. 1994, Duboule and Dollé 1989, Dekker et al. 1993, Godsave et al. 1994), with the vertebrate hindbrain expressing *Hox* genes in paralogous groups 1 through 4 during development (Alexander, Nolte, and Krumlauf 2009, Lumsden and Krumlauf 1996, Tumpel, Wiedemann, and Krumlauf 2009, Parker, Bronner, and Krumlauf 2016).

Despite the large amount of body plan diversity found in animals, *Hox* genes have maintained a great level of conservation throughout the animal kingdom both in terms of sequence and function (reviewed in (Gehring, Affolter, and Bürglin 1994, Burglin and Affolter 2016, Holland 2013)). This level of conservation first documented in *Hox* genes, and subsequently found in other core developmental gene families, has been hypothesized to occur because major changes in coding regions of *Hox* genes will be detrimental to the development of the organism. The proposed mechanism of conservation of antagonistic pleiotropy occurs because coding regions of these master developmental regulators have numerous downstream targets, and as a result, mutation in coding regions will be removed by selection because the consequences for phenotype and fitness will be so severe (Carroll 2008, Hoekstra and Coyne 2007). For example, one of the earliest homeotic mutations identified occurs because of alterations in the homeodomain of the *Antennapedia* gene and transforms antenna into legs (Struhl 1981).

The significant antagonistic pleiotropy observed in *Hox* coding region mutations led some researchers to hypothesize that mutations in such core developmental regulators are unlikely to contribute to evolution over short time scales (Carroll 2008, Hoekstra and Coyne 2007, Stern 2000). The relative paucity of nonsynonymous genetic variation in binding domains of *Hox* genes segregating in natural populations supports this argument. Alternatively, mutations of one or a small number of *cis*-regulatory elements (CRE) of *Hox* genes that cause shifts in expression of these conserved developmental genes may create traits that evolution can act upon while still working within the boundaries of developmental constraint (Wilkins 2002, Raff 2012). As a consequence, while mutations in CREs of *Hox* genes are also likely to exhibit antagonistic pleiotropy, it is predicted to be relatively lower than those in coding regions. As a result, we might predict that regulation of *Hox* genes may contribute to macroevolution—especially of body plan traits.

A key aspect of A-P axis formation in vertebrates is the repeated structures in the hindbrain called rhombomeres, which play key roles as units of anterior boundaries for overlapping patterns of expression of *Hox* genes. The hindbrain is organized into eight morphologically distinct rhombomeres (Kiecker and Lumsden 2005, Lumsden 2004). All jawed vertebrates have these repeated morphological units, which form through a progressive of segmentation during early development. Processes that include the formation of cytoskeletal barriers, cell adhesion and repulsion keep each rhombomere a distinctive unit. This leads to each rhombomere containing separate population of cells that follow different developmental pathways and neurons that are rhombomere specific (reviewed by (Parker, Bronner, and Krumlauf 2016)).

Rhombomeres are a source of cranial neural crest cells and are important regulators for craniofacial and nerve development (reviewed in (Parker, Bronner, and Krumlauf 2016)). Experimental manipulation of these anterior *Hox* genes have led to cranial phenotypes (Minoux and Rijli 2010, Santagati and Rijli 2003, Trainor and Krumlauf 2000, 2001).

Of particular interest, the *hoxa2* gene is expressed in the hindbrain during development (first described by (Prince and Lumsden 1994)). *Hoxa2* genes are known to send important patterning signals to pharyngeal arch 2 through rhombomere 4 during development via migratory streams of neural crest cells (Minoux and Rijli 2010, Santagati and Rijli 2003, Parker, Bronner, and Krumlauf 2014). Inactivation of the *hoxa2* gene has led to various craniofacial phenotypes. Loss-of-function experiments of *Hoxa2* in mice, in *hoxa2a* zebrafish, and *hoxa2a* and *hoxa2b* in Nile tilapia led to duplications of jaw elements (Gendron-Maguire et al. 1993, Rijli et al. 1993, Santagati et al. 2005, Hunter and Prince 2002, Le Pabic, Scemama, and Stellwag 2010). Gain-of-expression experiments with *hoxa2* led to repression of jaw formation in mice, *Xenopus*, and chicken (Grammatopoulos et al. 2000, Kitazawa et al. 2015, Pasqualetti et al. 2000).

Hoxa2 have several *cis*-regulatory factors that have been described over a series of studies (Maconochie et al. 1999, Maconochie et al. 2001, Nonchev, Maconochie, et al. 1996, Nonchev, Vesque, et al. 1996, McEllin et al. 2016, Tumpel et al. 2007, Tumpel et al. 2006, Parker, Bronner, and Krumlauf 2014). This described list currently includes a rhombomere 3/5 enhancer, a neural crest cell enhancer that is found upstream of the *hoxa2* gene, a rhombomere 4 enhancer element found in the intron and first exon of *Hoxa2*, and a rhombomere 2 enhancer element found in the second exon of *hoxa2* (Parker, Bronner, and Krumlauf 2016, Tumpel, Wiedemann, and Krumlauf 2009). The

knockout of this rhombomere 4 enhancer element in *hoxa2b* in fugu led to differential expression of *hoxa2b* in rhombomere 4 (Tumpel et al. 2006). In a previous study by Tumpel et al. (2007), various combinations of the binding site elements for this enhancer was knockout in chicken and mouse using site directed mutagenesis. They reported that using site directed mutagenesis on any one of these binding sites (with the exception of the fourth Pbx/Hox site located in exon 1 which not described at the time of the Tumpel et al. 2007 study) resulted in reduced efficiency of expression of *hoxa2* in rhombomere 4.

Due to their whole genome duplication, teleost fish typically have two copies of the *hoxa2* gene—called *hoxa2a* and *hoxa2b*. Expression of these two paralogs within the hindbrain varies among the different species of teleost. In zebrafish, *hoxa2a* is a pseudogene and *hoxa2b* is expressed in the pharyngeal arches 2–7 and rhombomeres 2–5. In striped bass, *hoxa2a* is known to be expressed in rhombomeres 2–7, and pharyngeal arch 2 and *hoxa2b* is expressed in rhombomeres 2–5 (Le Pabic et al. 2007, Scemama, Vernon, and Stellwag 2006). In Nile tilapia, *hoxa2a* and *hoxa2b* is expressed in pharyngeal arch 2 in the hindbrain during development (Le Pabic et al. 2007). In fugu, *hoxa2a* is expressed in rhombomere 1–2 and *hoxa2b* is expressed in rhombomeres 2–5 (Amores et al. 2004, McEllin et al. 2016, Tumpel et al. 2006). Examination of the *cis*-regulatory elements of *hoxa2* in highly derived fish lineages could be informative to understanding the evolution and function of this element.

In a previous paper (Fuiten et al. chapter IV) we documented that syngnathids are missing this rhombomere 4 enhancer element of *hoxA2*. The absence of this highly conserved and well described enhancer begged many questions. How is this enhancer modified in syngnathid fish? When was this enhancer lost? What are the possible downstream morphological consequences to the loss of this enhancer element?

The family Syngnathidae includes species of pipefish, seahorses, pipehorses, and seadragons. This charismatic teleost family displays a remarkable level of morphological diversity and phenotypic novelties such as a highly derived head and body plan, elongated body, prehensile tail, and the presence of male pregnancy (Small, Harlin-Cognato, and Jones 2013, Neutens et al. 2014, Bruner and Bartolino 2008). Syngnathid fishes are known for their highly divergent body plans, including the elongate form of many pipefishes and seadragons and the vertical body axis and reduced craniovertebral angle of seahorses (Herald 1959, Teske and Beheregaray 2009, Wilson and Rouse 2010). Derived characters such as leafy appendages, prehensile tails, bony body armor, male somatic brooding and loss of ribs, caudal, and pelvic fins are common across the family and in many cases have evolved independently in multiple lineages (Herald 1959, Wilson and Rouse 2010, Neutens et al. 2014).

In addition to variation in the body axis, syngnathid fish display a highly modified vertebrate skull which is an adaptation for suction feeding (Muller 1987, Muller and Osse 1984, de Lussanet and Muller 2007, Roos et al. 2009). This adaptive trait results from modified cranial bones in the ethmoid region and Meckel's cartilage. This includes the vomeral, mesethmoid, antorbitolacrimal, second infraorbital, quadrate, metapterygoid, preopercular, interopercular, and symplectic bones (Leysen et al. 2010). In the Gulf pipefish, the superior orientation of the mouth happens early in development prior to eight days post fertilization, while the elongation takes place relatively late in development between 12 to 17 days post fertilization after the bones have condensed into cartilage (Brown 2010). Whereas the morphology is well described for the adult crania of the pipefish, the genetic mechanism underlying the modification of the cranial bones remains unknown. Together, such extreme changes in body axis and craniofacial

structure beg the question as to whether modification of *Hox* gene expression may play a role.

In this study, we asked how this enhancer is modified in syngnathid fish, and to infer possible downstream morphological consequences to the loss of this enhancer element. We describe the binding sites of this element in syngnathid fish and the expression of the gene that it regulates during development. We find that the binding element sequence motifs and spacing between the binding elements have been modified for this enhancer. One binding motif has been lost and a second binding site has been partially lost. Subsequently, we show expression of this gene in rhombomere 4 is lower relative to the surrounding rhombomeres, reflecting previously published functional tests for this enhancer, and this change in expression is consistent with causing effects on the cranial neural crest. Our data support the hypothesis that natural mutations can occur in these deeply conserved pathways in ways potentially related to phenotypic diversity.

MATERIALS AND METHODS

Noncoding identification

Genomes used for comparison

CNEs identified using mVISTA analyses based on levels of sequence conservation within *Hox* clusters across *Gasterosteus aculeatus*, *Takifugu rubripes*, *Oryzias latipes*, *Thunnus orientalis*, *Hippocampus erectus*, *Hippocampus comes*, *Syngnathus scovelli*, *Boleophthalmus pectinirostris*, *Gadus morhua*, *Danio rerio*, *Lepisosteus oculatus*, *Mus musculus*, and *Homo sapiens* (Frazer et al. 2004, Mayor et al. 2000, Brudno, Do, et al. 2003, Brudno, Malde, et al. 2003). Sequences for *D. rerio*, *L. oculatus*, *M. musculus*, and *H. sapiens* were downloaded from Ensembl. *T. orientalis*

sequence was extracted from the *T. orientalis* genome ((Yasuike et al. 2016); http://nrifs.fra.affrc.go.jp/ResearchCenter/5_AG/genomes/Tuna_DNAmicroarray/index.html). *G. morhua* sequence was extracted from the *G. morhua* genome ((Torresen et al. 2017); https://figshare.com/articles/Transcript_and_genome_assemblies_of_Atlantic_cod/3408247). *S. scovelli* sequence was extracted from the *S. scovelli* genome ((Small et al. 2016); https://creskolab.uoregon.edu/pipefish/). The *H. erectus* sequence was extracted from the *H. erectus* genome ((Lin et al. 2017); NCBI with the project accession PRJNA347499). The *H. comes* sequence was extracted from the *H. comes* genome ((Lin et al. 2016); NCBI with the project accession PRJNA314292). The *B. pectinirostris* sequence was extracted from the *B. pectinirostris* genome ((You et al. 2014); NCBI with the project accession PRJNA232434). The *T. rubripes* sequences were retrieved from Genbank ((Lee et al. 2006); Genbank accessions DQ481663–9). The *O. latipes* sequences were retrieved from Genbank ((Kurosawa et al. 2006); AB232918–24). The *G. aculeatus* sequences were from BAC clones, which were make available by Angel Amores. Sequences were softmasked using RepeatMasker.

Noncoding VISTA analysis

G. aculeatus and *S. scovelli* was set as the reference sequence for the VISTA analysis. Alignment of each sequence from these species were aligned using the shuffle-LAGAN algorithm and the LAGAN algorithm through the mVISTA website with Minimum conservation identity set to 65% and Minimum length for a CNS set to 50.

All conserved noncoding sequences annotated within the *S. scovelli Hox* clusters were queried against the NCBI NR database to identify coding exons, against RFAM,

refseq_rna, and the miRBase Sequence Databases (Release 21) for mature miRNA chordate sequences and miRNA chordate hairpins (downloaded from miRBase). BBMapSkimmer was used to query against the miRBase Sequence Databases in order to identify RNA genes. Kmer index size was set to 7, max indel set to 0, approximate minimum alignment identity set to 0.50, secondary site score ratio set to 0.25, behavior on ambiguously-mapped reads set to retain all top-scoring sites, and maximum number of total alignments to print per read set to 4 million.

Additional syngnathid taxonomic sampling

In addition to the *Syngnathus scovelli, Hippocampus erectus*, and *H.* comes genomic sequences, degenerate primers were designed and used to sequence the *hoxa2b* enhancer region for the dwarf seahorse (*H. zostrae*), the messmate pipefish (*Corythoichthys haematopterus*), bluestripe pipefish (*Doryrhamphus excisus*), sculptured pipefish (*Choeroichthys sculptus*), and the robust ghost pipefish (*Solenostomus cyanopterus*) (Table 5.1). The dwarf seahorse and messmate pipefish bring additional taxonomic sampling from the Syngnathinae subfamily of Syngnathidae. The sculptured and bluestripe pipefish are members of the Nerophinae subfamily of Syngnathidae. In order to investigate when the loss of this enhancer element happened and whether its degenerated state unique is to syngnathid fish, the degenerate primers were designed and used to sequence the *hoxa2b* enhancer region for a species from the immediate outgroup to Syngnathidae, the ghost pipefish (genus *Solenostomus*). This additional taxonomic sampling provided further insight into the loss of this enhancer element in this teleost fish family (Figure 5.1).

Table 5.1: Degenerate primer pairs used on syngnathid species for *hoxa2b*.

species	Forward Primer	Reverse Primer
robust ghost pipefish	TGGCCTAGAAAGYGGTTTTATCAA	TACTTGTTGAAGTGGAACTCTT
messmate pipefish	TGGCCTAGAAAGYGGTTTTATCAA	AAATCCAACMAGGMGGCTATCT
dwarf seahorse	GGAGGAGATGAATTACGCATT	TACTTGTTGAAGTGGAACTCTT
sculptured pipefish	TGGCCTAGAAAGYGGTTTTATCAA	TACTTGTTGAAGTGGAACTCTT
bluestripe pipefish	TGGCCTAGAAAGYGGTTTTATCAA	TACTTGTTGAAGTGGAACTCTT

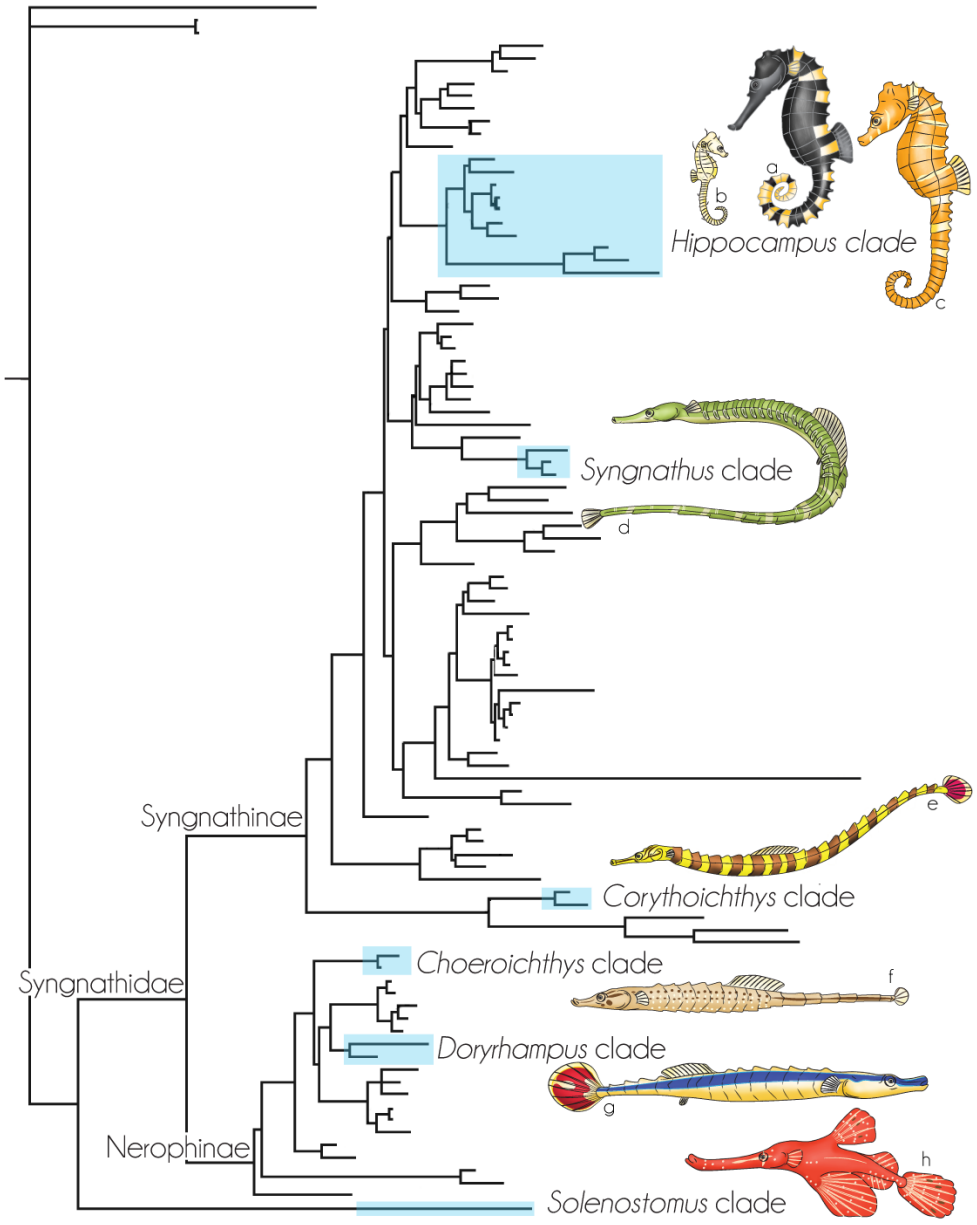


Figure 5.1: Syngnathid phylogeny, with samples used in this study marked. Illustrations depict representative species: (a) *Hippocampus zostrae* (b) *H. comes* (c) *H. erectus* (d) *Syngnathus scovelli* (e) *Corythoichthys haematopterus* (f) *Choeroichthys* sculptus (g) *Doryrhamphus excisus* (h) *Solenostomus cyanopterus.* Syngnathidae is divided into two subfamilies—the tail brooding Syngnathinae and the trunk brooding Nerophinae. Cladogram based on molecular phylogeny published by Hamilton et al. 2017.

Sequence alignments and identification of enhancer binding sites

Hoxa2, *hoxa2b*, and *hoxa2a* sequences from coelacanth (*Latimeria chalumnae*), anole (*Anolis carolinensis*), chicken (*Gallus gallus*), *D. rerio*, *L. oculatus*, *M. musculus*, and *H. sapiens* were downloaded from Ensembl. The Australian ghostshark (*Callorhinchus milii*) sequence was retrieved from Genbank. The tamar wallaby (*Notamacropus eugenii*) sequence was retrieved from Genbank. The *T. rubripes* sequences were retrieved from Genbank ((Lee et al. 2006); Genbank accessions DQ481663–9). The *O. latipes* sequences were retrieved from Genbank ((Kurosawa et al. 2006); AB232918–24). The *G. aculeatus* sequences were from BAC clones, which were make available by Angel Amores. *T. orientalis* sequence was extracted from the *T. orientalis* sequence was extracted from the *T. orientalis* genome ((Yasuike et al. 2016); http://nrifs.fra.affrc.go.jp/ResearchCenter/5_AG/genomes/Tuna_DNAmicroarray/index.html). *G. morhua* sequence was extracted from the *G. morhua* genome ((Torresen et al. 2017); https://figshare.com/articles/Transcript_and_genome_assemblies_of_Atlantic_cod/3408247). *S. scovelli* sequence was extracted from the *S. scovelli* genome ((Small et al. 2016); https://creskolab.uoregon.edu/pipefish/). The *H. erectus* sequence was extracted from the *H. erectus* genome ((Lin et al. 2017); NCBI with the project accession PRJNA347499). The *H. comes* sequence was extracted from the *H. comes* genome ((Lin et al. 2016); NCBI with the project accession PRJNA314292). The *B. pectinirostris* sequence was extracted from the *B. pectinirostris* genome ((You et al. 2014); NCBI with the project accession PRJNA232434). Primers were designed and used to obtain the *hoxa2b* sequences from the robust ghost pipefish (*Solenostomus cyanopterus*), messmate pipefish (*Corythoichthys haematopterus*), bluestripe pipefish (*Doryrhamphus excisus*), sculptured

pipefish (*Choeroichthys* sculptus) and dwarf seahorse (*Hippocampus zostrae*). Tissue samples from the robust ghost pipefish, messmate pipefish, and dwarf seahorse were obtained from the Adam Jones Lab at the University of Idaho. Tissue samples from the bluestripe pipefish (KU 7147) and sculptured pipefish (KU 5054) were obtained from the University of Kansas fish tissue collection. The sequences were aligned using MUSCLE through the Geneious software (Edgar 2004). Alignments were corrected manually.

Binding site sequences for Pbx/Hox and Prep/Meis were obtained from (Tumpel et al. 2007, Berthelsen et al. 1998, Ferretti et al. 2005, Ferretti et al. 2000). The binding motifs identified in a previous study for *hoxa2* in human, chicken, mouse, baboon, rat, bat, dog, coelacanth, shark, and for *hoxa2b* in zebrafish, fugu, and medaka, and for *hoxa2a* in fugu and medaka were used as guides in aligning and identifying the Pbx/Hox and Prep/Meis binding sites in the species included in this study.

Cloning and synthesis of riboprobes

Antisense riboprobes were made from syngnathid clones. Genes sequences for targeted genes were obtained from the Gulf pipefish genome. For design of the *in situ* probe, functional domains were identified on targeted gene, and the probe was designed around those sequences. Amplified fragments were cloned into TOPO PCR-IV vector (Invitrogen) and the inserts were confirmed by Sanger sequencing. The resulting plasmids were linearized with the either NotI or SpeI restriction enzymes, depending on insert orientation. Antisensedigoxigenin (DIG)-labeled RNA probes were prepared using DIG-RNA labeling mix (Fermentas), Ribolock RNase inhibitor (Fermentas) and either T7 RNA polymerase or T3 RNA polymerase (depending on insert orientation) and incubating at 37°C for 2 hours. The plasmid was digested using DNase I, RNAse-free

(Fermentas) and a portion of the resultant RNA was run on a gel (1.0% agarose, 10 cm gel, 1.0X TBE, 110 V) to confirm the synthesis of adequate probe. Probe concentration was also measured using Quantit RNA broad range assay kit on a Qubit fluorometer (Invitrogen). For *krox20a*, the probe sequence used was 5'-gcgcctccttgtacgcacgcgcacctccacccgccctcgtcgtacacgtgcatcagtgacgtgtaccaggaatcctctgatgaggttacctggccgtacccacctgcagcgcggtgacttatcacatggcgccagcctataactcggcgccaaaagccccgctggtggctgactacggcgtgggggggagtctacgccccacaggccaccttcccggaccggaagtcagtggcggcgtacgccttggactccctccgcgtggcccctccgctcacacc-3'. For hoxa2a, the probe sequence used was 5'-tggaatccacgcagcaggtccacaatagcagctcggcgagctttgctgctgcaccgctgaacagcaatgagaaaaatctgaaacattttcccaacccgtcacccactgttcccggctgcgtgtcaacaatgggcccaggctcggcatccgtgccggacaatggcgacagtccccccagctttggatgtttctatacacgacttccaagctttctcgtcggattcctgcttgcaactmtccgacgctgcctcgccgagcttgtctgaatcgctggacagtcccgtgg-3'. For hxa2b, the probe sequence used was 5'-gcgaaggacctttggaagagcagccagccaaggggcagaggtatttccaggaaaattgtttcaattcacaacattgtcctaatagccacaatggsgacaatgattcgactttgtgcataagtgagaaaaatgccaaacatcttccggactgcgctcccaccacggctcccttctgtgcgcccgaaataggcccggagaataatytttcccacgtctcgcacagtgaatactccccggatttggacgcctctttgcgggagcttcctcgagcatcctcgttctcgcaagactggtccgattcaactccgct-3'.

Whole mount *in situ* hybridization analysis

Embryos at various days post fertilization were extracted from the paternal brood pouch, anesthetized in 0.017% Tricaine-S, fixed in 4%PFA/PBS and stored in methanol. Whole-mount *in situ* hybridization analyses were performed as described in Thisse and Thisse (2008). One to five embryos from each stage were used in hybridization with each probe (*hoxa2b*, *hoxa2a*, *krox20a*). Hybridized specimens were placed in 50%

glycerol/50% PBSTw, mounted onto slides and photographed on a compound microscope.

Collection and maintenance of pipefish

Adult pipefish were collected in Tampa Bay, Florida on May 5, 2017. Breeding tanks were set up at the University of Tampa. Pregnant male pipefish were collected at 1, 2, 3, 4, 5, 6 dpf. Additionally, wild caught pregnant male Gulf pipefish were collected and euthanized. Threespine stickleback were raised at the University of Oregon and collected at various stages post fertilization. Embryos were euthanized with MS-222, and then fixed in 4% paraformaldehyde PFA either overnight at 4°C or for 5 hours at room temperature and stored in methanol. Experimental research conducted on these animals was performed according to protocols approved by the Institutional Animal Care and Use Committees (IACUC) at the University of Oregon.

RESULTS

A unique loss of a *hoxa2b* enhancer is shared across syngnathid fish

Previously we showed a loss of the *hoxa2b* R4 enhancer in *S. scovelli*. We addressed the question whether that was unique to pipefish or shared across syngnathids. We used *Hippocampus erectus, H. comes* and *Syngnathus scovelli* as the syngnathid representatives and compared their CNE content to percomorph teleost fish (*Gasterosteus aculeatus, Takifugu rubripes*, *Oryzias latipes*, *Thunnus orientalis*), non-percomorph teleost fish (*Boleophthalmus pectinirostris*, *Gadus morhua*, *Danio rerio*), non-teleost fish (*Lepisosteus oculatus)*, and two non-fish vertebrates (*Mus musculus* and *Homo sapiens*) in the *Hox* clusters.

From examining the VISTA plots, a shared loss of a highly conserved noncoding element among the included syngnathid species was found (Figure 5.2). This missing element is located in the intron of *hoxa2b* in the *HoxAb* cluster of *Hox* genes. It was highly conserved in that it was present in all other species included in the VISTA analysis.

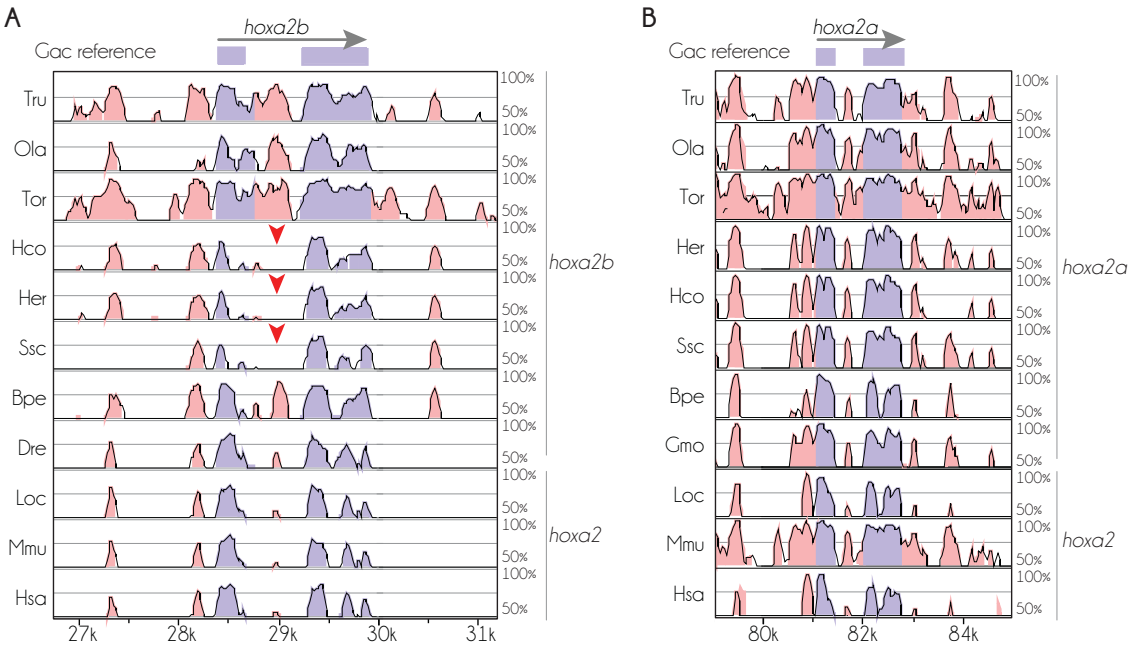


Figure 5.2: A conserved non-coding element is not detectable in the pipefish *HoxAb* cluster. a) One CNE present in other teleosts and mammals is missing from the intron of *hoxa2b* in the *S. scovelli, H. comes* and *H. erectus* assemblies (red arrows). b) Syngnathids are not missing CNEs from the intron of *hoxa2a* in the *S. scovelli, H. comes* and *H. erectus* assemblies. Exons are highlighted in blue, CNEs in pink. The reference, Gac, is stickleback; Tru, pufferfish; Ola, medaka; Tor, tuna; Hco, *tiger tail* seahorse; Her, lined seahorse; Ssc, pipefish; Bpe, mudskipper; Gmo, cod; Dre, zebrafish; Loc, spotted gar; Mmu, mouse; Hsa, human. Red arrows indicate missing CNE in syngnathid fish.

The CNE missing in syngnathid species is a previously described enhancer element for the *hoxa2b* gene in teleost fish. This enhancer element increases expression of *hoxa2* in rhombomere 4 during development. Teleost fish have two copies of *hoxa2*

called *hoxa2a* and *hoxa2b*. Previous research showed the *hoxa2a* paralog in fish, despite exhibiting the binding motifs of this enhancer, apparently does not drive expression of the *hoxa2a* gene in the hindbrain (McEllin et al. 2016).

A large degree of sequence changes to the Pbx/Hox syngnathid binding sites

The enhancer element consists of four Pbx/Hox binding sites and one Prep/Meis binding site. One of the four Pbx/Hox binding sites are located in the first exon of *hoxa2* and *hoxa2b* genes. The remaining binding sites are located in the intron of the *hoxa2* and *hoxa2b* genes (Figure 5.3) (Parker, Bronner, and Krumlauf 2014, Tumpel et al. 2006, Tumpel et al. 2007).

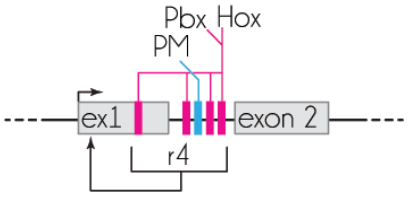


Figure 5.3: Rhombomeric regulatory modules in *hoxa2*. Pink boxes represent the Pbx/Hox binding sites and the blue box represents the Prep/Meis binding site. The gray boxes represent the exons.

In order to further examine the degree of conservation of this binding site among vertebrates, the enhancer element binding motifs were examined across Vertebrata using the Australian ghostshark, coelacanth, anole, chicken, tamar wallaby, human, mouse, spotted gar, zebrafish, takifugu, medaka, threespine stickleback, Pacific bluefin tuna, and mud skipper, along with syngnathid species (Figure 5.3, Table 5.2). Pbx/Hox dimers recognize the sequence 5'-TGATNNAT-3', with the *Hox* proteins recognizing the 5'-NNAT-3'. The Pbx proteins bind to the 5' part of the 5'-TGATNN-3' sequence, and the *Hox* protein contacts the NNAT sequence motif (Ferretti et al. 2005, Knoepfler, Lu, and Kamps 1996). The two NN bases tend to vary depending on the *Hox* gene that dimerizes with the Pbx (Chan et al. 1997, Chang et al. 1995, Knoepfler, Lu, and Kamps 1996, Manzanares et al. 2001).

Table 5.2: Binding site sequences for *hoxa2* enhancer element. Purple columns show Pbx/Hox binding sites. Pbx/Hox4 is found in exon 1 of hoxa2 genes while the other Pbx/Hox are located in the intron. Red letters indicate base pair changes that deviate from the consensus.

Species	paralog	Pbx/Hox4	Pbx/Hox1	Prep/Meis	Pbx/Hox2	Pbx/Hox3
Australian ghostshark	*hoxa2*	TGATACAT	TGATTTAT	TGACAG	TGATAGAT	TGATGCAG
coelacanth	*hoxa2*	TGATACAT	TGATTTAT	TGACAG	TGATAGAT	TGATGCAT
anole	*hoxa2*	TGATACAT	TGATTTAT	TGACAG	TGATAGAT	TGATGCAT
chicken	*hoxa2*	TGATACAT	TGATTTAT	TGACAG	TGATAGAT	TGATGCAT
tamar wallaby	*hoxa2*	TGATACAT	TGATTTAT	TGACAG	TGATAGAT	TGATGCAT
human	*HOXA2*	TGATACAT	TGATTTAT	TGACAG	TGATAGAT	TGACGCAT
mouse	*Hoxa2*	TGATACAT	TGATTTAT	TGACAG	TGATAGAC	TGATGCAT
spotted gar	*hoxa2*	TGATACAT	TGATTTAT	TGACAG	TGATAGAT	TGGCGCAT
zebrafish	*hoxa2b*	TGATGCAT	TGATATGT	TGACAG	TGATAGAT	TGGCGTGT
takifugu	*hoxa2b*	TGATGCGT	TGATTTAA	TGACAG	TGATAGAT	TGGCATGT
medaka	*hoxa2b*	TGATGCTT	TGATTTAA	TGACAG	TGATAGAT	TGGCACGT
threespine stickleback	*hoxa2b*	TGATGCTT	TGATTTAA	TGACAG	TGATAGAT	TGGCATGT
Pacific bluefin tuna	*hoxa2b*	TGATGCAT	TGATTTAA	TGACAG	TGATAAGG	TGGCATGT
mud skipper	*hoxa2b*	TGATGCGT	TGAATCAT	TGACAG	TGATAGAT	TGGCATGT
ghost pipefish	*hoxa2b*	CGATGCGT	TGATTTAG	TGACAG	TGATCGAT	TGGATCTA
bluestripe pipefish	*hoxa2b*	CCATGCTT	CGGATTTT	—ACAA	TGATGGAT	TGGC——
sculptured pipefish	*hoxa2b*	CGATGCCT	TGGATTTT	—ACAA	TGATGGAT	TGGC——
messmate pipefish	*hoxa2b*	CGATGCCT	GGATTTGG	———	TGATGGAT	TGGC——
Gulf pipefish	*hoxa2b*	CGATGCGT	AGATTTGG	———	TGATGGAT	TGGC——
tiger tail seahorse	*hoxa2b*	CGATGCGT	TGATATGT	———	TGATGGAT	TGGC——
lined seahorse	*hoxa2b*	CGATGCCT	TGATATGT	———	TGATGGAT	TGGC——
dwarf seahorse	*hoxa2b*	CGATGCCT	TGAAATGT	———	TGATGGAT	TGGC——

We found that teleost fish have the 5'-TGAT-3' motif in the Pbx/Hox 1 site, with the exception of the mudskipper, the bluestripe pipefish, the sculptured pipefish, the messmate pipefish, the Gulf pipefish, and the dwarf seahorse. Teleost fish, except for the mudskipper, did not have the NNAT sequence motif. The Pbx/Hox 2 have stayed the most conserved relative to the other Pbx/Hox binding sites for this enhancer. The binding sequence had stayed 5'-TGATAGAT-3' with the exception of mouse, that had 5'-TGATAGAC-3' and Pacific bluefin tuna which had 5'-TGATAAGG-3'. The ghost pipefish had 5'-TGATCGAT-3' and the syngnathid species all had 5'-TGATGGAT-3'. The Pbx/Hox 3 binding site displayed the most sequence variation. Teleost fish did not

follow the 5'-TGAT-3' or the 5'-NNAT-3' rules established by (Ferretti et al. 2005). Based on alignments, the second half of the binding sequence appeared to have been lost in the syngnathid species. Teleost fish had the 5'-TGAT-3' motif in Pbx/Hox 4 except for ghost pipefish and Syngnathidae fish that did not have 5'-TGAT-3' motif. All teleost fish, with the exception of the Pacific bluefin tuna, did not have the 5'-NNAT-3' sequence motif.

Loss of Prep/Meis in syngnathid species

We found that the Prep/Meis binding site had stayed conserved across taxa examined, including in the ghost pipefish, with the exception of the syngnathid species. Members of the Syngnathinae subfamily (*Hippocampus erectus, H. comes*, *H.* zostrae, *Corythoichthys haematopterus*, and *Syngnathus scovelli*) were missing the Prep/Meis binding site. Based on alignments, it appeared that the two species from the Nerophinae subfamily, *Doryrhamphus excisus* and *Choeroichthys sculptus* had only the "ACA" nucleotides remaining from this binding site (Figure 5.4, Table 5.1).

Truncated spacing between binding sites in the syngnathid binding sites

The spacing of the binding elements have also been modified in the syngnathid lineages. Overall, the intron was shorter in syngnathid lineages relative the other vertebrates included for comparison (Table 5.3). The intron lengths spanned from 924 bases in *Anolis carolinensis* to 417 bases in *Takifugu rubripes*. The intron length in syngnathid species were all less than 275 bases.

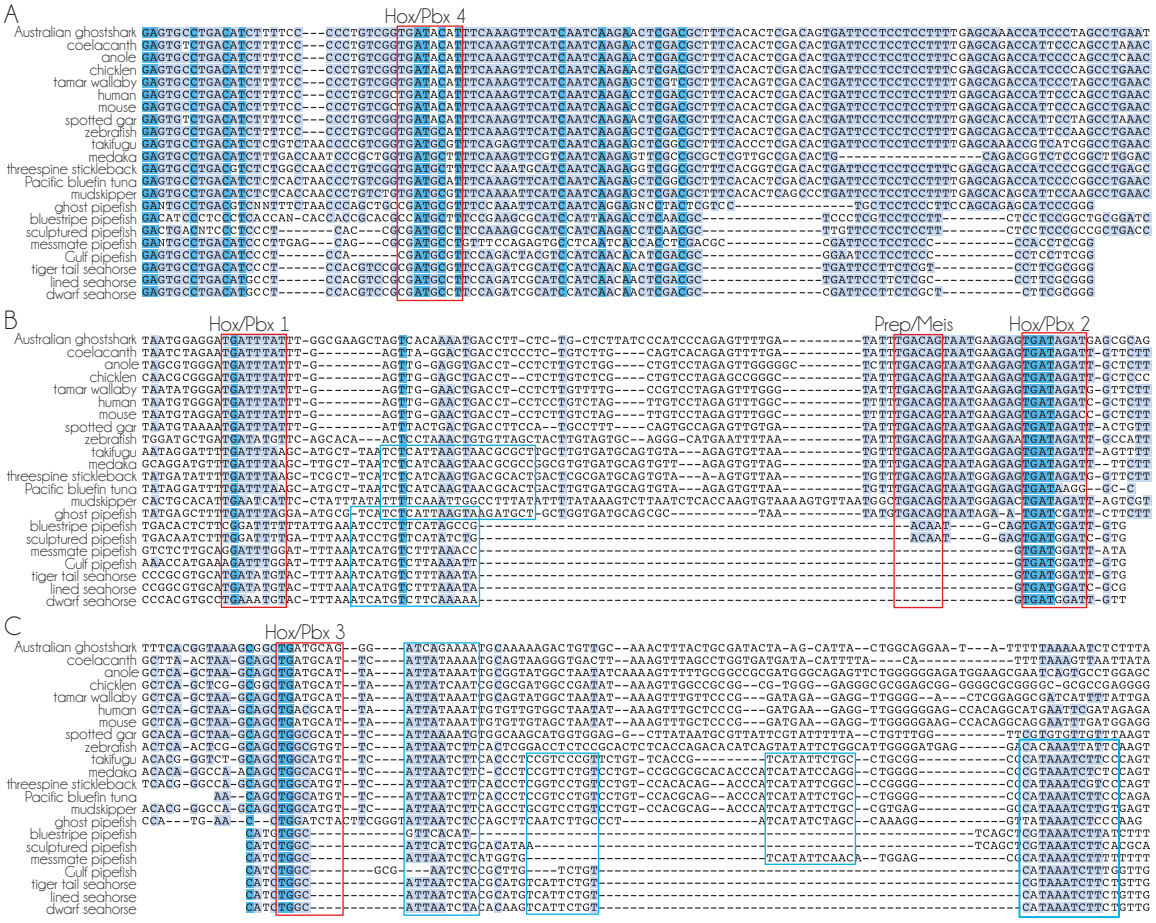


Figure 5.4: Sequence alignment of *hoxa2* rhombomere 4 enhancer across Vertebrata. Shown are the sequence alignments around the four Pbx/Hox and one Prep/Meis binding sites (red boxes) for the r4 *hoxa2* enhancer. The hoxa2 sequence was used for Australian ghostshark, coelacanth, anole, chicken, tamar wallaby, human, mouse, and spotted gar. The *hoxa2b* sequence was used for the rest of the included taxa. a) alignments surrounding the Pbx/Hox4 binding site. This binding site is upstream to the other binding sites and is located in the first exon of *hoxa2*/*hoxa2b.* b) alignments surrounding the Pbx/Hox1, Prep/Meis and Pbx/Hox2 binding sites located in the intron of *hoxa2*/*hoxa2b*. c) is an immediate continuation of the alignment starting in b) and includes the Pbx/Hox3 binding site alignment. It is also located within the *hoxa2*/*hoxa2b* intron. Blue boxes highlight key areas of sequence across different subsets of the taxa.

The spacing between each of the binding sites was also shorter in the syngnathid species relative to the other species (Table 5.3). In vertebrates, the spacing between Pbx/Hox binding sites 1 to 2 was between 66 and 110, except in syngnathids when it shortened to 33 and 32 bases in the bluestripe and sculptured pipefish and to 24 bases in all other syngnathids examined. The nucleotides between Pbx/Hox binding sites 2 to 3 was consistently at 22 bases, with the exception of medaka at 21, the Australian ghostshark at 16 and the Pacific bluefin tuna at nine. Syngnathids had the spacing of eight bases. Overall the distance between the first binding site of this enhancer element to the last binding site of this enhancer element typically ranged from 682 to 384, with the exception of the anole having the distance of 924 bases. The syngnathids included in this analysis had a spacing of 267 to 338 (Table 5.3).

Ghost pipefish had a space of 66 bases between Pbx/Hox binding sites 1 and 2, while the other syngnathid fish had a space of 24 bases. The nucleotides between Pbx/Hox binding sites 2 to 3 was at 16 bases with the ghost pipefish and the other syngnathid fish had a spacing of eight bases. Overall the distance between the first binding site of this enhancer element to the last binding site of this enhancer element for the syngnathids included in this analysis ranged from 267 to 289 bases, with the exception of the ghost pipefish which had a longer spacing of 356 bases (Table 5.3).

Table 5.3: Binding site spacing for *hoxa2* enhancer element. PH4 = Pbx/Hox4, PH1 = Pbx/Hox1, PH2 = Pbx/Hox2, PH3 = Pbx/Hox3, and PM = Prep/Meis binding sites. Intron length for *hoxa2* or *hoxa2b* genes is recorded in last column.

Species	paralog	PH4 to PH1	PH1 to PH2	PH2 to PH3	PH1 to PM	PM to PH2	intron length
Australian ghostshark	*hoxa2*	*417*	78	25	62	10	533
coelacanth	*hoxa2*	*398*	67	22	51	10	478
anole	*hoxa2*	*682*	69	22	53	10	924
chicken	*hoxa2*	*386*	66	22	50	10	644
tamar wallaby	*hoxa2*	*581*	67	22	51	10	658
human	*HOXA2*	*553*	67	22	51	10	644
mouse	*Hoxa2*	*537*	67	22	51	10	640
spotted gar	*Hoxa2*	*443*	67	22	51	10	535
American eel	*hoxa2b*	*393*	110	22	94	10	537
zebrafish	*hoxa2b*	*491*	75	22	59	10	597
takifugu	*hoxa2b*	*395*	77	22	61	10	417
medaka	*hoxa2b*	*384*	77	21	61	10	437
threespine stickleback	*hoxa2b*	*404*	76	22	60	10	473
Pacific bluefin tuna	*hoxa2b*	*431*	77	9	61	10	455
mud skipper	*hoxa2b*	*356*	91	22	75	10	418
robust ghost pipefish	*hoxa2b*	*305*	66	16	52	8	350
bluestripe pipefish	*hoxa2b*	*325*	33	8	24	5	218
sculptured pipefish	*hoxa2b*	*338*	32	8	23	5	254
messmate pipefish	*hoxa2b*	*267*	24	8	------	------	274
Gulf pipefish	*hoxa2b*	*279*	24	8	------	------	257
tiger tail seahorse	*hoxa2b*	*287*	24	8	------	-------	265
lined seahorse	*hoxa2b*	*281*	24	8	------	-------	257
dwarf seahorse	*hoxa2b*	*289*	24	8	------	-------	263

Loss of Prep/Meis and further space shortening happened after ghost pipefish split from the rest of the syngnathid clade

We found that the missing Prep/Meis binding site and modified state of the Pbx/Hox binding sites of this enhancer element was also found in species sampled from both subfamilies of Syngnathidae. We concluded that this particular extreme modification of the *hoxa2b* enhancer is mostly likely shared across the family of Syngnathidae (Figure 5.1).

We found that the robust ghost pipefish, *Solenostomus cyanopterus,* had all five binding sites for this enhancer element and an intermediately sized intron of 350 bases (Tables 5.2 and Table 5.3). This can be interpreted as that the loss of the Prep/Meis binding site happened after the ghost pipefish diverged from Syngnathidae clade (Figure 5.1). The spacing of the motifs were already shortening before ghost pipefish split from Syngnathidae, but more extreme shortening of the binding site spacing after ghost pipefish diverged from Syngnathidae.

Pattern of expression of *hoxa2b* in rhombomere 4 in syngnathid is similar to expression in knockout studies

In a previous study by Tumpel et al. (2007), various combinations of the binding site elements for this enhancer was knocked out in chicken and mouse using site directed mutagenesis. Based on this study, we hypothesized that the modification and reduction of this enhancer element in syngnathid fish would result in reduced expression of *hoxa2b* in rhombomere 4.

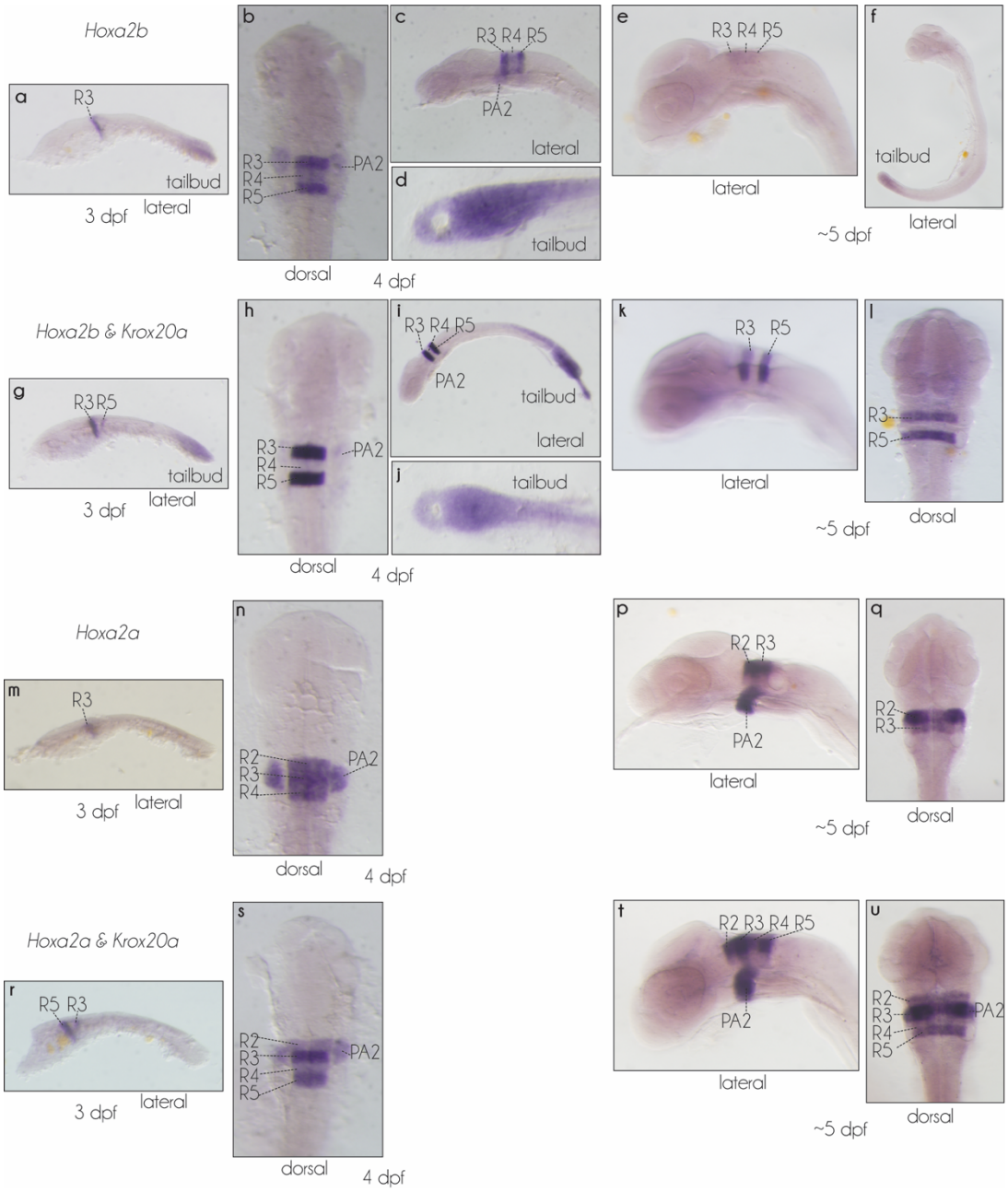
We examined the expression of *hoxa2a* and *hoxa2b* over development in the Gulf pipefish. We found that *hoxa2b* was expressed in the hindbrain and in the tailbud during

development. At three days post fertilization, *hoxa2b* was expressed in rhombomere 3 and in the tailbud. At four days post fertilization, *hoxa2b* was expressed in rhombomeres 3, 4, and 5, in the pharyngeal arch 2, and in the tailbud. There is less expression of *hoxa2b* in rhombomere 4 relative to rhombomeres 3 and 5. At five days post fertilization, *hoxa2b* is expressed in rhombomeres 3, 4, and 5 and in the tailbud (Figure 5.5).

We found that *hoxa2a* was expressed in the hindbrain during development. At four and five days post fertilization, *hoxa2a* is expressed in rhombomeres 2, 3, and 4 and pharyngeal arch 2 (Figure 5.5). Remarkably, this follows predictions based on functional tests previously published by Tumpel et al. 2007, expression of *hoxa2b* appears to be reduced in rhombomere 4 relative to neighboring rhombomeres 3 and 5 in Gulf pipefish.

In zebrafish, *hoxa2a* is a pseudogene and *hoxa2b* is expressed in the pharyngeal arches 2–7 and rhombomeres 2–5. In striped bass, *hoxa2a* is expressed in rhombomeres 2–7, and pharyngeal arch 2 and *hoxa2b* is expressed in rhombomeres 2–5 (Scemama, Vernon, and Stellwag 2006). In fugu, *hoxa2a* is expressed in rhombomere 1–2 and *hoxa2b* is expressed in rhombomeres 2–5 (Amores et al. 2004, McEllin et al. 2016, Tumpel et al. 2006).

Figure 5.5: (next page) *In situ* expression of *hoxa2a* and *hoxa2b* in Gulf pipefish. Images a–f show expression of *hoxa2b* in Gulf pipefish embryos. Images g–l show expression of *hoxa2b* in Gulf pipefish embryos co-stained for *krox20a*. Images m–q show expression of *hoxa2a* in Gulf pipefish embryos. Images r–u show expression of *hoxa2a* in Gulf pipefish embryos co-stained for *krox20a*. (a) *hoxa2b* 3dpf lateral; (b) *hoxa2b* 4dpf dorsal; (c) *hoxa2b* 4dpf right lateral; (d) *hoxa2b* 4dpf tailbud; (e) *hoxa2b* ~5dpf left lateral; (f) *hoxa2b* ~5dpf full embryo lateral; (g) *hoxa2b* with *krox20a* 3dpf lateral; (h) *hoxa2b* with *krox20a* 4dpf dorsal; (i) *hoxa2b* with *krox20a* 4dpf right lateral; (j) *hoxa2b* with *krox20a* 4dpf tailbud; (k) *hoxa2b* with *krox20a* ~5dpf left lateral; (l) *hoxa2b* with *krox20a* ~5dpf dorsal; (m) *hoxa2a* 3dpf lateral; (n) *hoxa2a* 4dpf dorsal; (p) *hoxa2a* ~5dpf left lateral; (q) *hoxa2a* ~5dpf dorsal; (r) *hoxa2a* with *krox20a* 3dpf lateral; (s) *hoxa2a* with *krox20a* 4dpf dorsal; (t) *hoxa2a* with *krox20a* ~5dpf left lateral; (u) *hoxa2a* with *krox20a* ~5dpf dorsal. Krox20a marks rhombomeres 3 and 5. R3 = Rhombomere 3, R5 = Rhombomere 5, PA2 = Pharyngeal Arch 2.


Hoxa2b
a
R3
tailbud
3 dpf
lateral
b
R3
R4
R5
PA2
dorsal
4 dpf
c
R3 R4 R5
PA2
lateral
d
tailbud
e
R3 R4 R5
lateral
f
tailbud
~5 dpf
lateral
Hoxa2b & Krox20a
g
R3 R5
tailbud
3 dpf
lateral
h
R3
R4
R5
PA2
dorsal
4 dpf
i
R3 R4 R5
PA2
tailbud
lateral
j
tailbud
k
R3 R5
lateral
l
R3
R5
~5 dpf
dorsal
Hoxa2a
m
R3
3 dpf
lateral
n
R2
R3
R4
PA2
dorsal
4 dpf
p
R2 R3
PA2
lateral
q
R2
R3
~5 dpf
dorsal
Hoxa2a & Krox20a
r
R5 R3
3 dpf
lateral
s
R2
R3
R4
R5
PA2
dorsal
4 dpf
t
R2 R3 R4 R5
PA2
lateral
u
R2
R3
R4
R5
PA2
~5 dpf
dorsal

DISCUSSION

Loss of the *hoxa2b* R4 enhancer is a synapomorphy of syngnathid fish.

Syngnathid fish all share a modified rhombomere 4 *hoxa2b* enhancer element. We find that the Pbx/Hox binding element sequence motifs and spacing between the binding elements have been modified for this enhancer. One Prep/Meis binding motif has been lost. One of the Pbx/Hox binding motifs is partially lost. Ghost pipefish, the immediate outgroup to the teleost family Syngnathidae, has all the expected binding sites for this enhancer element, which means that the total loss of the Prep/Meis binding site must have occurred after ghost pipefish split from Syngnathidae. Interestingly, the length of the spacing of the binding sites in the ghost pipefish falls between the typical vertebrate spacing lengths (with the exception of the space between PH2 and PH3) and the reduced spacing length found in the examined syngnathid fish (Figure 5.6).

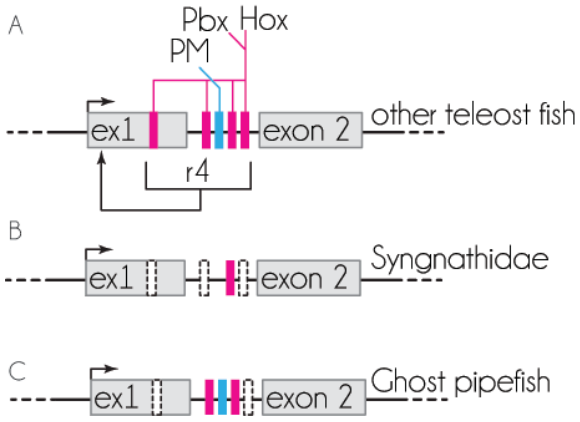


Figure 5.6: Schematic of rhombomeric regulatory modules in *hoxa2b* in Syngnathid. a) binding sites present in other teleost fish. b) binding sites in syngnathid fish. Dashed boxes indicate site with a high amount of sequence change. c) binding sites in the ghost pipefish

Bases on a previous study by Tumpel et al. (2007), which examined the variation of this enhancer in 12 vertebrates, it was found that the Prep/Meis sequence stayed conserved across the vertebrates examined. The Pbx/Hox binding site, PH2. stayed highly conserved. The PH1 and PH3 sites were very conserved across amniotes, but showed more various in the fish species examined, to the point where it deviated from the TGAT and NNAT motifs. This previous examination of the binding motifs of the *hoxa2b* enhancer in zebrafish, fugu, and medaka showed that there is more sequence variation in the teleost version of this enhancer, to the degree that they do not fit the 5'-TGATNNAT-3'. Curiously, even though teleost fish seem to defy the 5'-TGATNNAT-3' or the 5'-NNAT-3' rules established in (Ferretti et al. 2005), experiments using the zebrafish and fugu version of this enhancer still led to expression in the hindbrain (Tumpel et al. 2007, Tumpel et al. 2006). One can assume that either the teleost *Hox* and *Pbx* can bind to these sites without the 5'-TGATNNAT-3' or the 5'-NNAT-3' or this particular binding element in the teleost enhancer is not as critical.

The natural variants on this enhancer element has been previously reported in Tumpel et al. 2006, Tumpel et al. 2007 and Parker et al. 2014. Up until now, variation to this enhancer element was limited to slight modifications to the inter-elemental space between the critical Pbx/Hox and Prep/Meis bind sites and a small degree of base pair changes. Amniotes have very conserved motifs for PH1–3, with more various in these binding sites present in fish. The Prep/Meis site has stayed perfectly conserved in vertebrates examined, with no known variation (Tumpel et al. 2007). Complete loss of the Prep/Meis binding site, reduction in spacing between the binding sites, and the sequence changes to the Pbx/Hox sites have never been reported until now in syngnathid fish.

Loss of the *hoxa2b* R4 enhancer affects expression in a predictable fashion

Subsequently, we show expression of this gene in rhombomere 4 is lower relative to the surrounding rhombomeres and this change in expression is consistent with causing effects on the cranial neural crest. Other studies have reported changes to regulatory elements that have resulted in interesting phenotypic modifications to body plans (reviewed in (Rebeiz and Tsiantis 2017, Wray 2007, Gehrke and Shubin 2016, Carroll 2008)). Some examples include the *pitx1* regulatory mutations influencing the reduction of pelvic fin structure in stickleback fish (Chan et al. 2010), the inactivation of a *Tbx4* enhancer likely contributing to the evolution of limblessness in snakes (Infante et al. 2015), and regulatory mutations in *ovo/svb* affecting trichomes in *Drosophila* larvae (Stern and Frankel 2013). This study adds to the increasing evidence to that noncoding changes are linked to body plan changes.

Hoxa2 has been previously described as a "master regulator of craniofacial programs and jaw formations" (McEllin et al. 2016). Mouse, zebrafish and Nile tilapia *hoxa2* paralog mutants have homeotic mutation phenotypes that involve pharyngeal arch 2 cranial elements developing into pharyngeal arch 1 cranial elements (Le Pabic, Scemama, and Stellwag 2010, Hunter and Prince 2002, Gendron-Maguire et al. 1993, Rijli et al. 1993, Santagati et al. 2005). Although the requirement of *hoxa2* for proper pharyngeal arch 2 derivative development is well demonstrated, the mechanism is less understood.

Multiple perturbation studies have demonstrated that *Hox* genes and hindbrain segmentation play important roles in neural crest cell specification, migration and differentiation. This is possibly due to the fact that signals from rhombomeres influence neural crest migratory routes. Specific rhombomeres have different contributions to

streams of cranial neural crest cells.

Rhombomere 4 contributes to the stream of cranial neural crest cells that populate pharyngeal arch 2 and these neural crest cells continue to express *hoxa2* as they migrate to pharyngeal arch 2. *Hoxa2* can repress components of the ossification pathway like *sox9*, *phx1*, *runx2* in pharyngeal arch 2 in neural crest cells. Intriguingly, syngnathids have numerous modifications to their skulls, which include pharyngeal arch 1 derived Meckel's cartilage, quadrate and metapterygoid, and pharyngeal arch 2 derived preopercular, opercular, and symplectic bones (Leysen et al. 2010, Brown 2010, Kimmel, Small, and Knope 2017). Early in development, Gulf pipefish have a relatively expanded pharyngeal arch 1 derived palatoquadrate and Meckel's cartilage, and a relatively reduced pharyngeal arch 2 ceratohyal (Brown 2010). Potentially, the loss of the *hoxa2b* enhancer element is tied to the highly modified skull in syngnathid fish.

In addition to bones, rhombomere 4 is important for nerves and Mauthner cells development. Intriguingly, syngnathids have reportedly lost their Mauthner cells (Benedetti, Sassi, and Stefanelli 1991). *Hoxa2* -/- mouse mutants have been described to have an altered rhombomere 2 and 3 motor axons, which suggests that changes in expression in *hoxa2b* in rhombomere 4 could affect the alar plate of rhombomere 4 (Gavalas et al. 1997). Although, Mauthner cells are derivatives of the basal plate, not the alar plate which would against this connection.

CONCLUSION

Making use of the increasingly available *de novo* genome assemblies of highly derived animals like syngnathid fish allows us to take advantage of natural evolutionary

developmental models. Creatures like syngnathid fish can provide insight into how biodiversity evolved.

In this study, we asked how a *hoxa2b* enhancer is modified in syngnathid fish and infer possible downstream morphological consequences to the loss of this enhancer element. We described how this element has been modified in syngnathid fish and the expression of the *hoxa2b* that it regulates during syngnathid development. We find that the binding element sequence motifs and spacing between the binding elements have been modified for this enhancer. One binding motif has been lost and a second binding site has been partially lost. Subsequently, we show expression of this gene in rhombomere 4 is lower relative to the surrounding rhombomeres, reflecting previously published functional tests for this enhancer, and this change in expression is consistent with causing effects on the cranial neural crest.

Studying the genetic basis of morphological divergence in organisms with greatly derived morphologies provides an opportunity to explore the ways that conserved genetic pathways can be altered and how genetic changes can lead to the evolution of derived traits. Our data support the hypothesis that natural mutations can occur in these deeply conserved pathways in ways potentially related to phenotypic diversity.

CHAPTER VI

CONCLUSION

A central question in evolutionary biology concerns how organisms evolve highly derived and novel morphologies (Darwin 1859, Raff 2012, Carroll, Grenier, and Weatherbee 2013). An amazing diversity of phenotypes has evolved across multicellular organisms, but biologists are still largely unclear as to how highly novel phenotypes arise at the genetic level. Evolutionary origins of such things like the turtle's shell or the elongation of snake have been the subject of numerous studies over the years.

As a particular example, teleost fish have evolved numerous diverse characteristics including highly derived body plans. For instance, the dwarf cyprinids from the genus *Paedocypris* has one the smallest vertebrate skeletons with a large reduction in skeletal elements (Britz and Conway 2009). The *Mola mola* sunfish also has a reduced skeleton, but it is actually one of the largest species of teleost (Pan et al. 2016). Syngnathid fish also have evolved numerous modifications to their morphologies as well. This includes expansion of vertebral elements, leafy appendages, prehensile tails, male somatic brooding and loss of ribs, caudal, and pelvic fins (Neutens et al. 2014, Herald 1959, Wilson and Rouse 2010, Hoffman, Mobley, and Jones 2006).

Despite this diversity in body size and shape, research dating to just the 1980s has now clearly shown that all vertebrates share a common core of genes and pathways important for developmental processes that occur throughout ontogeny. A fundamental gap in our knowledge is how diverse phenotypes—and particularly highly derived novelties—evolve using this conserved genetic toolkit.

King and Wilson (1975) was one of several papers to first propose that evolutionary changes can be more often attributed to the change in gene expression rather

than the changes of the protein sequences (King and Wilson 1975, Zuckerkandl and Pauling 1965, Britten and Davidson 1971). More recent studies have shown the connection between changes in developmental gene expression and the evolution of derived morphological features (reviewed by (Carroll 2008, Hoekstra and Coyne 2007)).

For this dissertation, I identified the genetic changes that are responsible for the evolution of some of the unique vertebrate morphological characters present in syngnathid fish. I used comparative genomics, gene editing, and gene expression approaches to investigate the genetic and genomic changes to the developmentally important *Hox* genes in a group of fish that exhibit a striking departure from the typical fish body plan: the pipefish and seahorse family, Syngnathidae.

Looking back

Syngnathid fish provide an exceptional opportunity to study the evolution of novelties because they provide both a breadth of characters absent in all other teleost lineages and now, with several genomes available for this family—including the one present in Chapter I—we have the genomic tools limited to only a handful of fish species. *Hox* genes code for homeodomain transcription factors that are responsible for determining the body plan of an embryo along the anterior-posterior axis, and changes to these genes have paralleled the rise of morphological diversity in the vertebrate animals. The evolution of syngnathid fish involved major modifications to their vertebrate body plan, but the developmental genetic basis of those changes is unknown.

In Chapter II, I included the Gulf pipefish genome publication for which I am a co-author (Small et al. 2016). Production of a reference genome from this family of Syngnathidae was necessary for my proposed dissertation research. Therefore, I

significantly contributed to the production of the Gulf pipefish genome and its publication. I described the genomic organization of *Hox* clusters in a species of syngnathid pipefish—the Gulf pipefish (*Syngnathus scovelli*). I assess the phylogenetic placement of syngnathid fish relative to other representative fish taxa using ultraconserved elements and I compared the *Hox* cluster gene content of the Gulf pipefish against other teleost fish species. It was the first time that the *Hox* clusters were described from a member of the Syngnathidae family. Overall, I found that the *Hox* gene content has remained largely conserved relative to other teleost fish with annotated *Hox* clusters with a few key losses. The key losses included the convergent loss of *hox7* genes and the unique loss of *eve1*.

In Chapter III, I presented a preliminary investigation on phenotypic consequences to the loss of *hox7* genes in teleost fish—a group of *Hox* genes that are missing in syngnathids. I describe the successful use of the CRISPR/Cas9 system to induce indels in all *hox7* genes (*hoxa7a*, *hoxb7a)* in the threespine stickleback (*Gasterosteus aculeatus*) and established transgenic lines for the *hox7* gene knockouts. In addition, I described some preliminary results that indicate the possible role for *hox7* genes in rib and vertebrae development. This provided insight into the morphological consequences to the evolutionary loss of these genes in syngnathid fish.

Both Chapters II and III were focused on exploring the *Hox* gene content and the possible phenotypic impact of the evolutionary loss of some of these *Hox* genes. I found some key losses to the *Hox* genes that could have contributed some of the divergent skeletal features in these fish. Not finding large degrees of change to the *Hox* genes in pipefish and seahorses is maybe not that surprising. As discussed in the first chapter of this dissertation, *Hox* genes tend to maintain a high level of conservation throughout

animals (reviewed in (Gehring, Affolter, and Bürglin 1994, Burglin and Affolter 2016, Holland 2013)). This level of conservation in *Hox* genes and in other core developmental gene families has been hypothesized to occur because major changes will be detrimental to the development of the organism. Alternatively, slight shifts in expression of conserved developmental genes that may create traits that evolution can act upon while still working within the boundaries of developmental constraint (Wilkins 2002, Raff 2012). Perhaps modifications to these regulatory elements that have contributed to modified body plans of the pipefish and seahorse family.

Therefore, for Chapter IV, I wanted to explore the conserved noncoding elements within the boundaries of the syngnathid *Hox* clusters. These conserved noncoding elements are putative *cis*-regulatory element for the surrounding *Hox* genes. I used *Hippocampus erectus, H. comes* and *Syngnathus scovelli* as the syngnathid representatives and compared their CNE content to percomorph teleost fish (*Gasterosteus aculeatus, Takifugu rubripes*, *Oryzias latipes*, *Thunnus orientalis*), non-percomorph teleost fish (*Boleophthalmus pectinirostris*, *Gadus morhua*, *Danio rerio*), non-teleost fish (*Lepisosteus oculatus)*, and two non-fish vertebrates (*Mus musculus* and *Homo sapiens*).

I cataloged many noncoding elements that were found the Gulf pipefish, tiger tail and lined seahorse genomes. I found three unique CNE losses only found among the syngnathid species. Two of the three CNEs are undescribed in the literature and it is unknown whether or not these are regulatory elements. The third element is a known enhancer element for *hoxa2b* and its loss was further examined in the final experiment chapter of this thesis.

In Chapter V, I was able to further expand my syngnathid sampling to include two species of the Nerophinae subfamily—*Doryrhamphus excisus* and *Choeroichthys sculptus* and five species from the Syngnathinae subfamily—*Corythoichthys haematopterus*, *Syngnathus scovelli*, *Hippocampus erectus*, *H. comes*, and *H. zostrae*. I also incorporated sequence data from a species from the *Solenostomus* genus, which is the immediate outgroup to Syngnathidae. I found that the Pbx/Hox binding element sequence motifs and spacing between the binding elements have been modified for this enhancer. One Prep/Meis binding motif has been lost in Syngnathidae.

Subsequently, I showed expression of this gene in rhombomere 4 is lower relative to the surrounding rhombomeres in the Gulf pipefish and this change in expression is consistent with it causing effects on the cranial neural crest. Ghost pipefish, the immediate outgroup to the teleost family Syngnathidae, has all the expected binding sites for this enhancer element, which means that the total loss of the Prep/Meis binding site must have occurred after ghost pipefish split from Syngnathidae. Like the *Hox* gene content, I found no great shifts in the putative *cis*-regulatory elements of the *Hox* clusters in Syngnathidae. Singular changes, such as the loss of the *hoxa2b* enhancer element possibly contributed to the evolution to the divergent development of cranial facial elements.

Looking forward

The findings of this research have revealed intriguing examples of development gene and noncoding loss. A deeper survey of the individual *hox7* genotypes will provide much more insight into whether disrupting these genes are the causative factor for these recorded axial deformities. It would be interesting to also follow up some of the other

notable gene losses of the syngnathid family discovered by my dissertation work such as the tooth development gene *eve1* or the axial regulator *mir196b* with knockout experiments in stickleback.

Expanding the taxonomic sampling in the research involving the *hoxa2b* enhancer was very insightful. Future incorporation of more syngnathid species will be very helpful in timing when certain developmental genetic changes occurred and generating further genomes from this family will greatly aid in this. Producing several genomes from more closely related outgroups that exhibit more standard teleost morphologies such as members from the goatfish (Mullidae), flying gurnards (Dactylopteridae), and dragonets (Callionymidae) will also prove to be useful in adding evolutionary significance to certain changes described in Syngnathidae.

I found that the *Hox* clusters in syngnathids has remained relatively preserved. Given the large degree of change in body morphology of these fish, it is interesting to see that these amazing morphologies evolved from seemingly subtle changes to their *Hox* developmental toolkit. Results from this research indicate the divergent syngnathid body plan is not due to rampant change in throughout *Hox* clusters and support that certain key changes to the *Hox* genes, microRNAs, and regulatory elements have led these fish evolving a unique *Hox* cluster that probably had a significant impact on their body plan developmental evolution. This research in syngnathid fish show how gene and regulatory element loss can work as an important source of genetic variation that, in term, can contribute to adaptive phenotypic diversity.

APPENDIX A

SUPPORTING INFORMATION FOR CHAPTER III

Table S3. 1: Primers used for CRISPR indel screening.

Gene target	Primer name	Primer sequence
hoxa7a	Forward 1	GGTGTATTGCTGTCATATATCAC
hoxa7a	Forward 2	GAGTTCTTATTATGTGGATGGTC
hoxa7a	Reverse 1	CGAAATTAATTGAACCACTAACG
hoxa7a	Reverse 2	GGCTTTAAAATAGAACGTACGAG
hoxb7a	Forward 1	CTGTTTTCCAAATACCAGCTAG
hoxb7a	Forward 2	GATCCTTCAACTTCTCCTTCC
hoxb7a	Reverse 1	TCTTTCTATTCATATCCCTTCCC
hoxb7a	Reverse 2	GGCTCTTCCTCGTTTAACTG

Table S3.2: List of G1 crosses generated. G1 generation was made using parents that screened positive for CRISPR indels.

G1 type	Family	♀	♂	CRISPR alleles detected?
hoxa7a	3187	3131.0002	3131.0001	yes
hoxa7a	3189	3131.0002	3131.0003	yes
hoxa7a	3202	3127.0001	3135.0001	tba
hoxa7a	3221	3131.0004	3135.0002	tba
hoxa7a	3210	3127.0002	3135.0002	tba
hoxa7a	3228	3127.0004	3135.0003	tba
hoxb7a	3190	3133.0001	3129.0002	yes
hoxb7a	3194	3133.0002	3133.0003	yes
hoxb7a	3218	3129.0005	3133.0008	tba
hoxb7a	3216	3133.0007	3133.0008	yes
hoxb7a	3195	3129.0003	3133.0003	tba
hoxb7a	3197	3133.0004	3133.0005	tba
hoxb7a	3199	3129.0004	3133.0005	no
hoxb7a	3201	3129.0005	3133.0005	tba
hoxb7a	3203	3133.0006	3135.0001	tba
hoxb7a	3204	3133.0002	3129.0006	tba
hoxb7a	3206	3129.0003	3133.0005	tba
hoxb7a	3219	3133.0010	3133.0008	tba
hoxb7a	3217	3133.0009	3133.0008	tba
hoxa7a;hoxb7a	3229	3143.0001	3143.0002	tba
hoxa7a;hoxb7a	3223	3141.0002	3127.0003	tba
hoxa7a;hoxb7a	3222	3141.0001	3127.0003	yes
hoxa7a;hoxb7a	3240	3143.0003	3127.0005	yes
hoxa7a;hoxb7a	3241	3143.0004	3127.0006	yes
hoxa7a;hoxb7a	3247	3143.0004	3127.0007	tba

Table S3.3: Early stop codon status of CRISPR lesion. The columns from left to right lists the CRISPR target gene, the stickleback stock number, the type of indel lesion detected via Sanger sequencing of TOPO clones, and whether or not the CRISPR lesion led to a frameshift mutation that would cause an early stop codon to the gene peptide sequence.

CRISPR target	Family	Lesion type	Early stop codon?
hoxa7a	3187	3 bp deletion	no
hoxa7a	3187	14 bp insertion	yes, 33 nucleotides from lesion
hoxa7a	3187	18 bp insertion, 2 bp deletion	late stop codon with extra 11 amino acids
hoxa7a	3189	13 bp deletion	yes, 31 nucleotides from lesion
hoxa7a	3189	14 bp deletion	yes, 48 nucleotides from lesion
hoxb7a	3190	5 & 18 bp insertions	yes, 61 nucleotides from first lesion
hoxb7a	3194	5 bp deletion	yes, 184 nucleotides from lesion
hoxb7a	3216	4 bp deletion	yes, 35 nucleotides from lesion
hoxb7a	3216	9 bp insertion	no
hoxb7a	3216	2 bp deletion, 1 bp insertion	yes, 45 nucleotides from first lesion
hoxa7a	3222	3 & 3 bp insertions, 1 & 14 bp deletions	no
hoxa7a	3222	55 bp deletion	yes, 32 nucleotides from lesion
hoxa7a	3222	21 bp deletion	no
hoxb7a	3222	2 bp deletion	yes, 188 nucleotides from lesion
hoxb7a	3222	18 bp deletion	no
hoxa7a	3240	3 bp deletion	no
hoxa7a	3240	2 & 1 bp insertions, 2 bp deletion	late stop codon with extra 11 amino acids
hoxb7a	3240	1 bp insertion	yes, 190 nucleotides from lesion
hoxb7a	3240	15 bp deletion	no
hoxa7a	3241	6 bp deletion	no
hoxb7a	3241	6 bp deletion	no

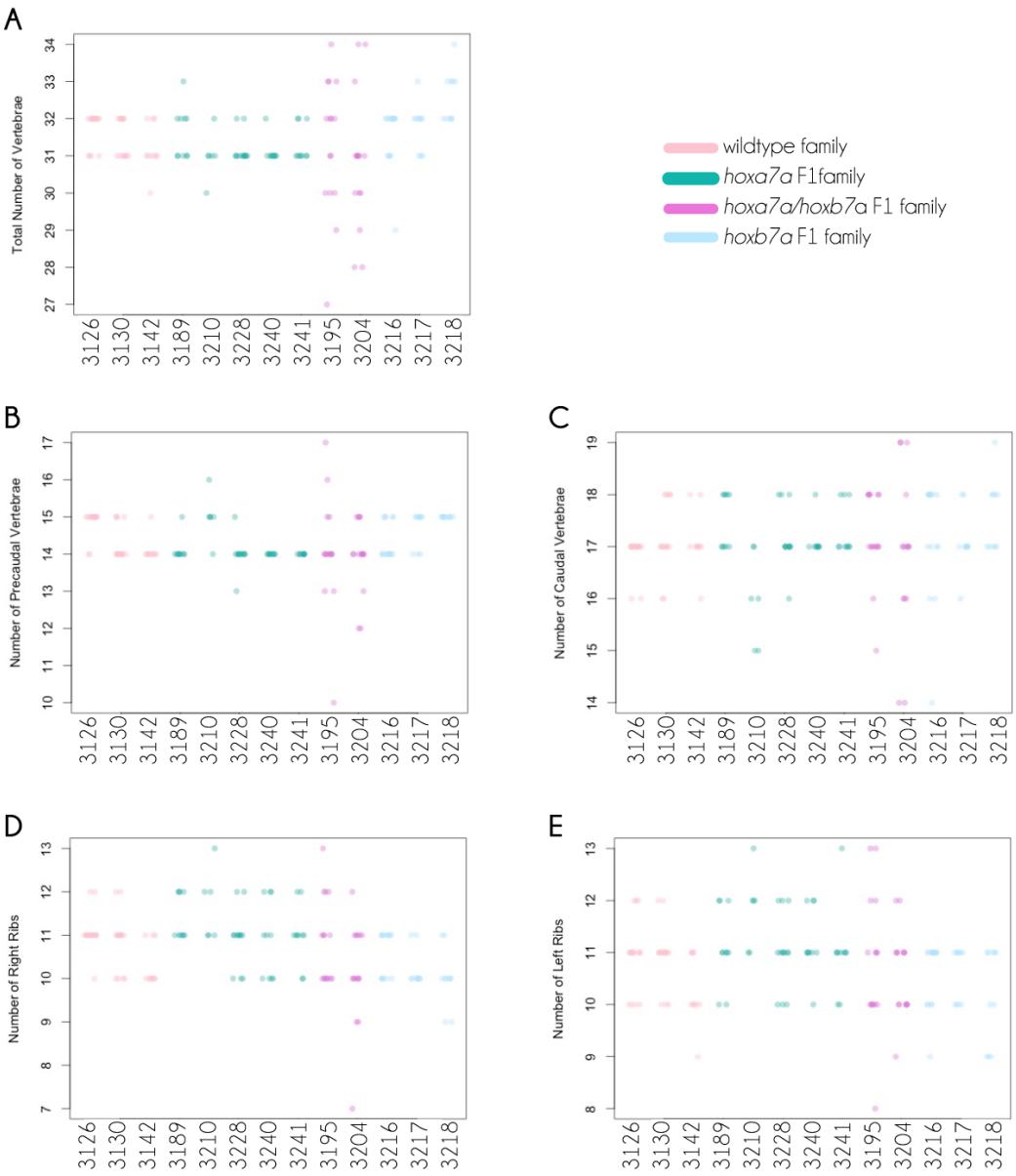


Figure S3.1: Distribution of axial character counts from G1 phenotypic screen. The G1 fish from the *hoxa7a* G1 families versus *hoxb7a* G1 families versus *hoxa7a* and *hoxb7a* families versus control fish have no significant difference in total vertebrae number total vertebrae number (χ^2=0.4729; d.f. =3, p=0.9248) (a), total precaudal vertebrae number (χ^2=0.6126; d.f.=3, p-value=0.8935)(b), total caudal vertebrae number (χ^2=0.0821; d.f.=3, p-value=0.9939)(c), total left pleural rib number (χ^2=0.9393; d.f.=3, p-value=0.8159)(d), total right pleural rib number (χ^2=1.436; d.f.=3, p-value=0.6971)(e).

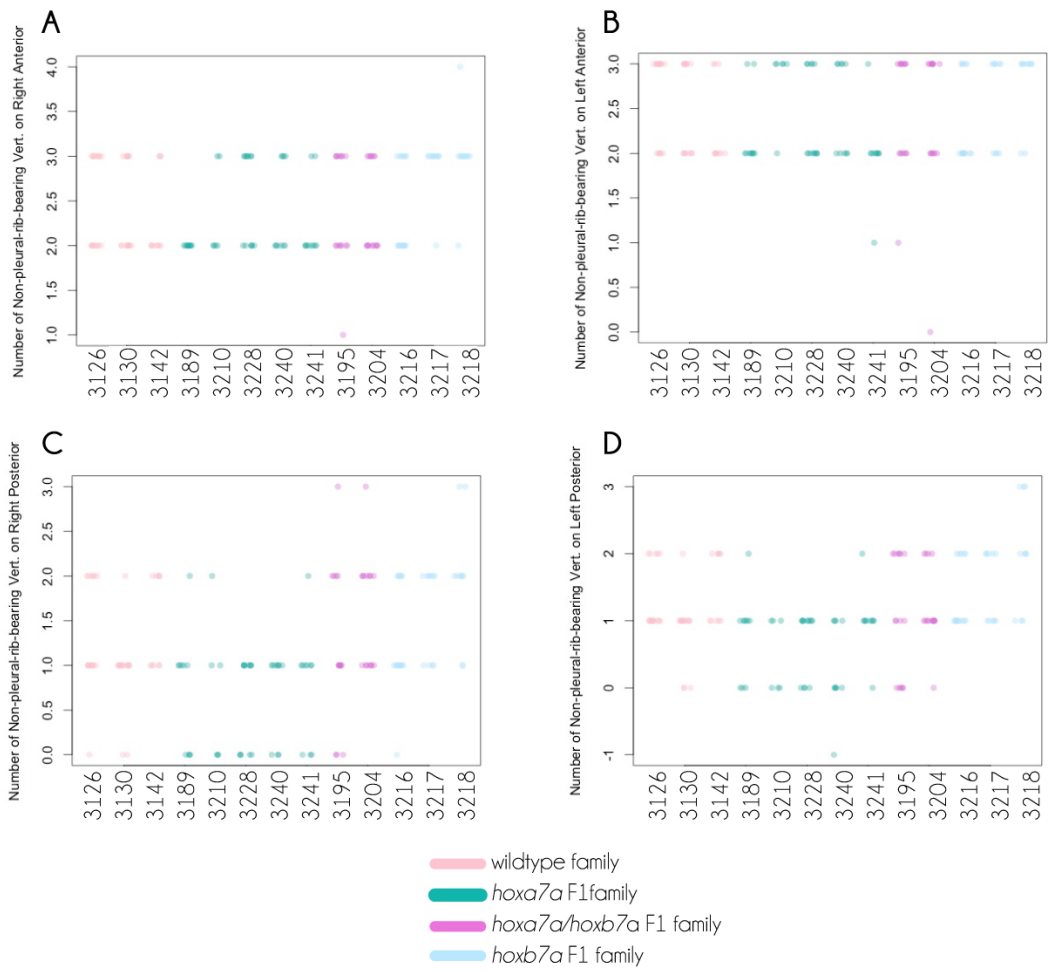


Figure S3.2: Distribution of non-pleural-rib-bearing precaudal vertebrate counts from G1 phenotypic screen. The G1 fish from the *hoxa7a* G1 families versus *hoxb7a* G1 families versus *hoxa7a* and *hoxb7a* families versus control fish have no significant difference total anterior precaudal vertebrate that do not bear pleural ribs on right (χ^2=0.4604; d.f.=3, p-value=0.9275) (a), and total anterior precaudal vertebrate that do not bear pleural ribs on left (χ^2=1.4365; d.f.=3, p-value=0.697)(b). Total posterior precaudal vertebrate that do not bear pleural ribs on right (c) and left (d) also included.

APPENDIX B

SUPPORTING INFORMATION FOR CHAPTER IV

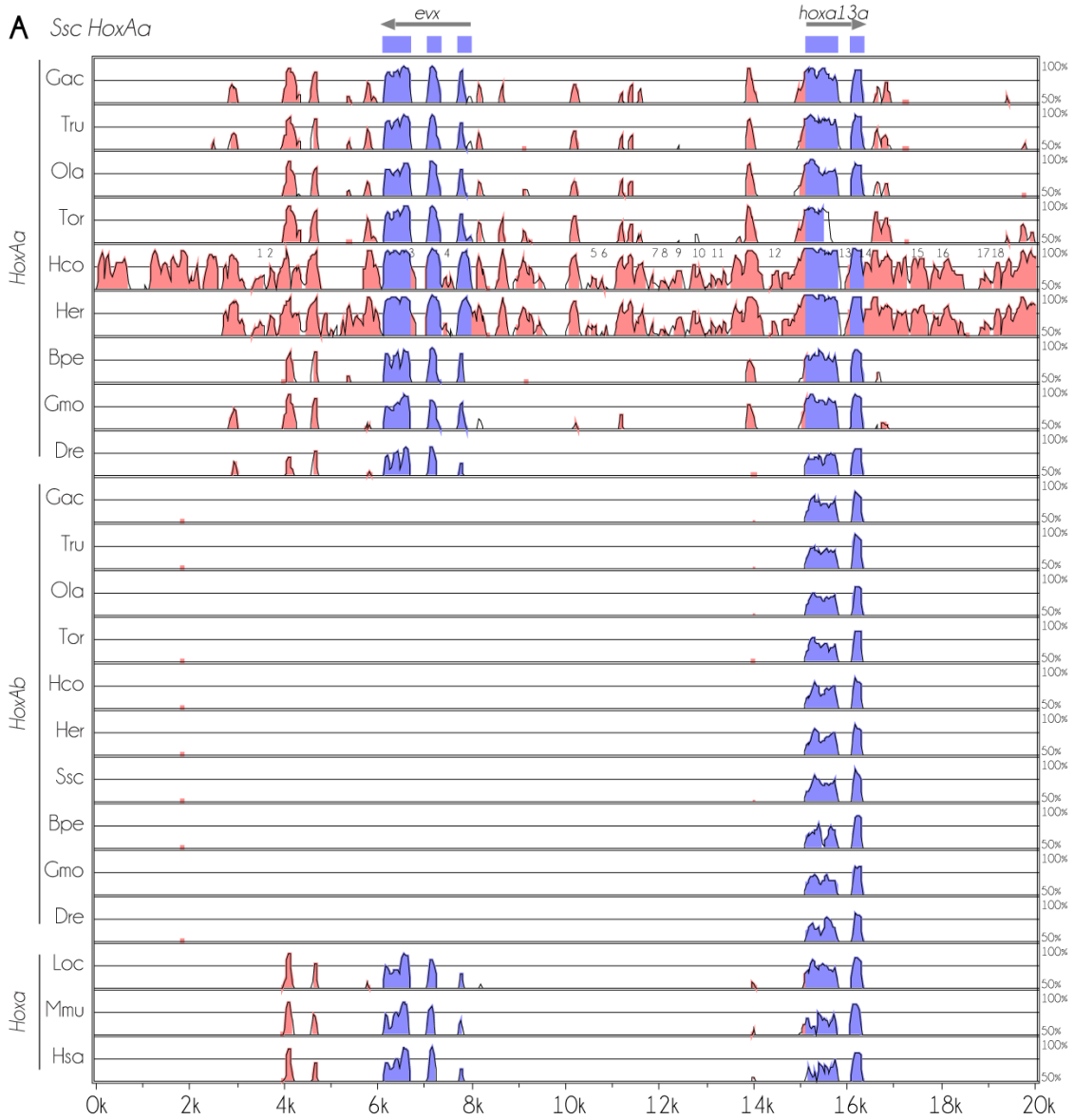


Figure S4. 1: VISTA plots for the *HoxA* clusters with Gulf pipefish *HoxAa* set as reference sequences. Exons are highlighted in blue, CNEs in pink, UTRs in teal, microRNAs are in the blue boxes. LAGAN alignment was used. The reference, Ssc, is Gulf pipefish; Tru, pufferfish; Ola, medaka; Tor, tuna; Gac, stickleback; Hco, *tiger tail* seahorse; Her, lined seahorse; Bpe, mudskipper; Gmo, cod; Dre, zebrafish; Loc, spotted gar; Mmu, mouse; Hsa, human. Syngnathid specific CNEs are numbered on the *tiger tail* seahorse.

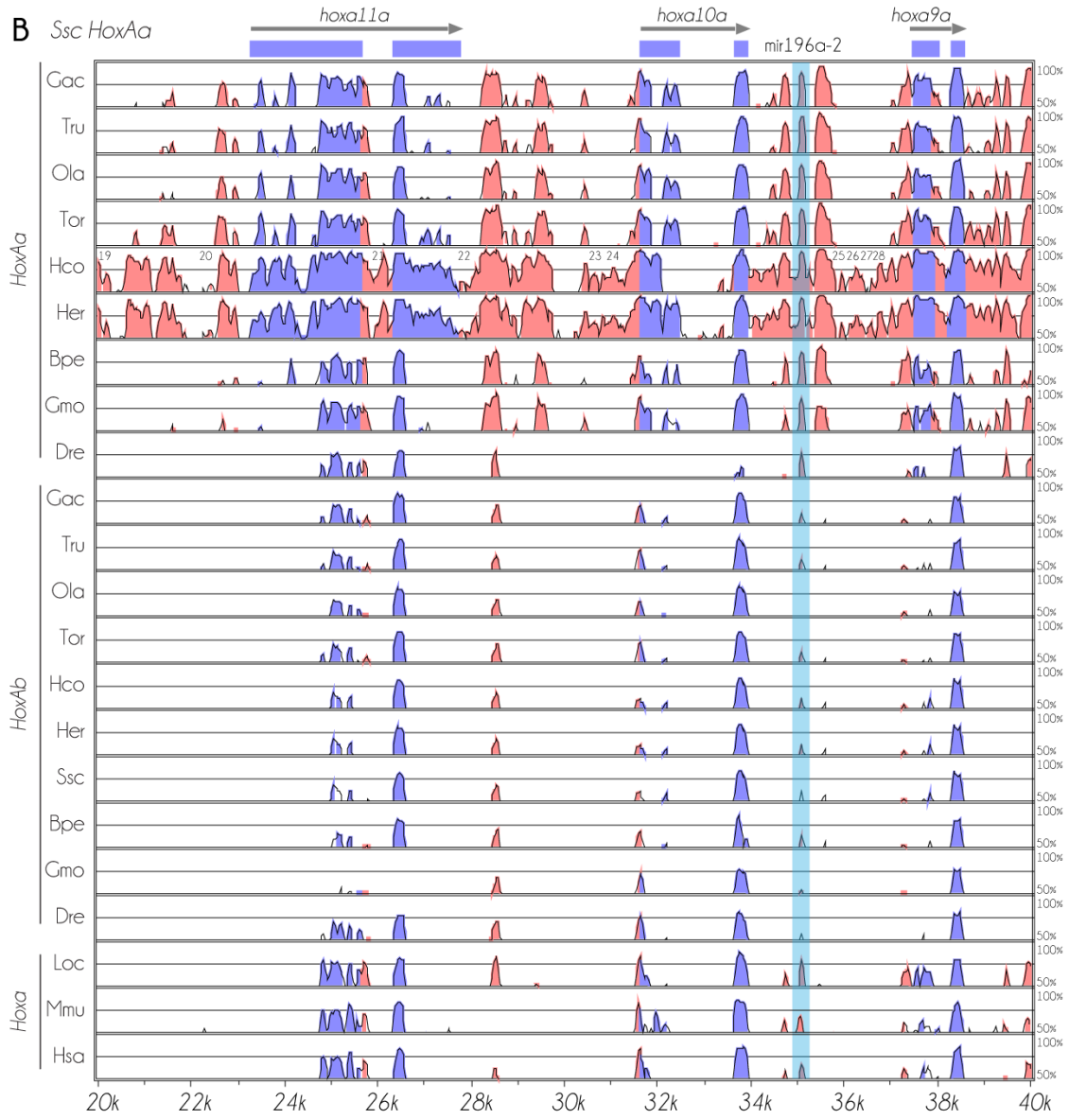


Figure S4.1 continued.

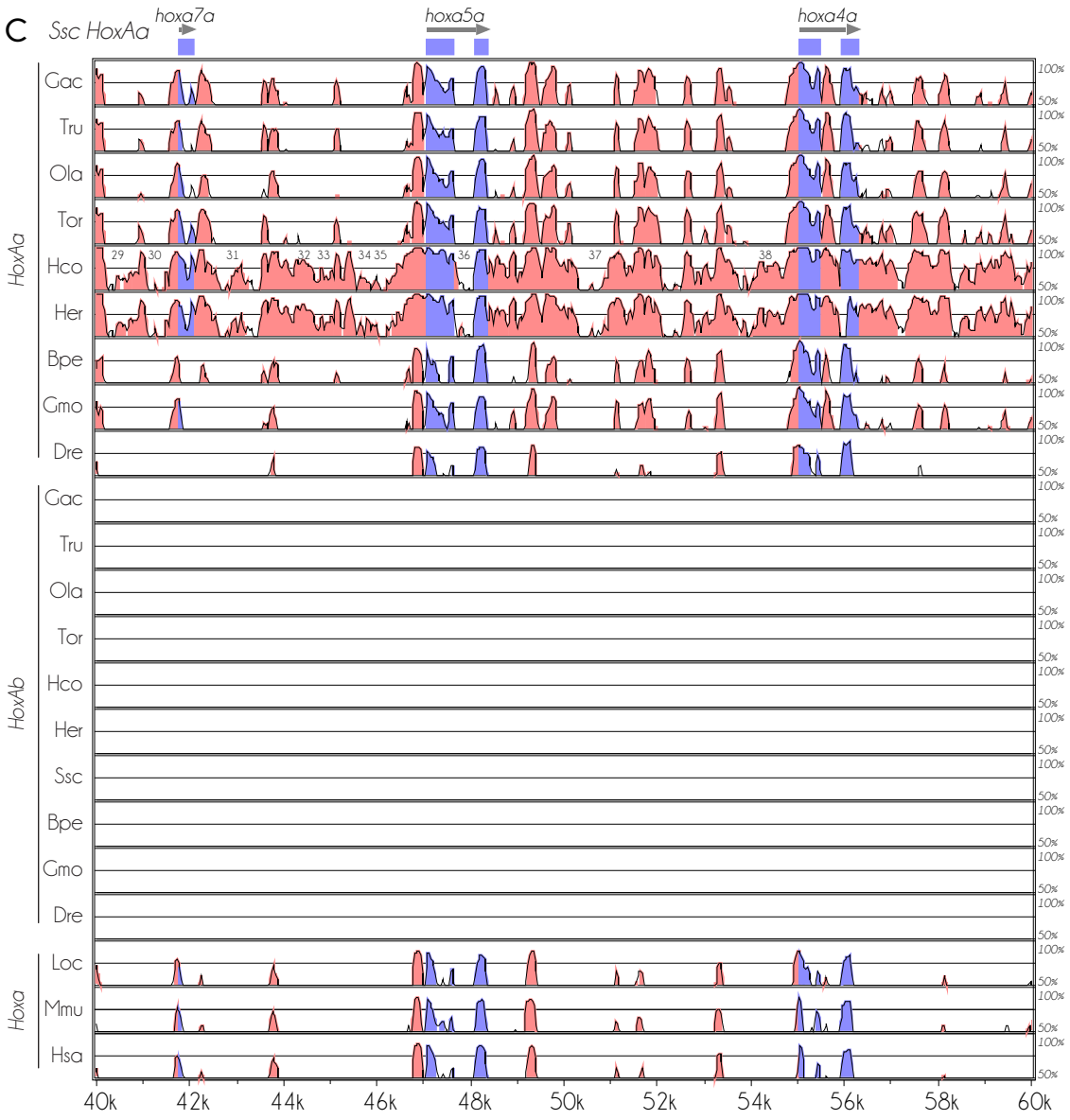


Figure S4.1 continued.

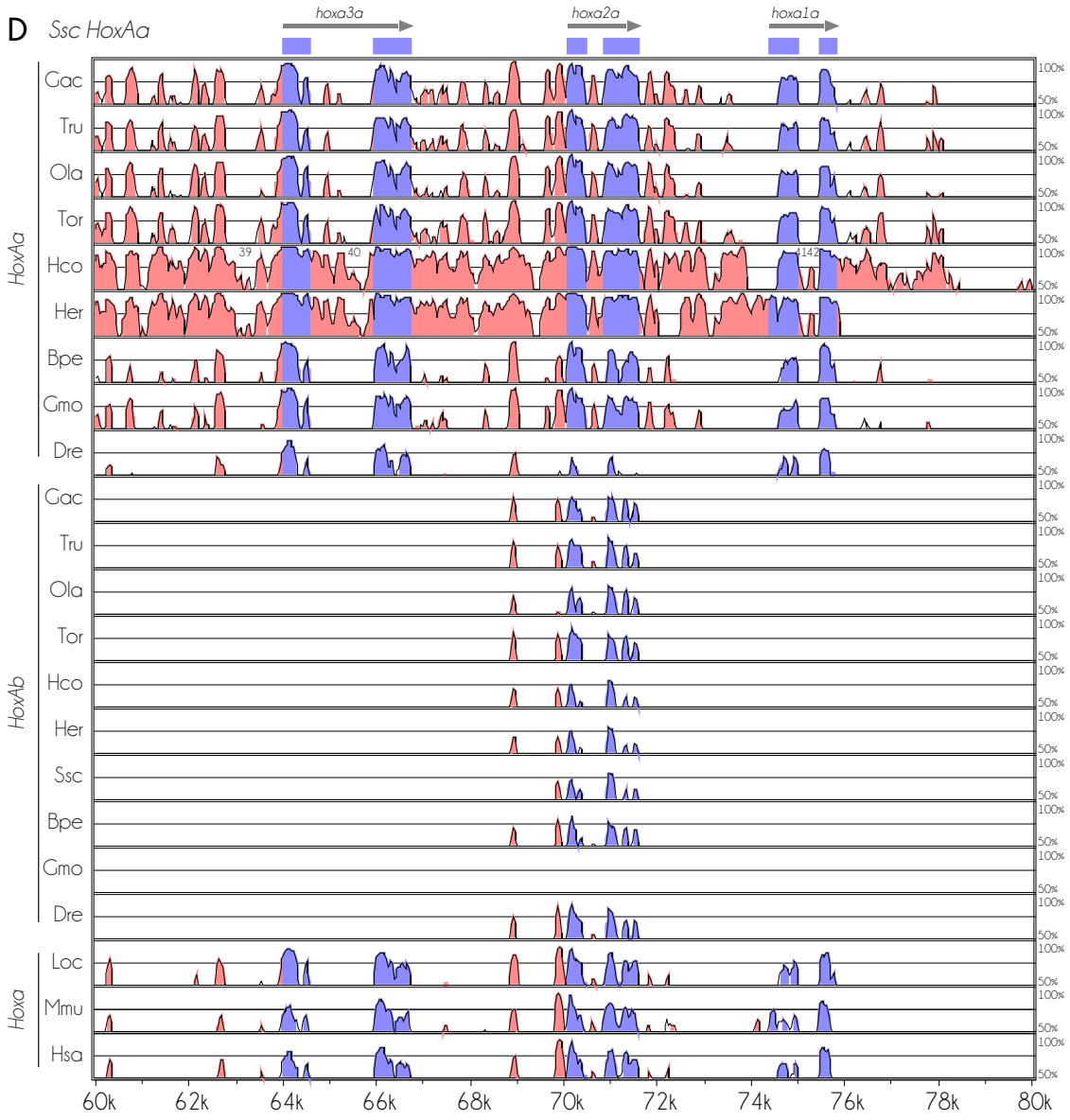


Figure S4.1 continued.

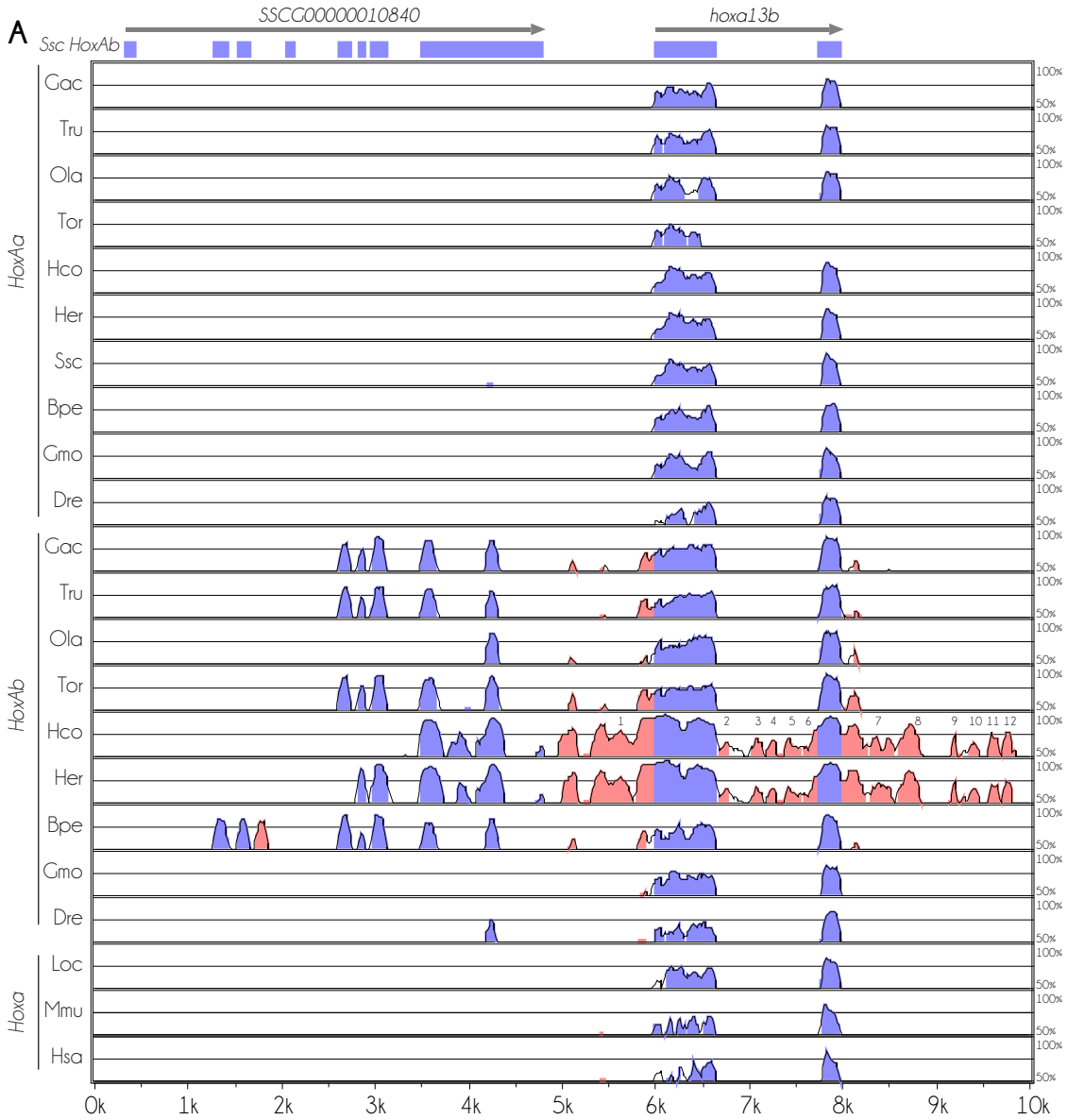


Figure S4.2: VISTA plots for the *HoxA* clusters with Gulf pipefish *HoxAb* set as reference sequence. Exons are highlighted in blue, CNEs in pink, UTRs in teal, microRNAs are in the blue boxes. LAGAN alignment was used. The reference, Ssc, is Gulf pipefish; Tru, pufferfish; Ola, medaka; Tor, tuna; Gac, stickleback; Hco, *tiger tail* seahorse; Her, lined seahorse; Bpe, mudskipper; Gmo, cod; Dre, zebrafish; Loc, spotted gar; Mmu, mouse; Hsa, human. Syngnathid specific CNEs are numbered on the *tiger tail* seahorse.

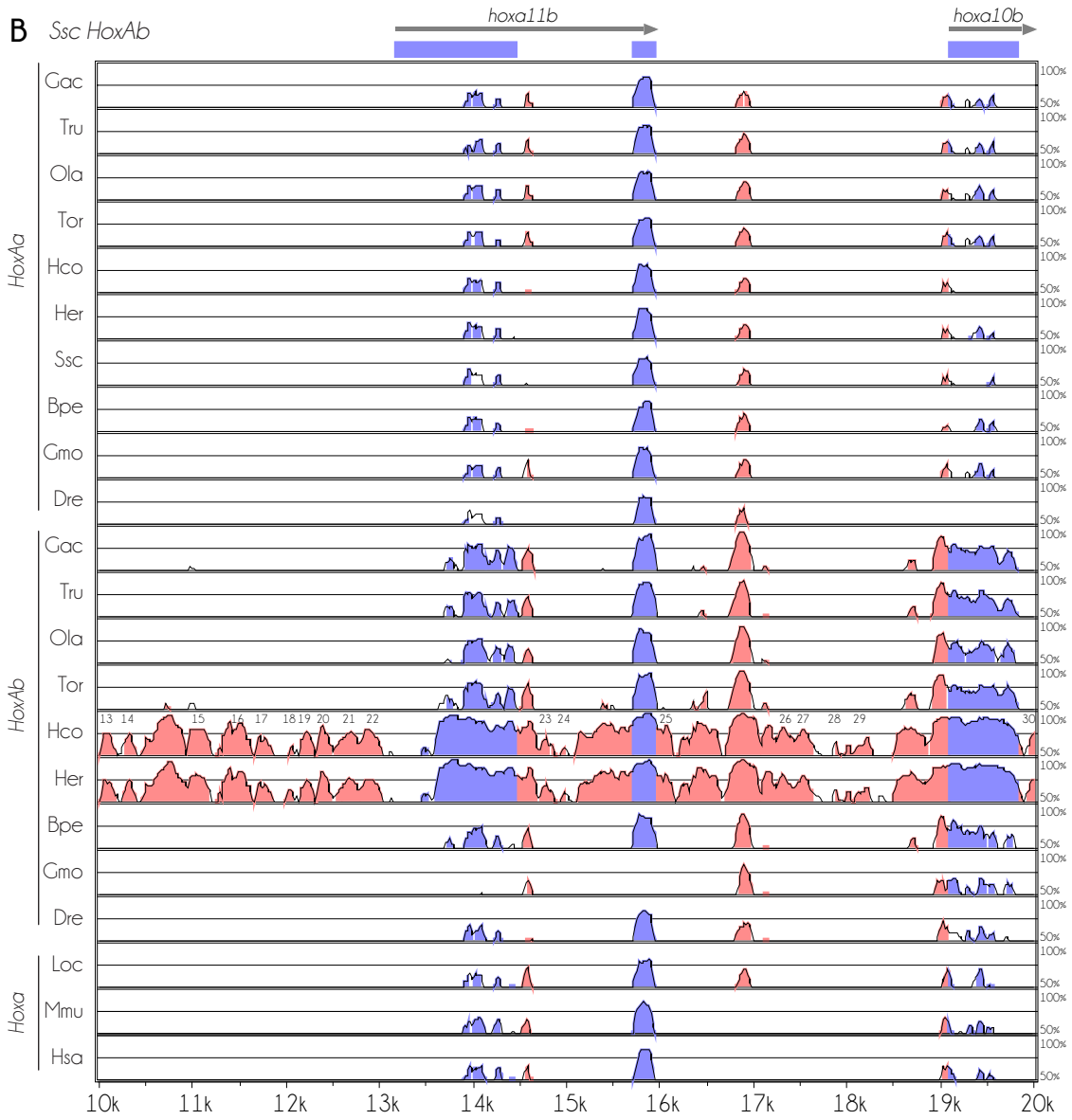


Figure S4.2 continued.

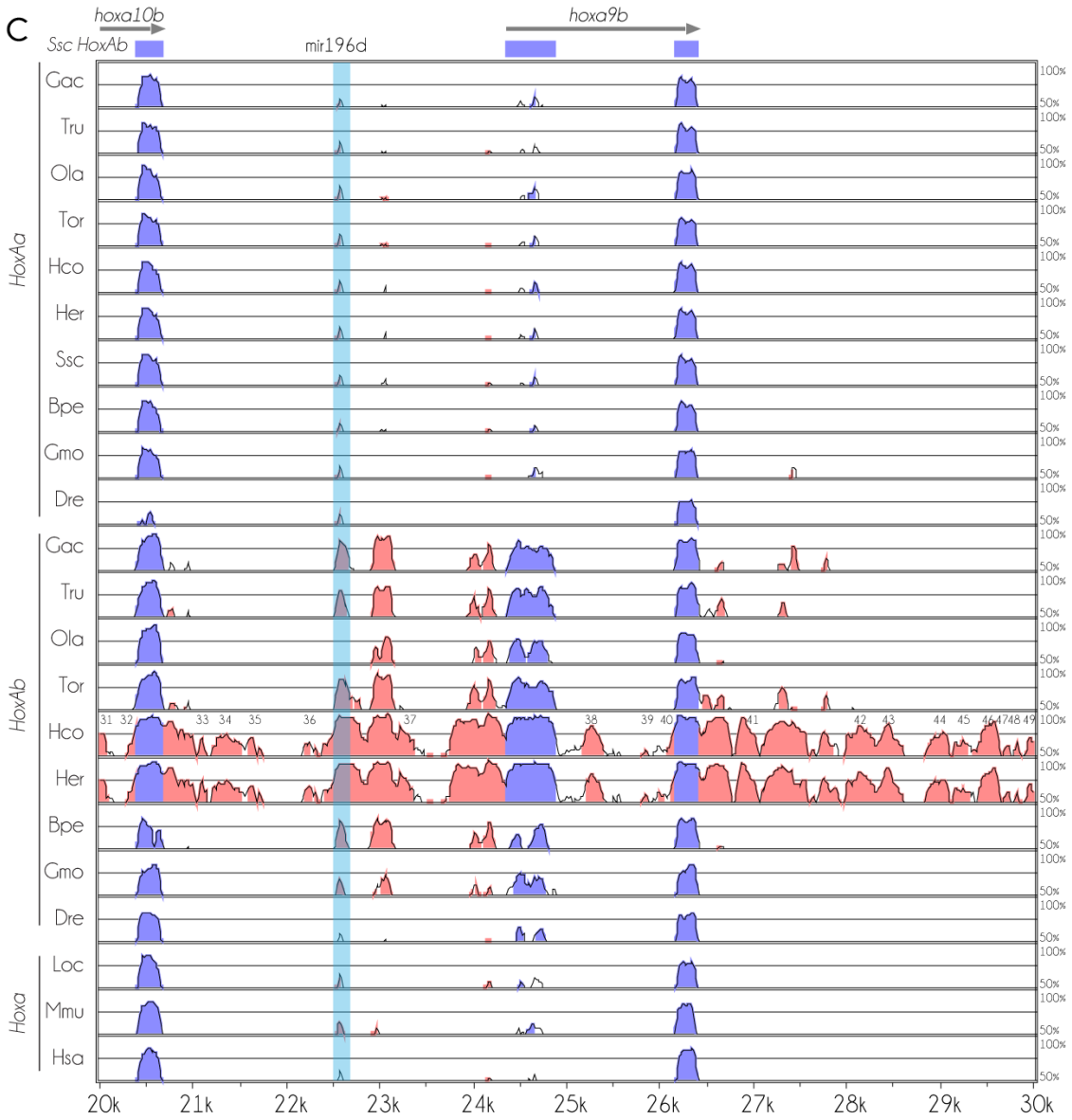


Figure S4.2 continued.

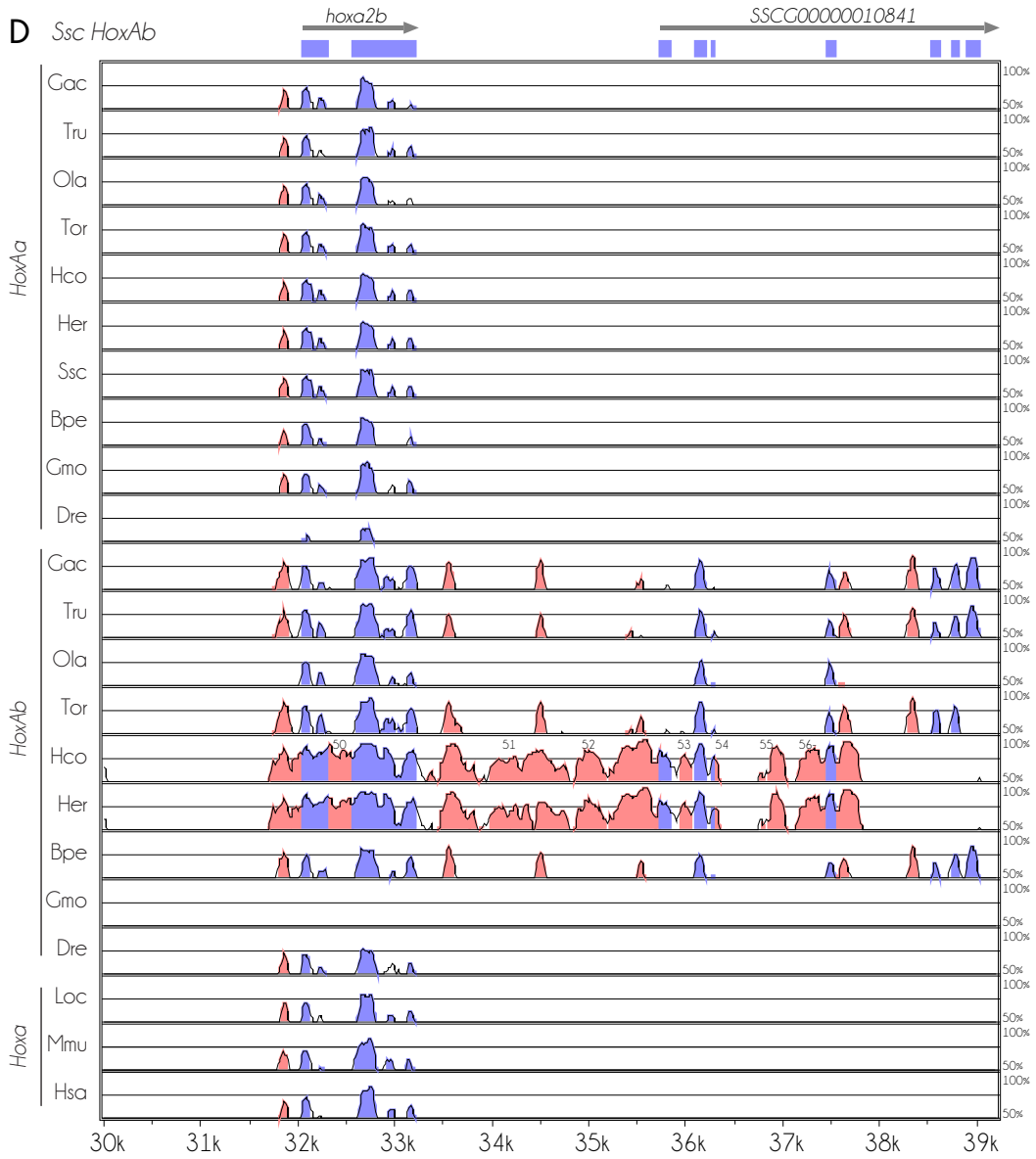


Figure S4.2 continued.

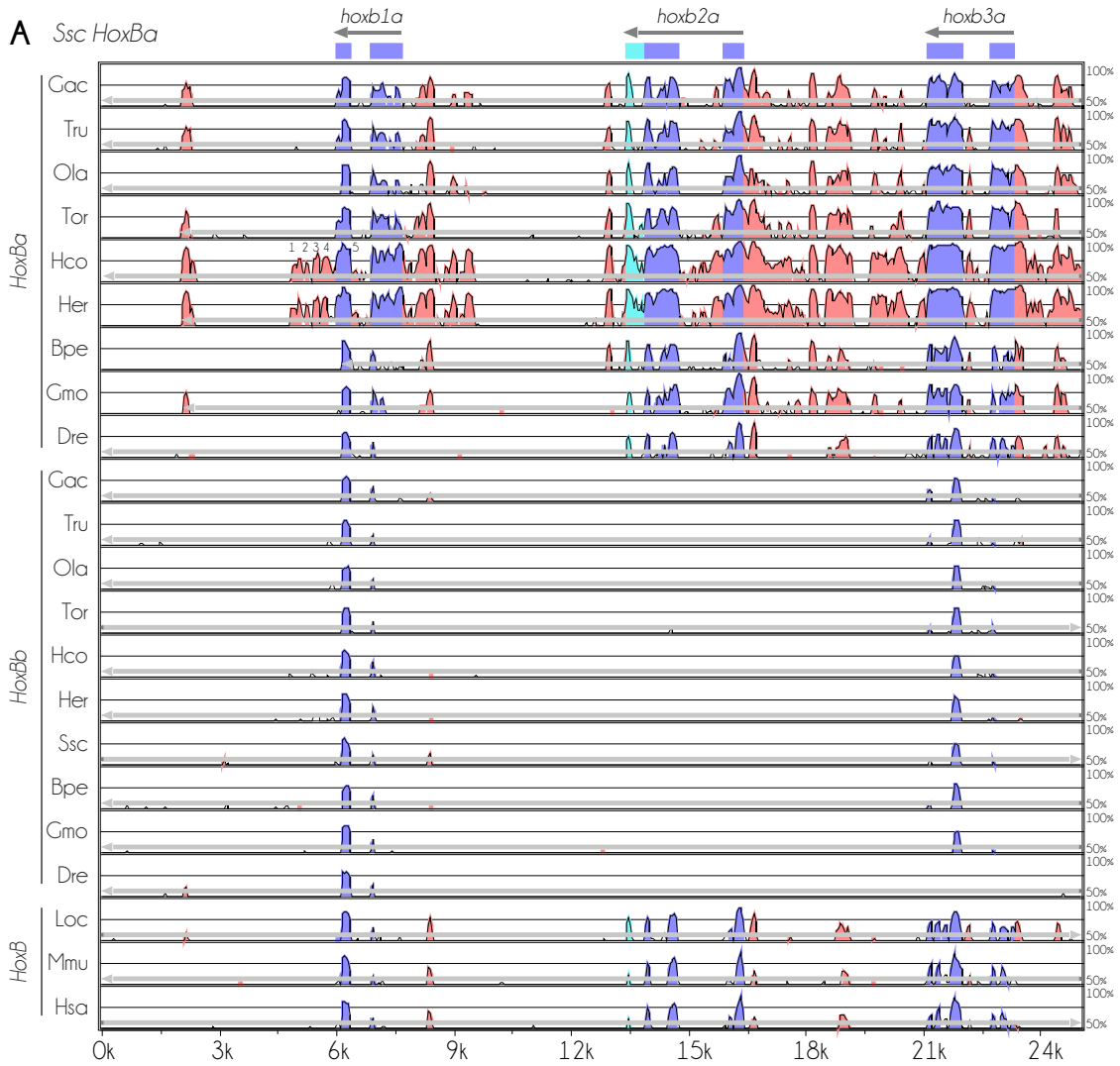


Figure S4.3: VISTA plots for the *HoxB* clusters with Gulf pipefish *HoxBa* set as reference sequences. Exons are highlighted in blue, CNEs in pink, UTRs in teal, microRNAs are in the blue boxes. The gray lines indicate stretches of continuous sequence. The reference, Ssc, is Gulf pipefish; Tru, pufferfish; Ola, medaka; Tor, tuna; Gac, stickleback; Hco, *tiger tail* seahorse; Her, lined seahorse; Bpe, mudskipper; Gmo, cod; Dre, zebrafish; Loc, spotted gar; Mmu, mouse; Hsa, human.

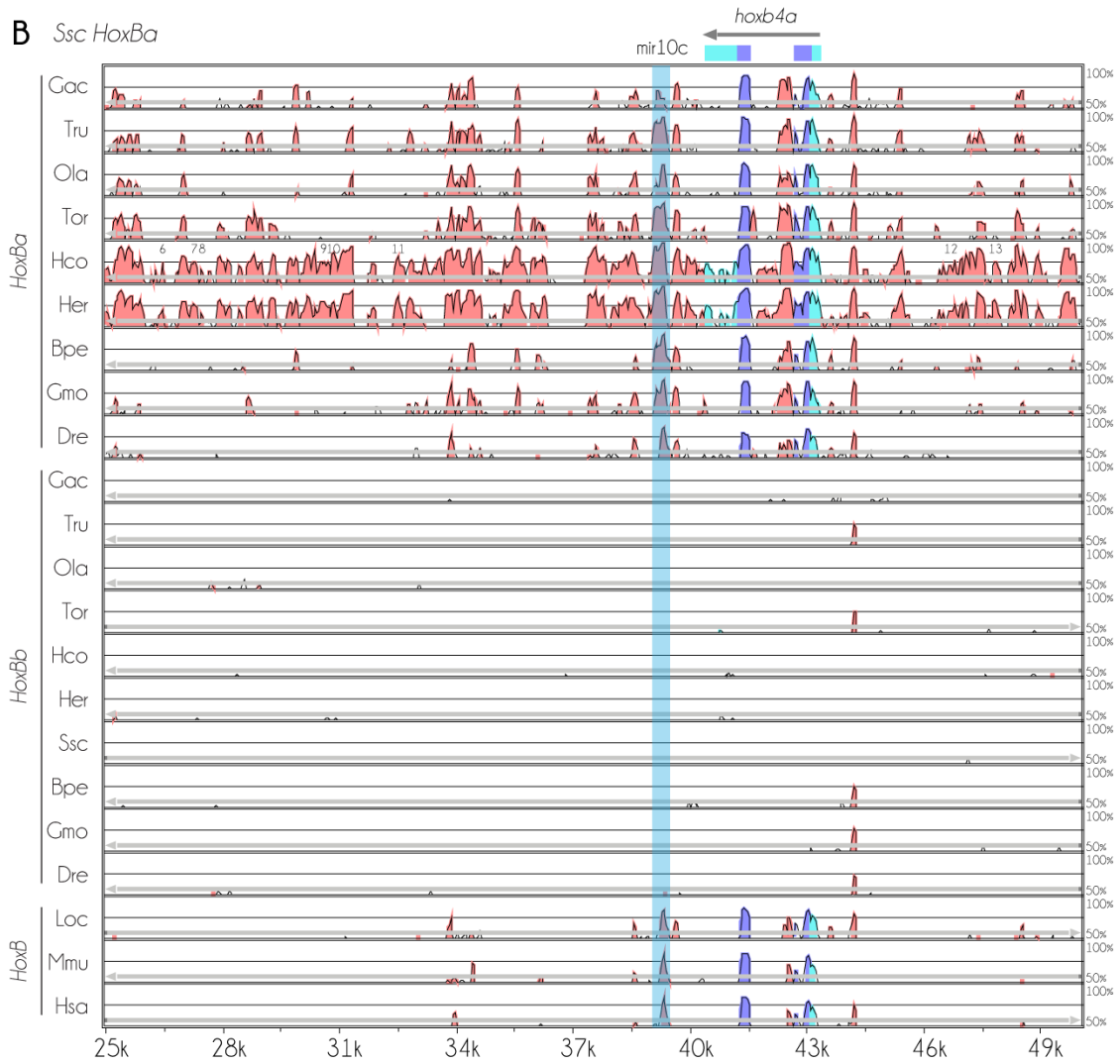


Figure S4.3 continued.

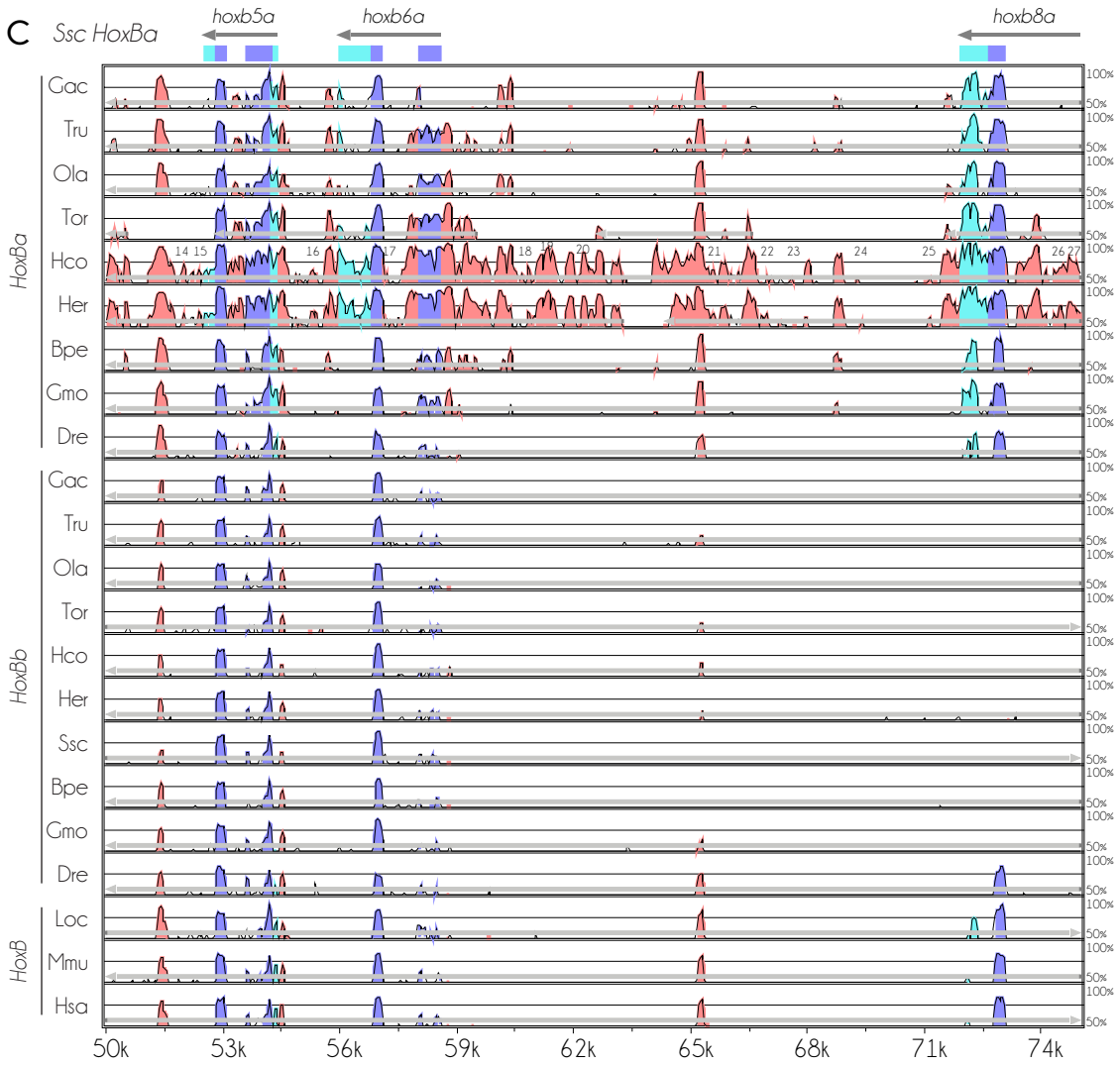


Figure S4.3 continued.

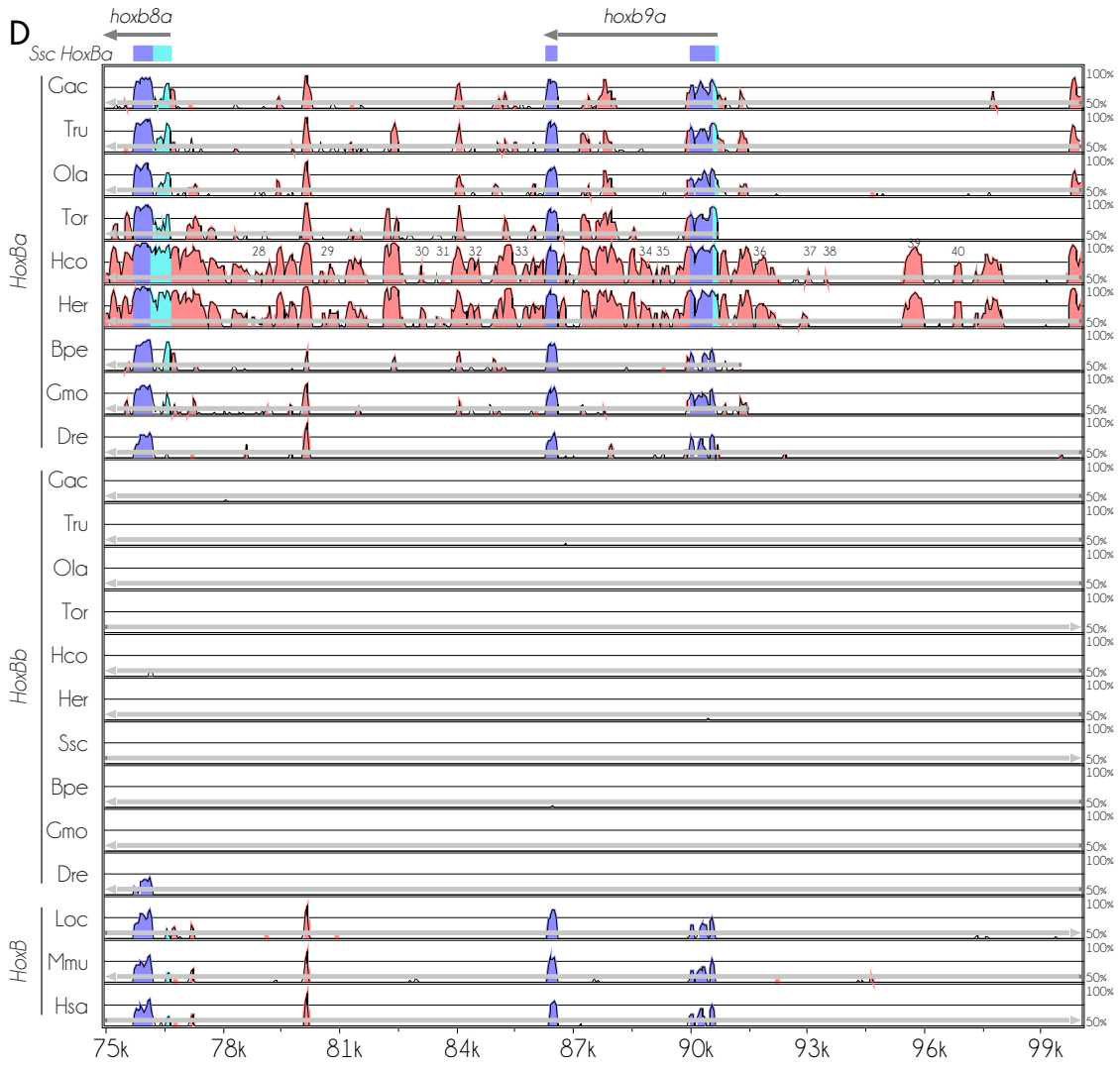


Figure S4.3 continued.

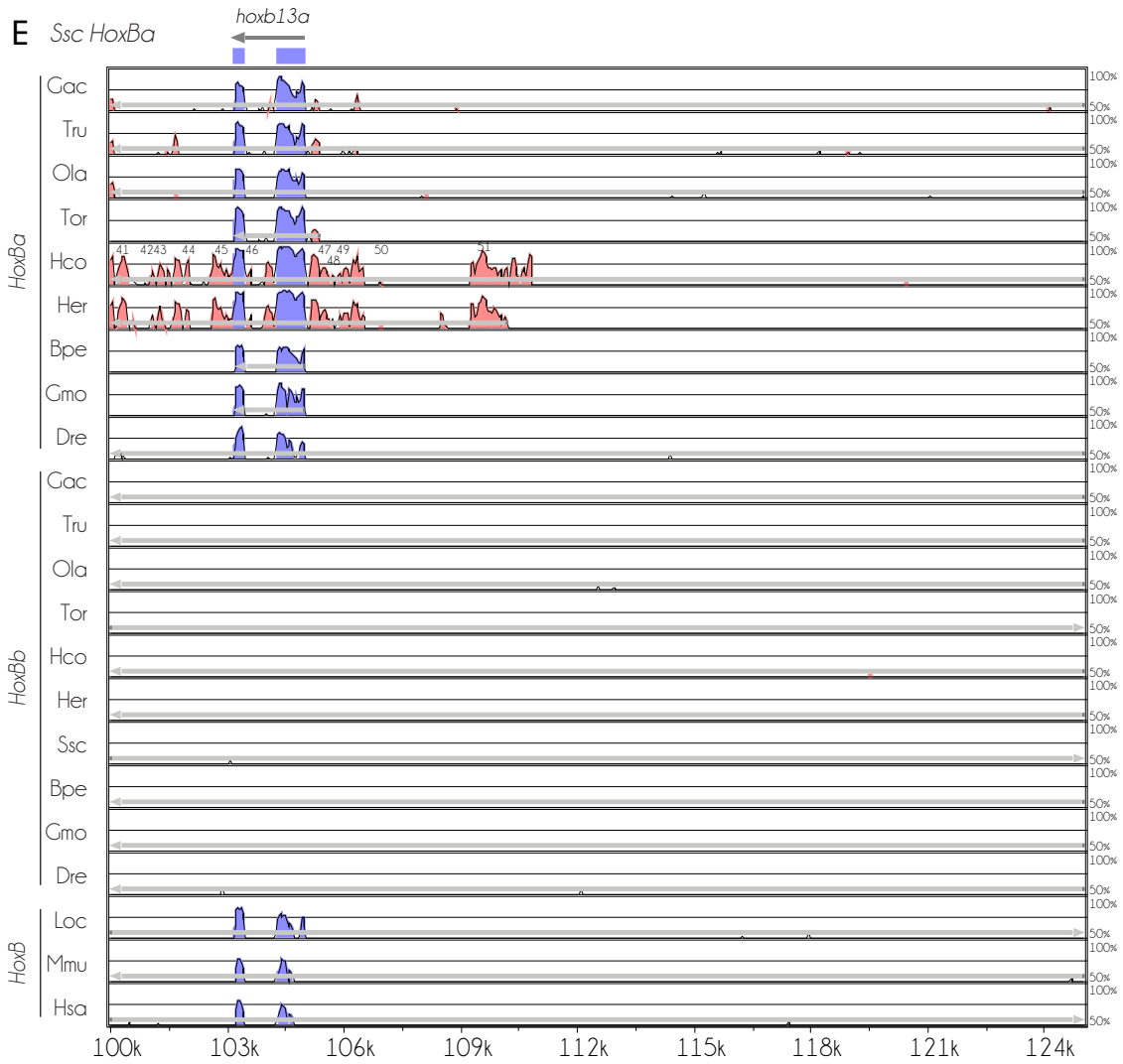


Figure S4.3 continued.

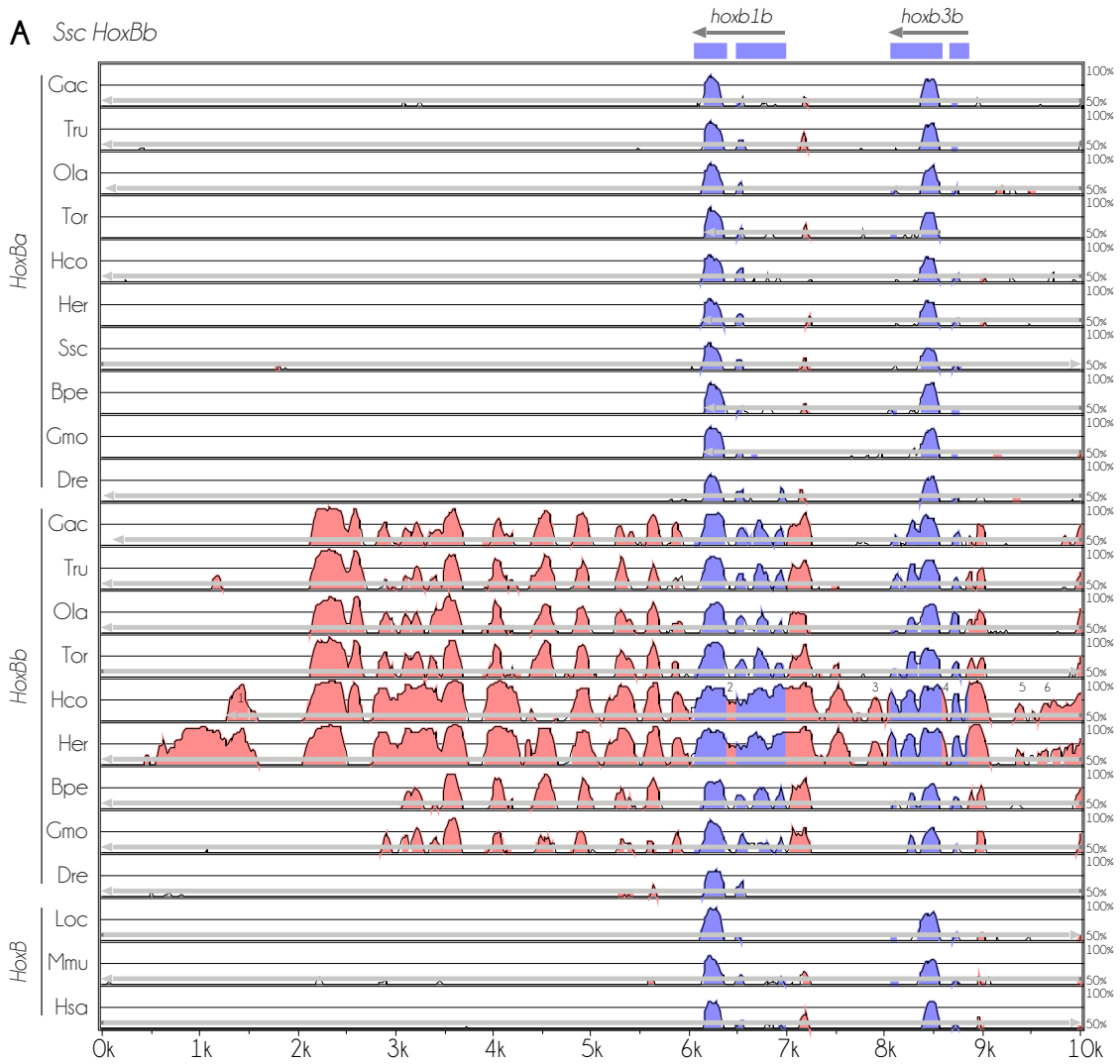


Figure S4.4: VISTA plots for the *HoxB* clusters with Gulf pipefish *HoxBb* set as reference sequence. Exons are highlighted in blue, CNEs in pink, UTRs in teal, microRNAs are in the blue boxes. Shuffle LAGAN alignment was used with gray lines indicate stretches of continuous sequence. The reference, Ssc, is Gulf pipefish; Tru, pufferfish; Ola, medaka; Tor, tuna; Gac, stickleback; Hco, *tiger tail* seahorse; Her, lined seahorse; Bpe, mudskipper; Gmo, cod; Dre, zebrafish; Loc, spotted gar; Mmu, mouse; Hsa, human. Syngnathid specific CNEs are numbered on the *tiger tail* seahorse.

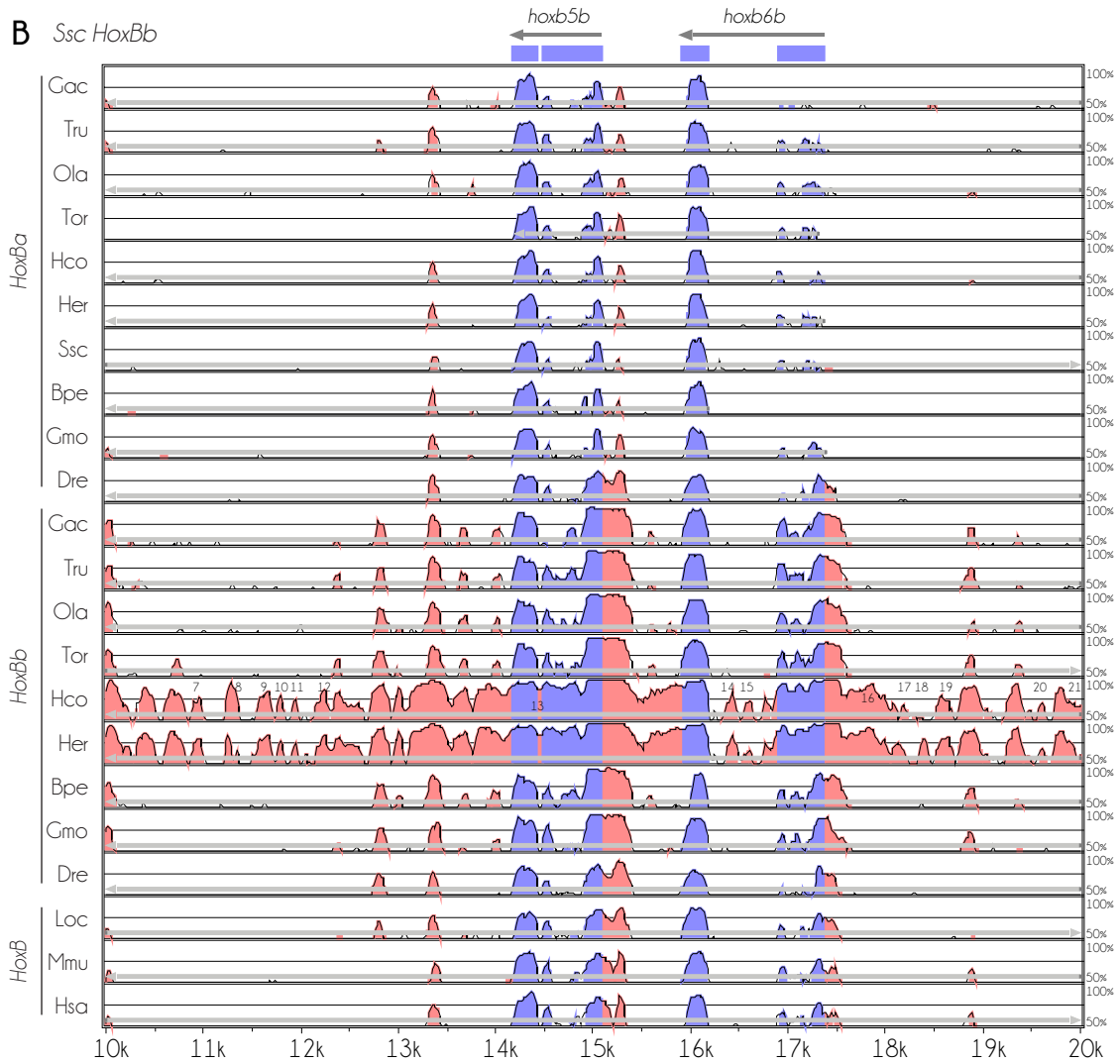


Figure S4.4 continued.

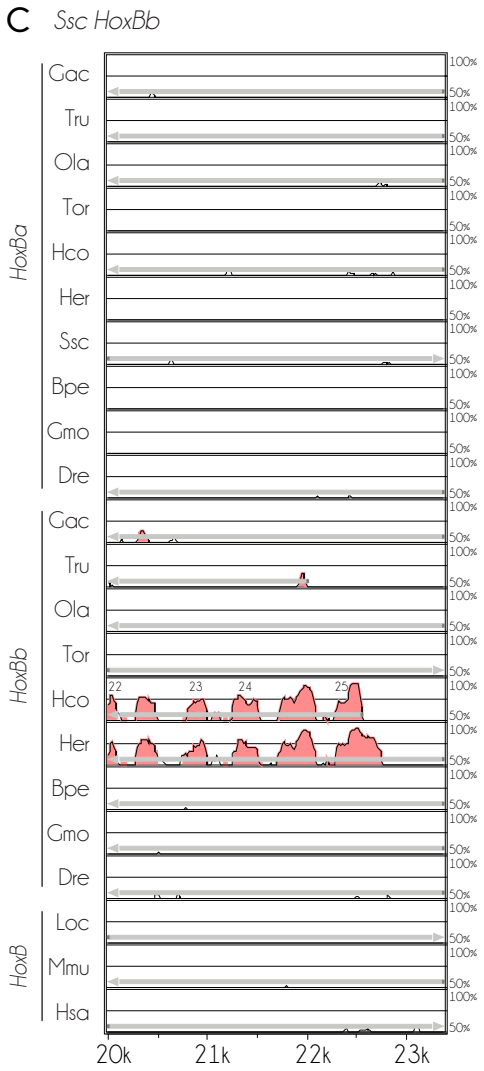


Figure S4.4 continued.

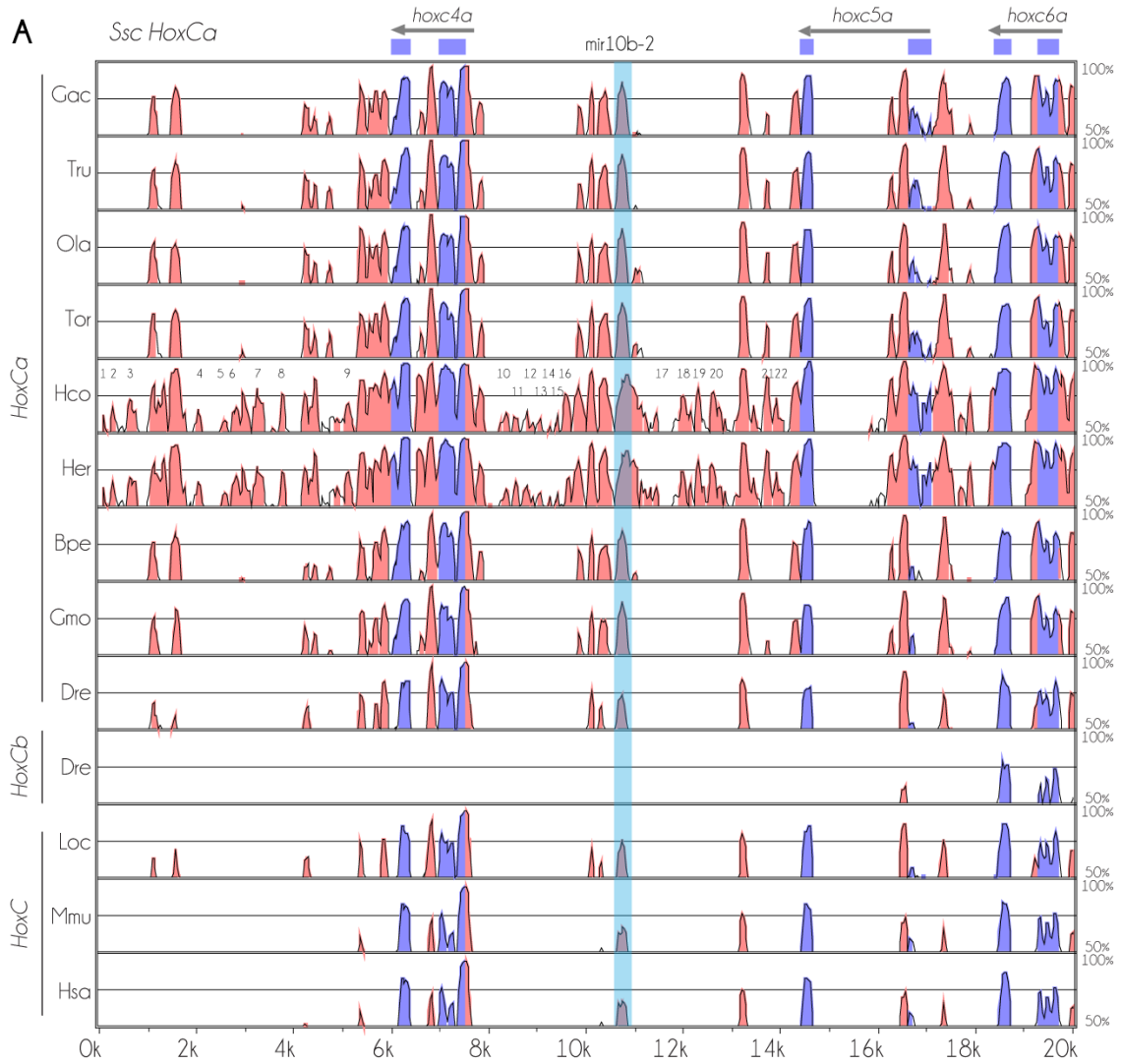


Figure S4.5: VISTA plots for the *HoxC* clusters with Gulf pipefish *HoxCa* set as reference sequence. Exons are highlighted in blue, CNEs in pink, UTRs in teal, microRNAs are in the blue boxes. LAGAN alignment was used. The reference, Ssc, is Gulf pipefish; Tru, pufferfish; Ola, medaka; Tor, tuna; Gac, stickleback; Hco, *tiger tail* seahorse; Her, lined seahorse; Bpe, mudskipper; Gmo, cod; Dre, zebrafish; Loc, spotted gar; Mmu, mouse; Hsa, human. Syngnathid specific CNEs are numbered on the *tiger tail* seahorse.

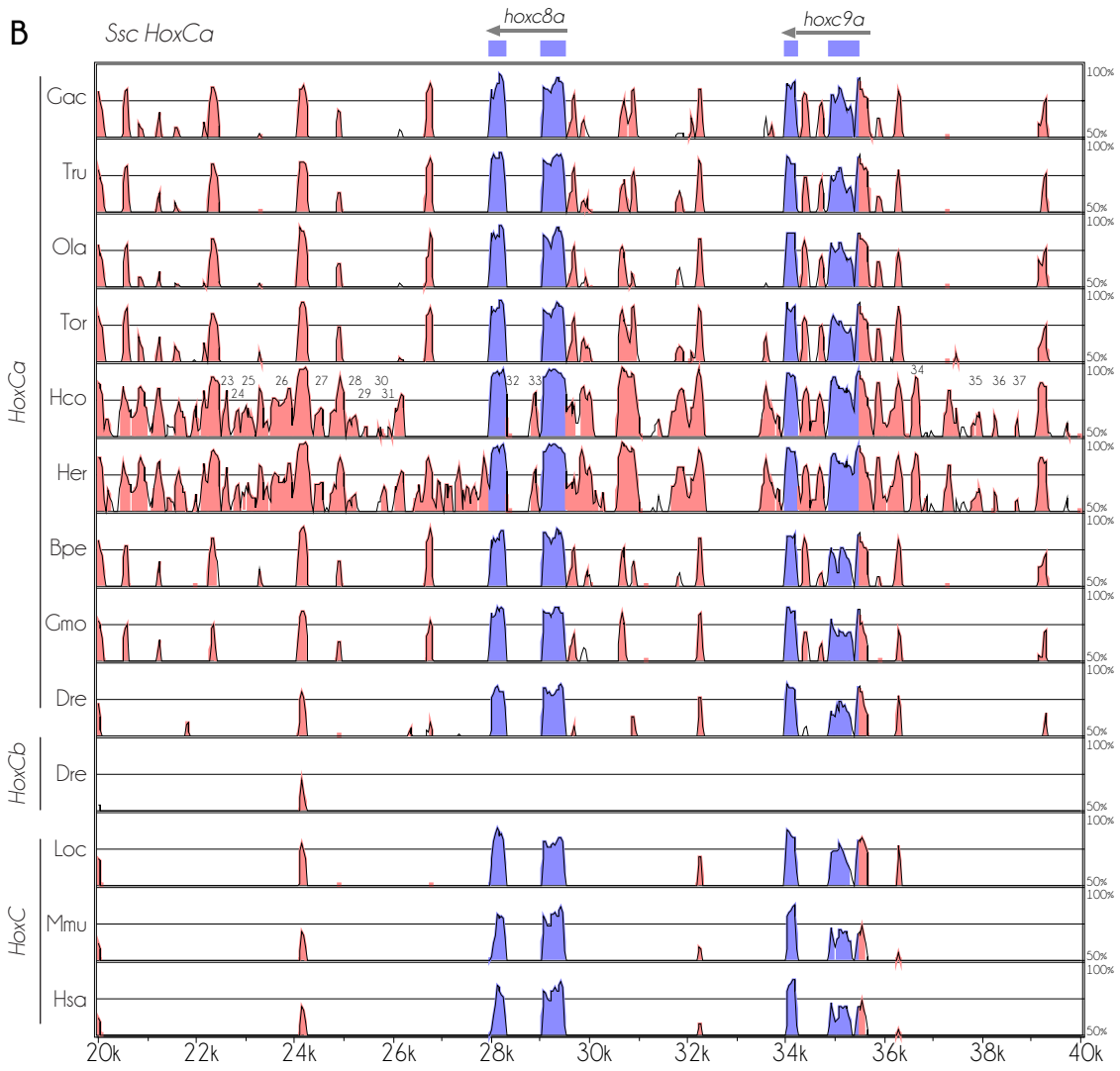


Figure S4.5 continued.

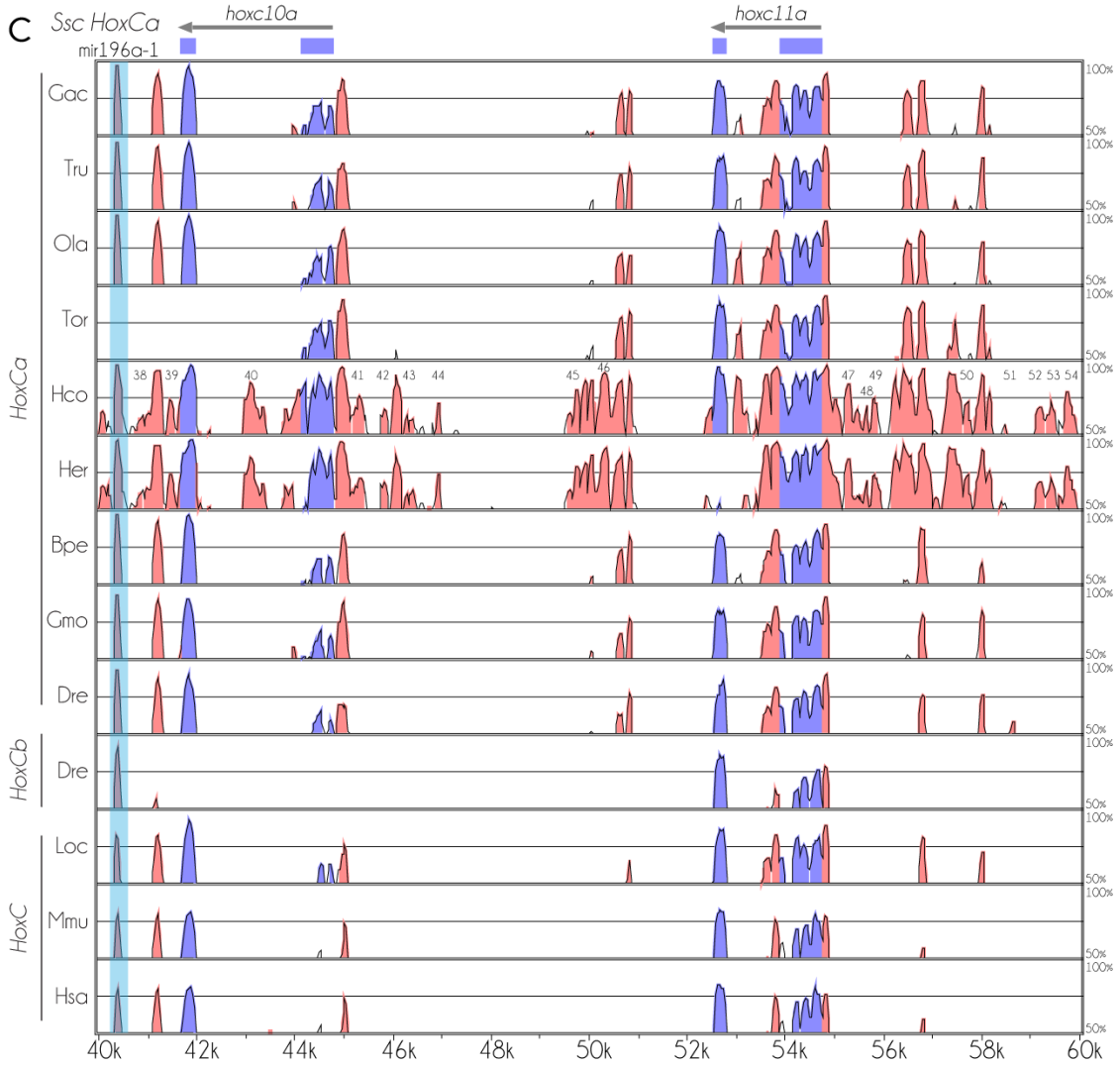


Figure S4.5 continued.

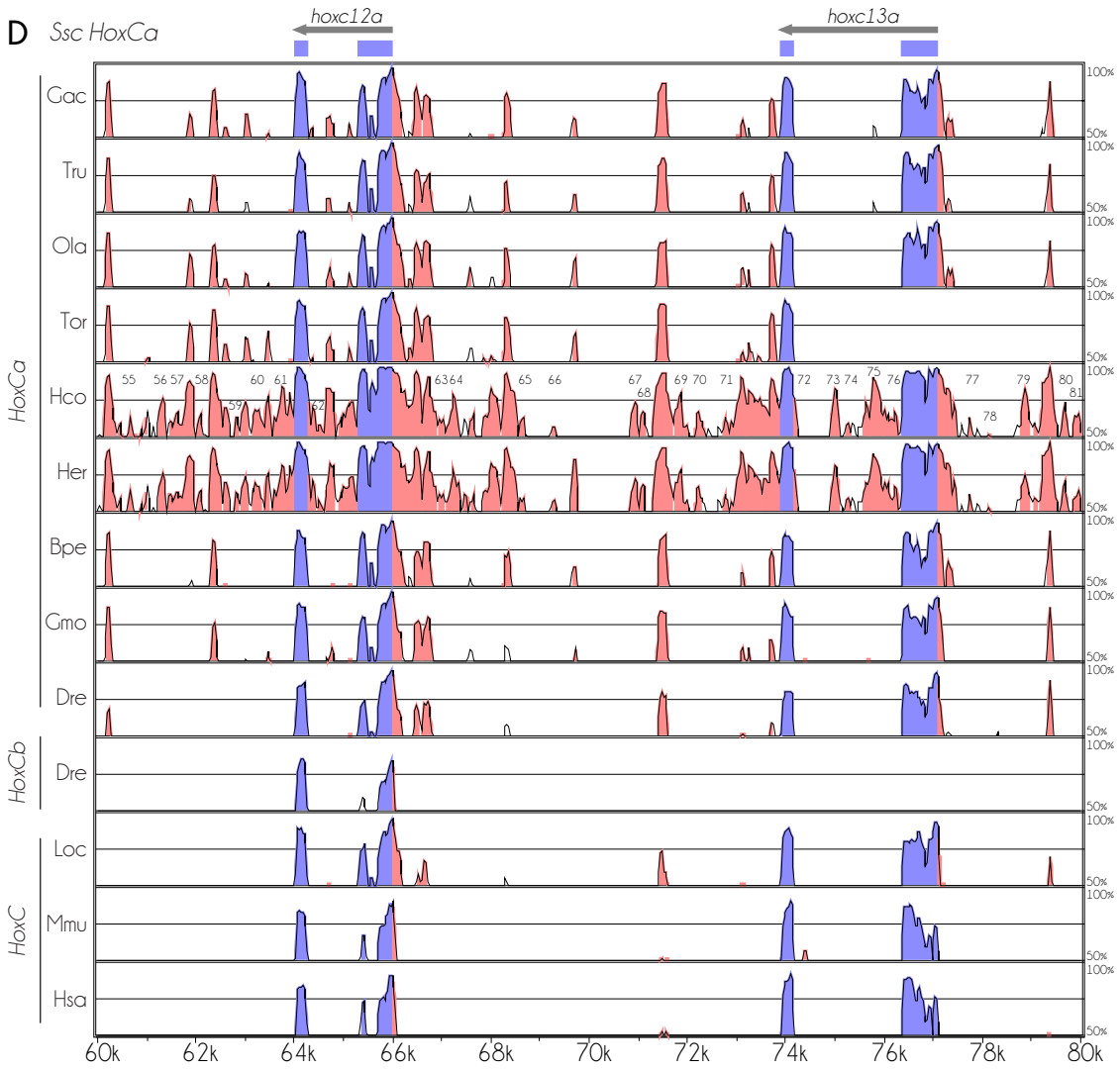


Figure S4.5 continued.

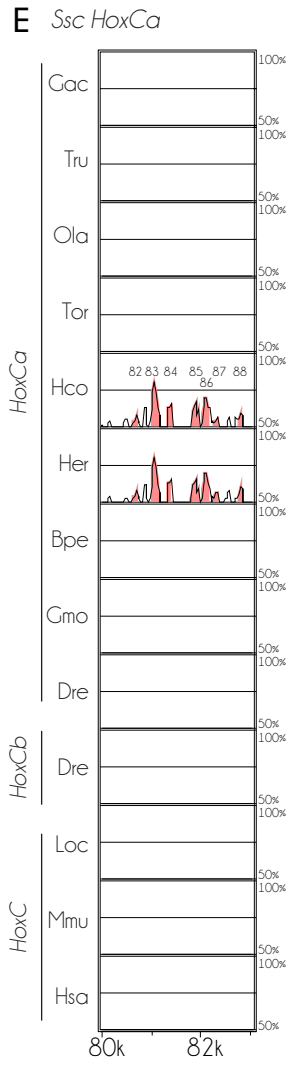


Figure S4.5 continued.

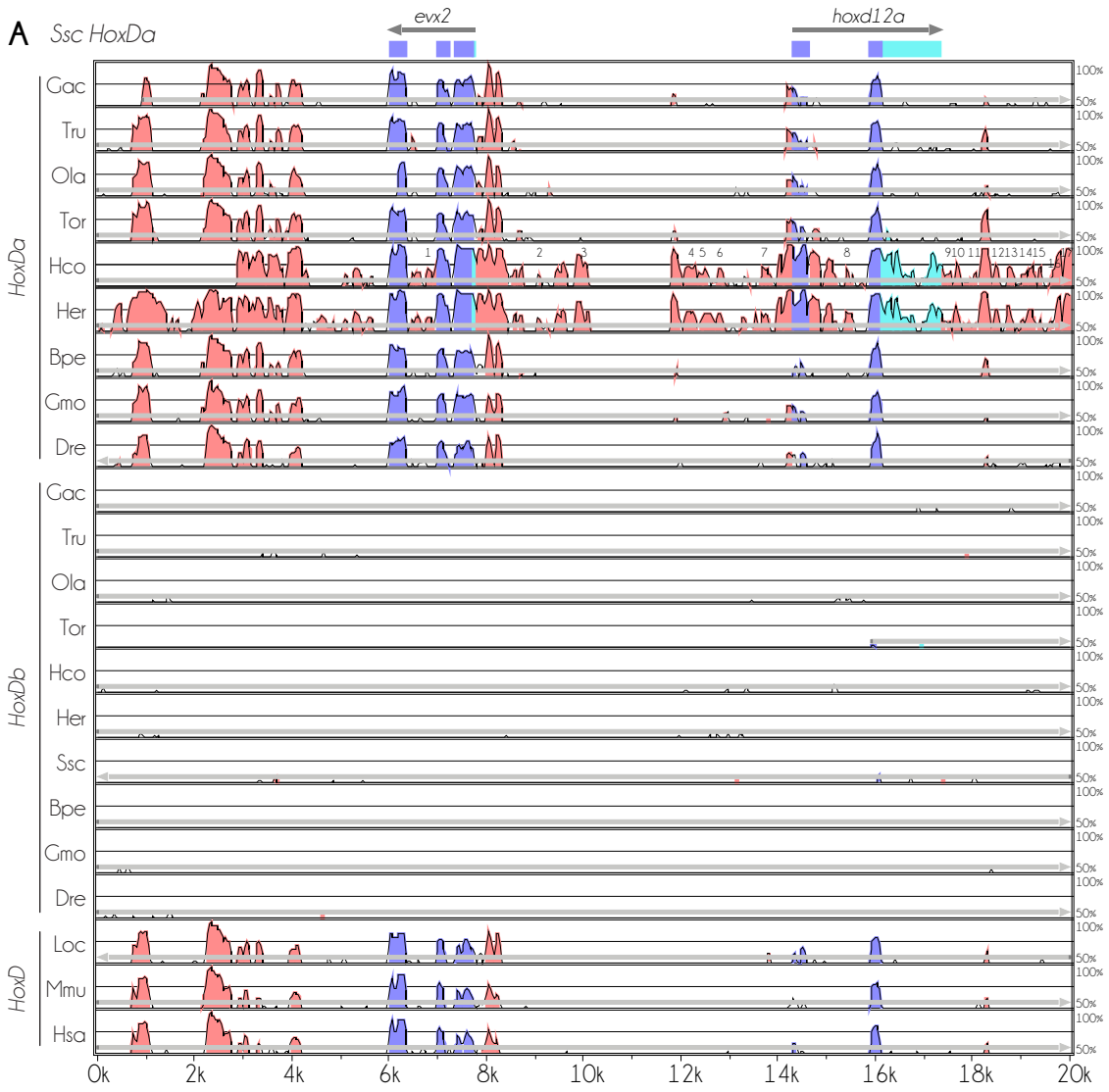


Figure S4. 6: VISTA plots for the *HoxD* clusters with Gulf pipefish *HoxDa* set as reference sequence. Exons are highlighted in blue, CNEs in pink, UTRs in teal, microRNAs are in the blue boxes. Shuffle LAGAN alignment was used with gray lines indicate stretches of continuous sequence. The reference, Ssc, is Gulf pipefish; Tru, pufferfish; Ola, medaka; Tor, tuna; Gac, stickleback; Hco, *tiger tail* seahorse; Her, lined seahorse; Bpe, mudskipper; Gmo, cod; Dre, zebrafish; Loc, spotted gar; Mmu, mouse; Hsa, human. Syngnathid specific CNEs are numbered on the *tiger tail* seahorse.

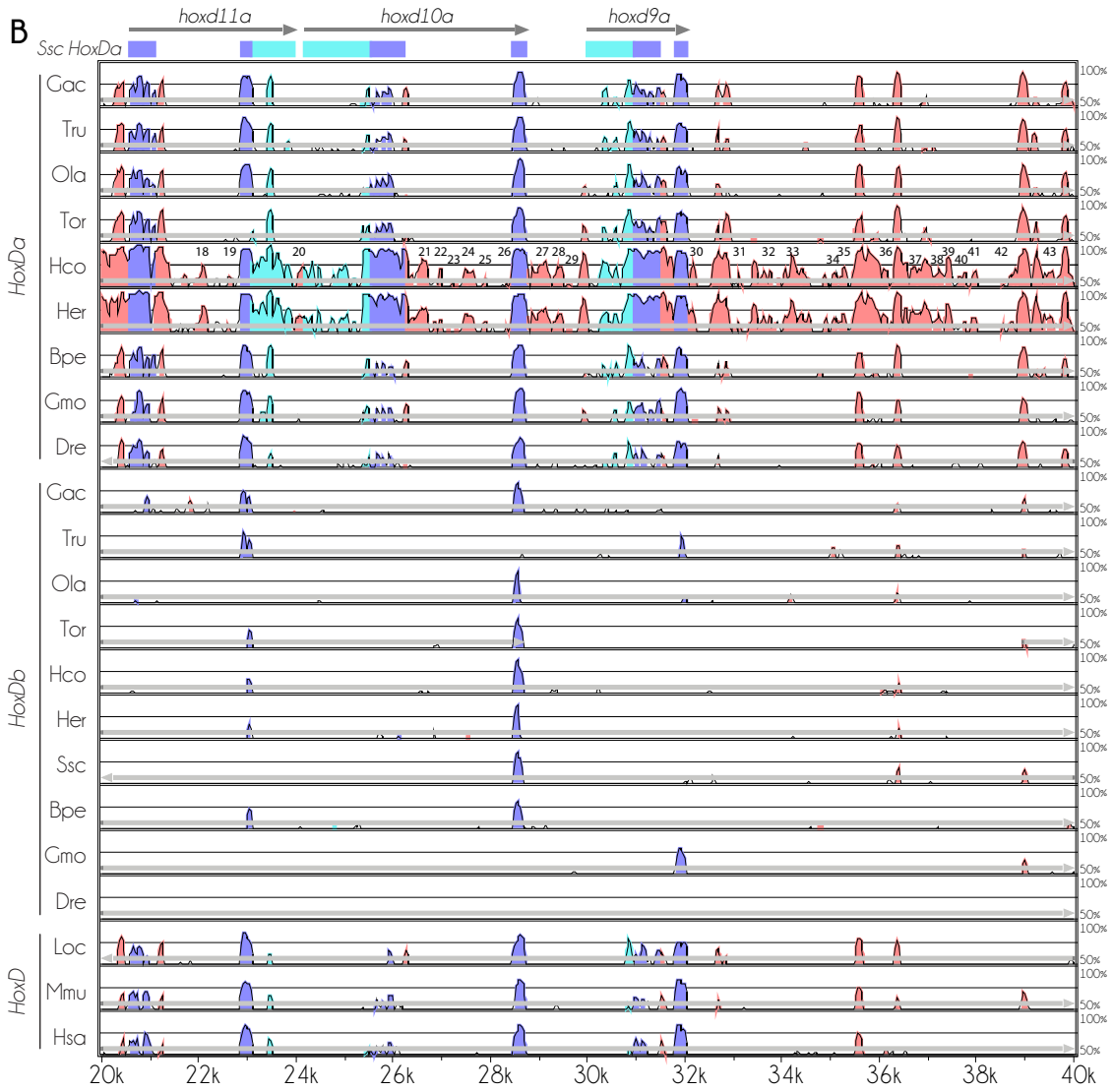


Figure S4.6 continued.

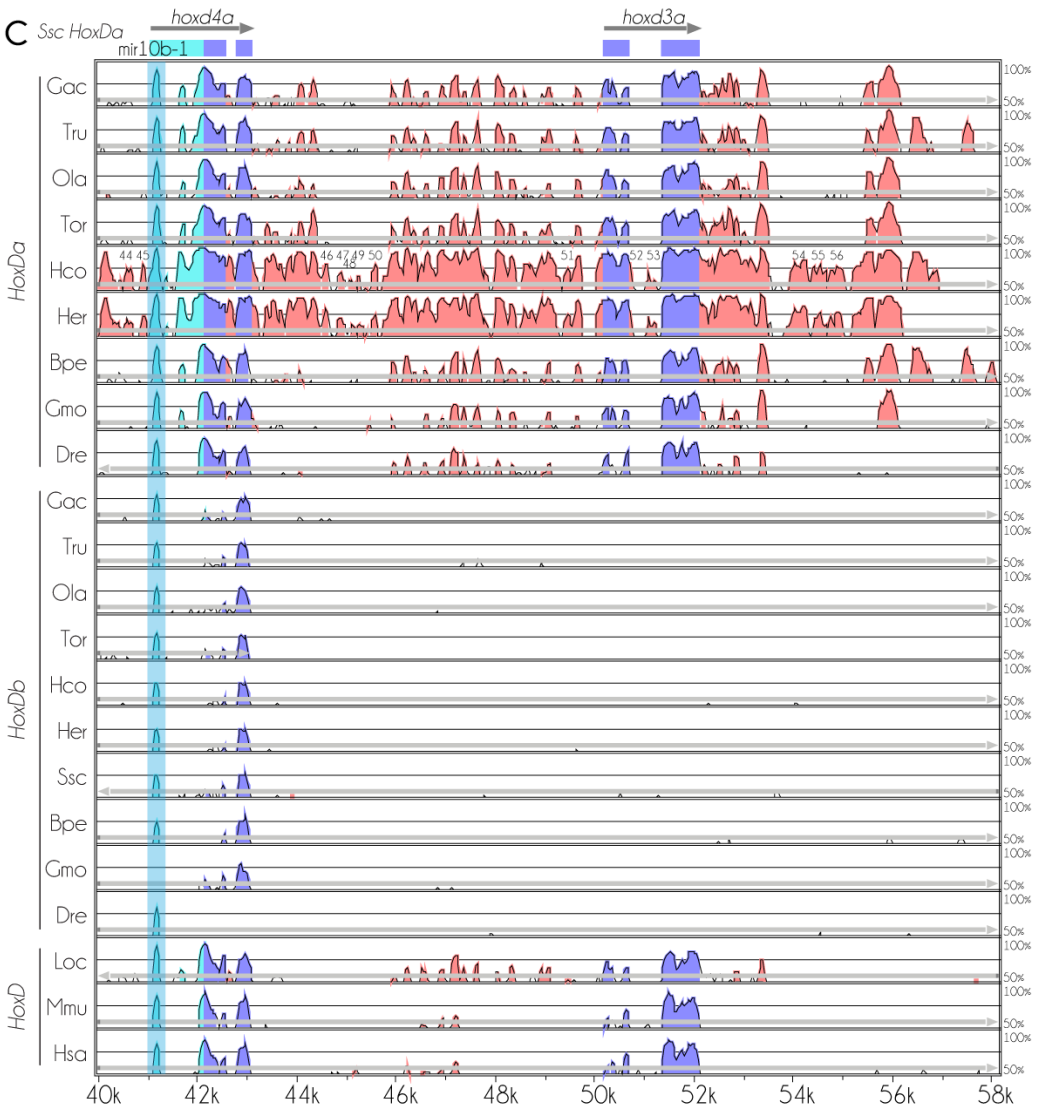


Figure S4.6 continued.

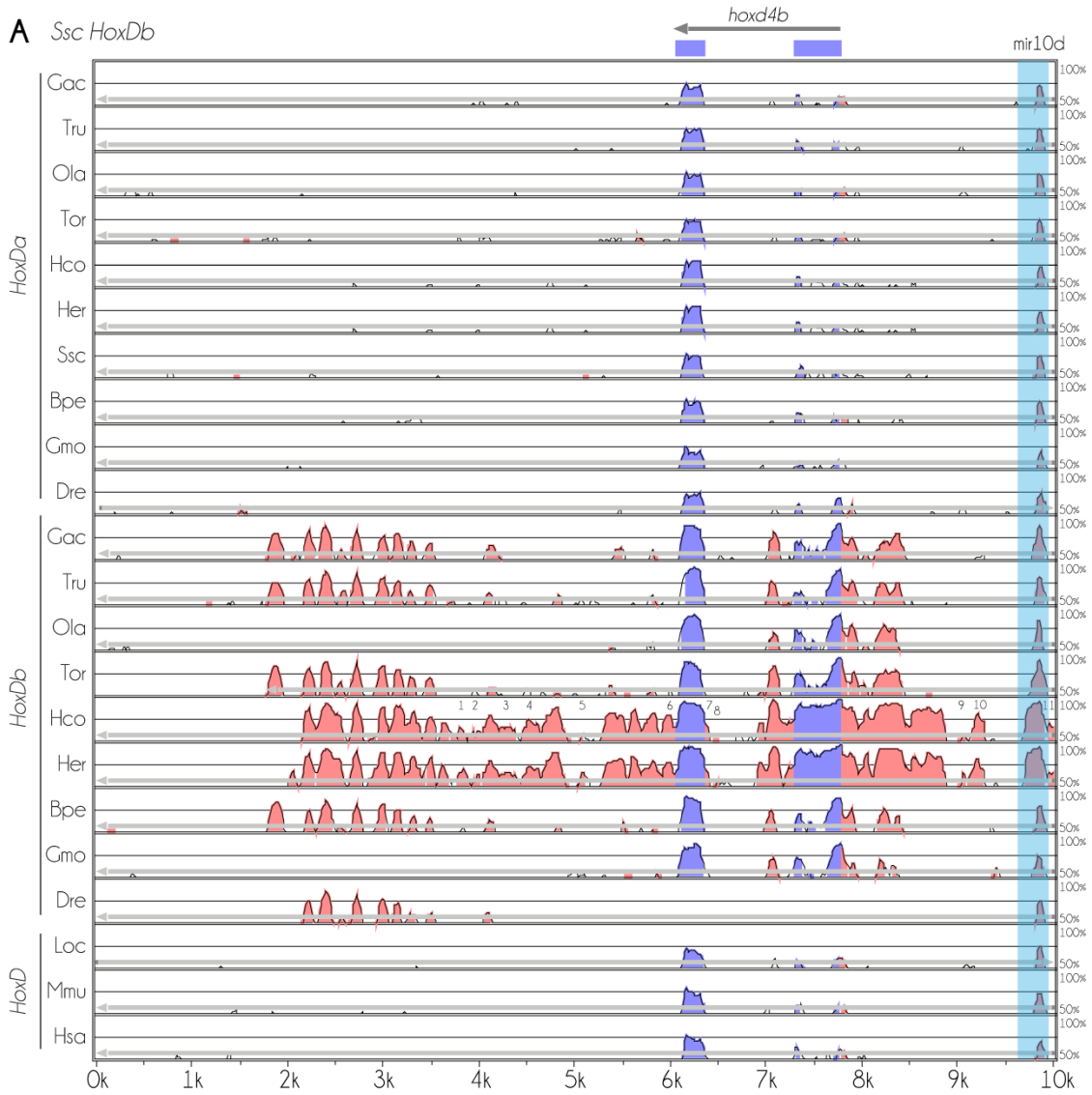


Figure S4.7: VISTA plots for the *HoxD* clusters with Gulf pipefish *HoxDb* set as reference sequence. Exons are highlighted in blue, CNEs in pink, UTRs in teal, microRNAs are in the blue boxes. Shuffle LAGAN alignment was used with gray lines indicate stretches of continuous sequence. The reference, Ssc, is Gulf pipefish; Tru, pufferfish; Ola, medaka; Tor, tuna; Gac, stickleback; Hco, *tiger tail* seahorse; Her, lined seahorse; Bpe, mudskipper; Gmo, cod; Dre, zebrafish; Loc, spotted gar; Mmu, mouse; Hsa, human. Syngnathid specific CNEs are numbered on the *tiger tail* seahorse.

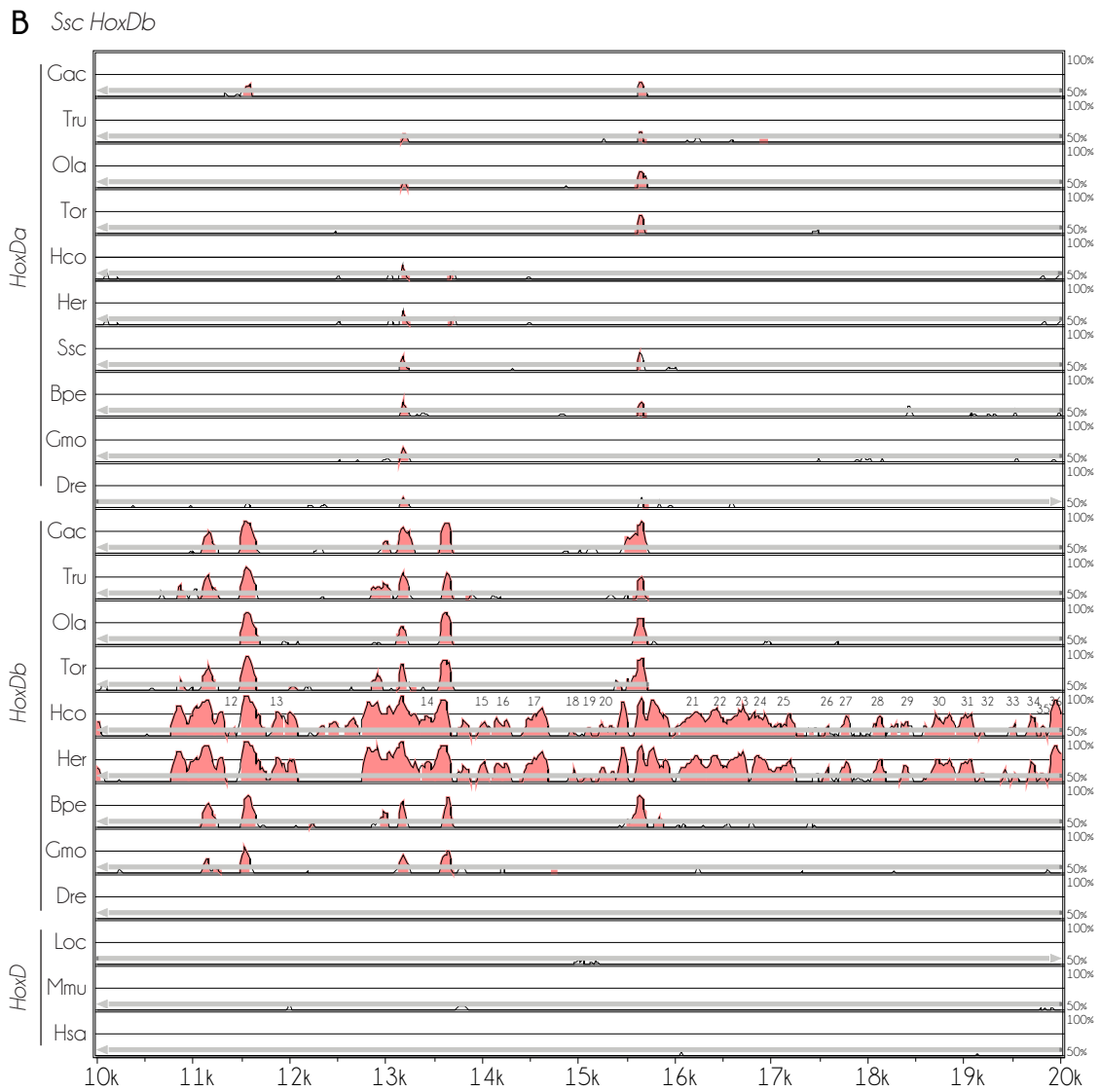


Figure S4.7 continued.

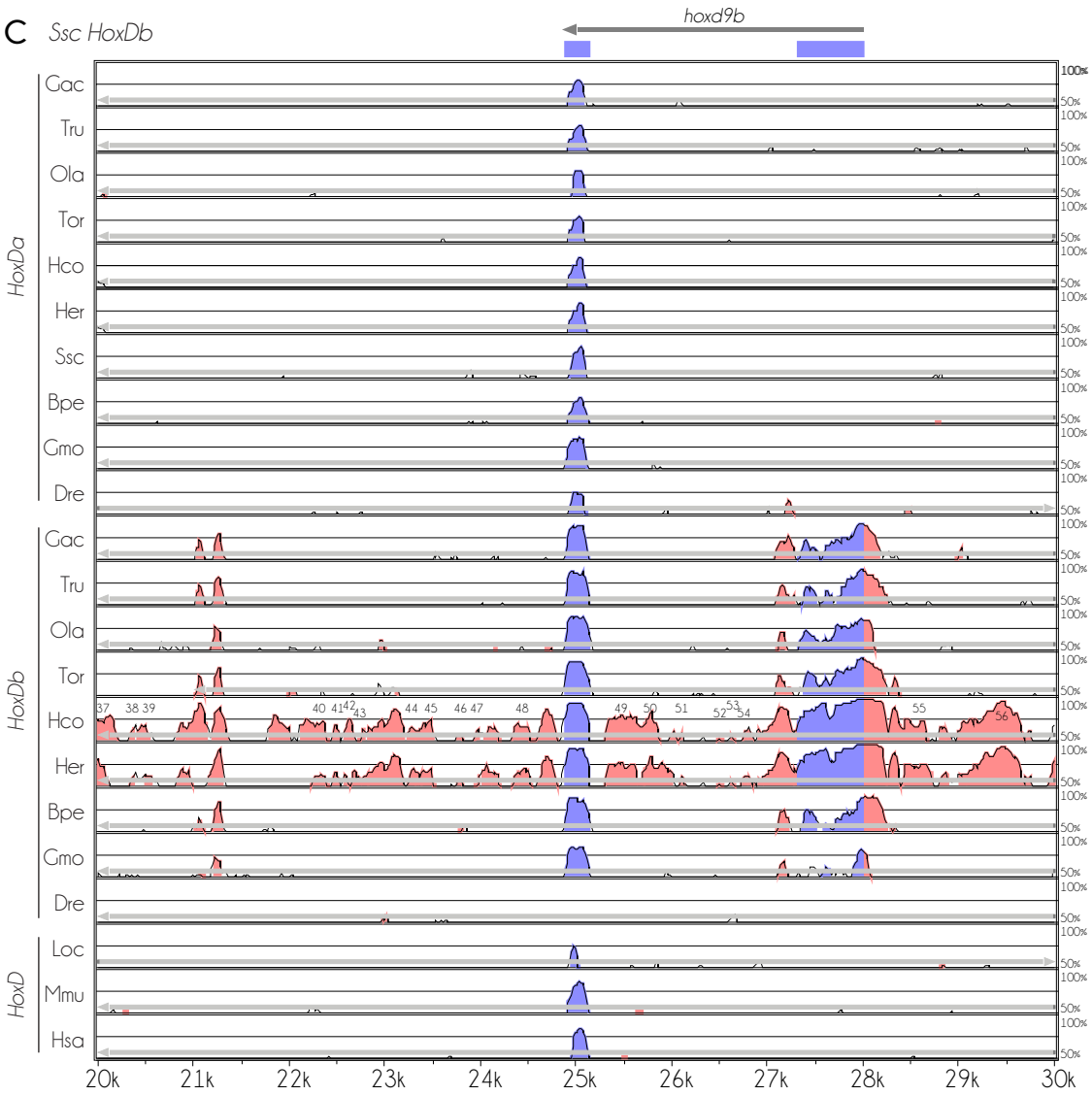


Figure S4.7 continued.

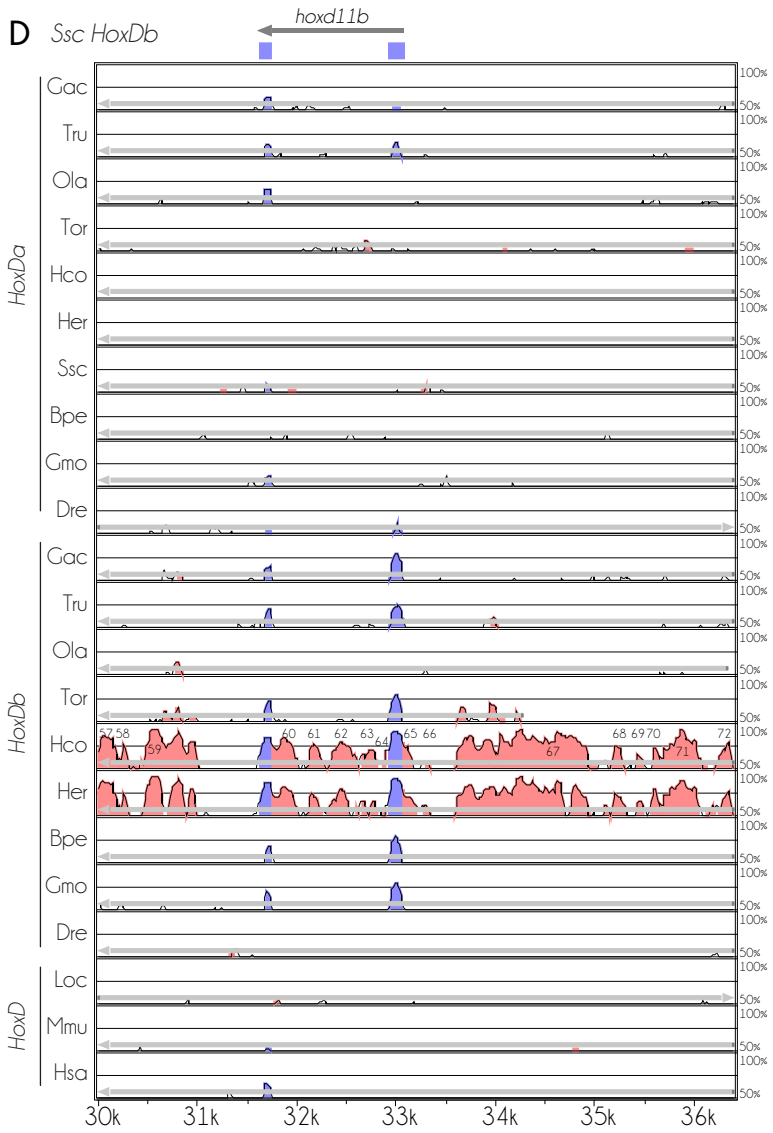


Figure S4.7 continued.

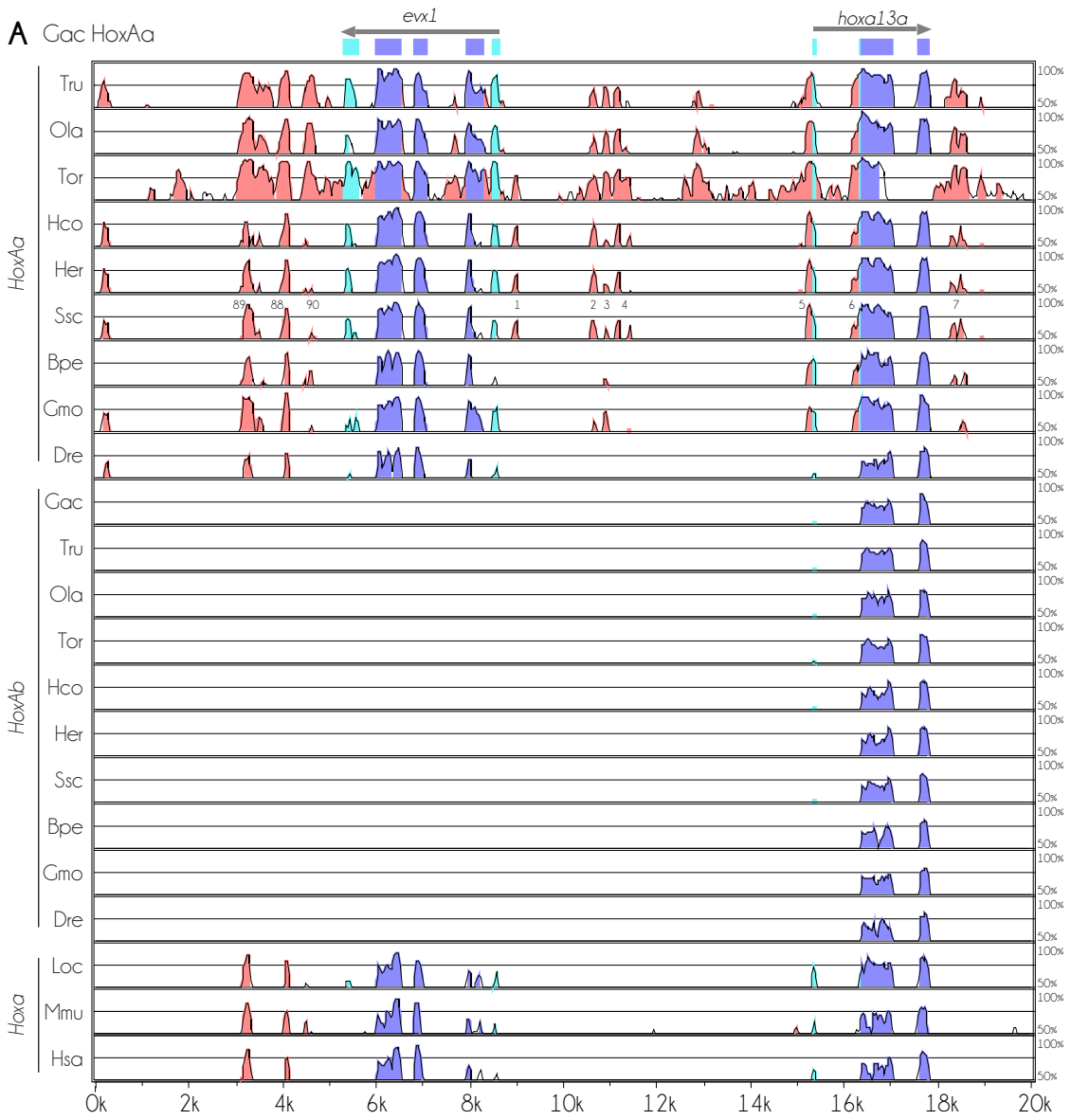


Figure S4.8: VISTA plots for the *HoxA* clusters with threespine stickleback *HoxAa* set as reference sequence. Exons are highlighted in blue, CNEs in pink, UTRs in teal, microRNAs are in the blue boxes. LAGAN alignment was used. The reference, Gac, is the threespine stickleback; Tru, pufferfish; Ola, medaka; Tor, tuna; Ssc, pipefish; Hco, *tiger tail* seahorse; Her, lined seahorse; Bpe, mudskipper; Gmo, cod; Dre, zebrafish; Loc, spotted gar; Mmu, mouse; Hsa, human. Syngnathid CNEs are numbered on the Gulf pipefish. Syngnathid specific losses are in red boxes.

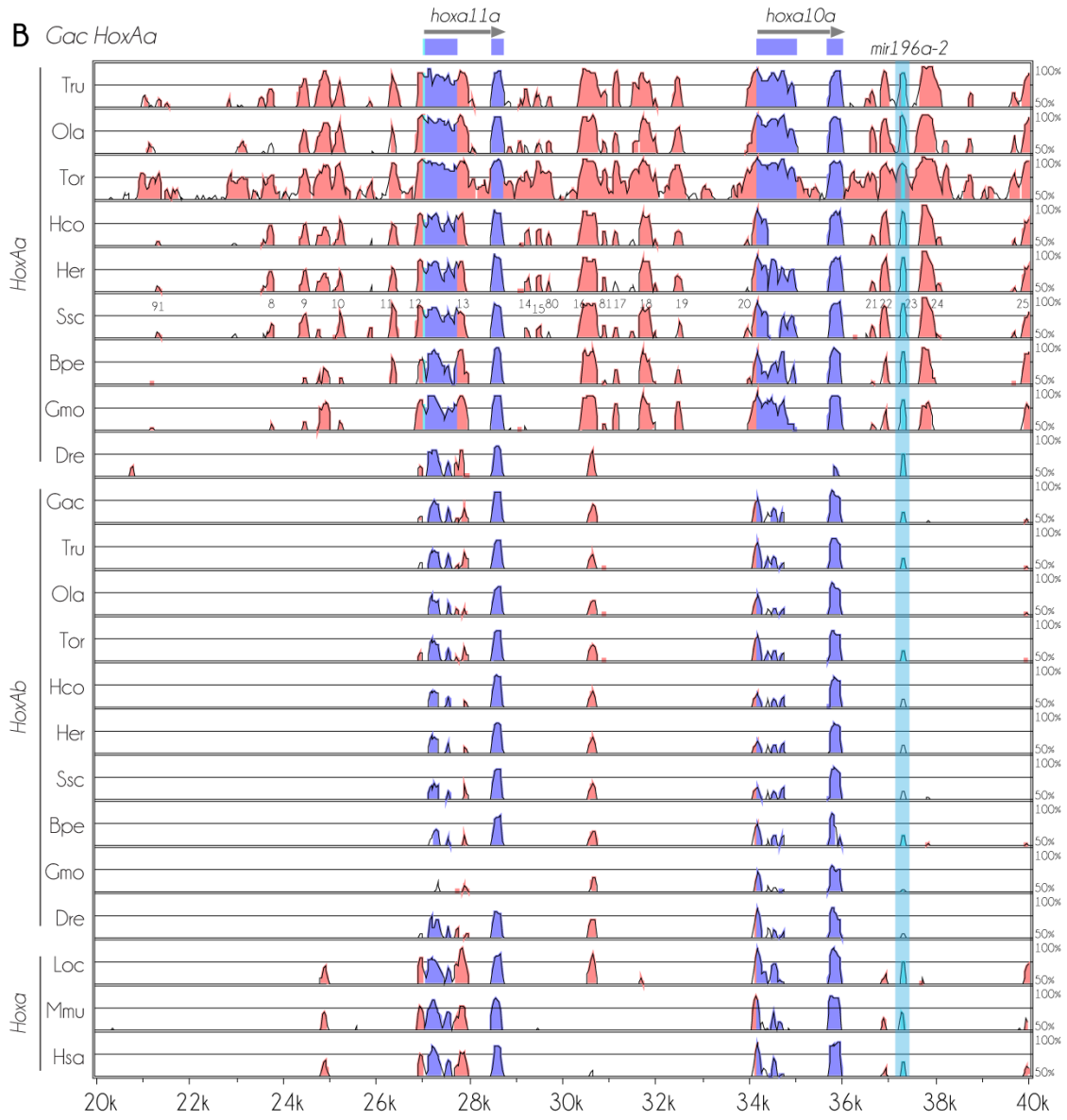


Figure S4.8 continued.

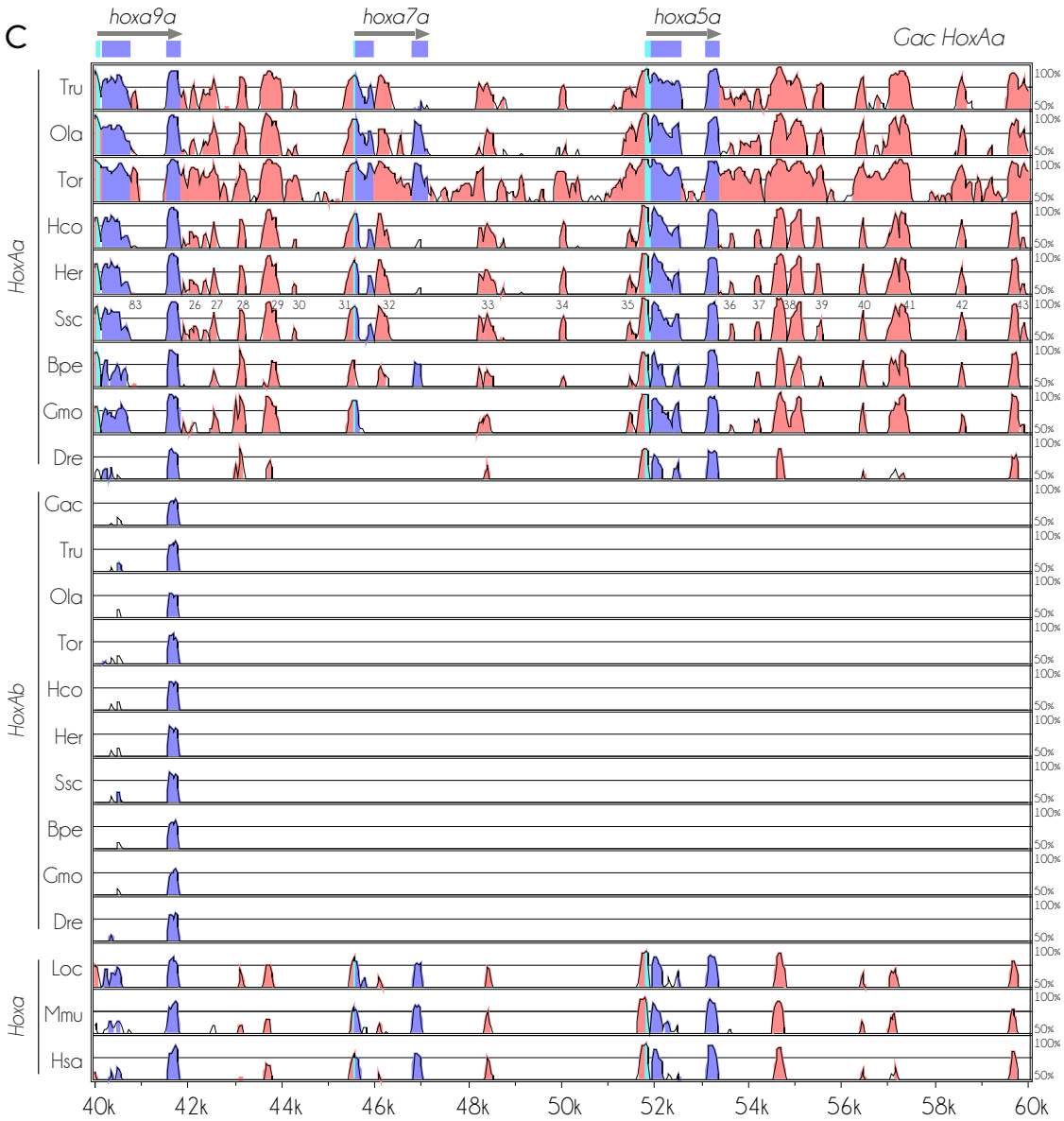


Figure S4.8 continued.

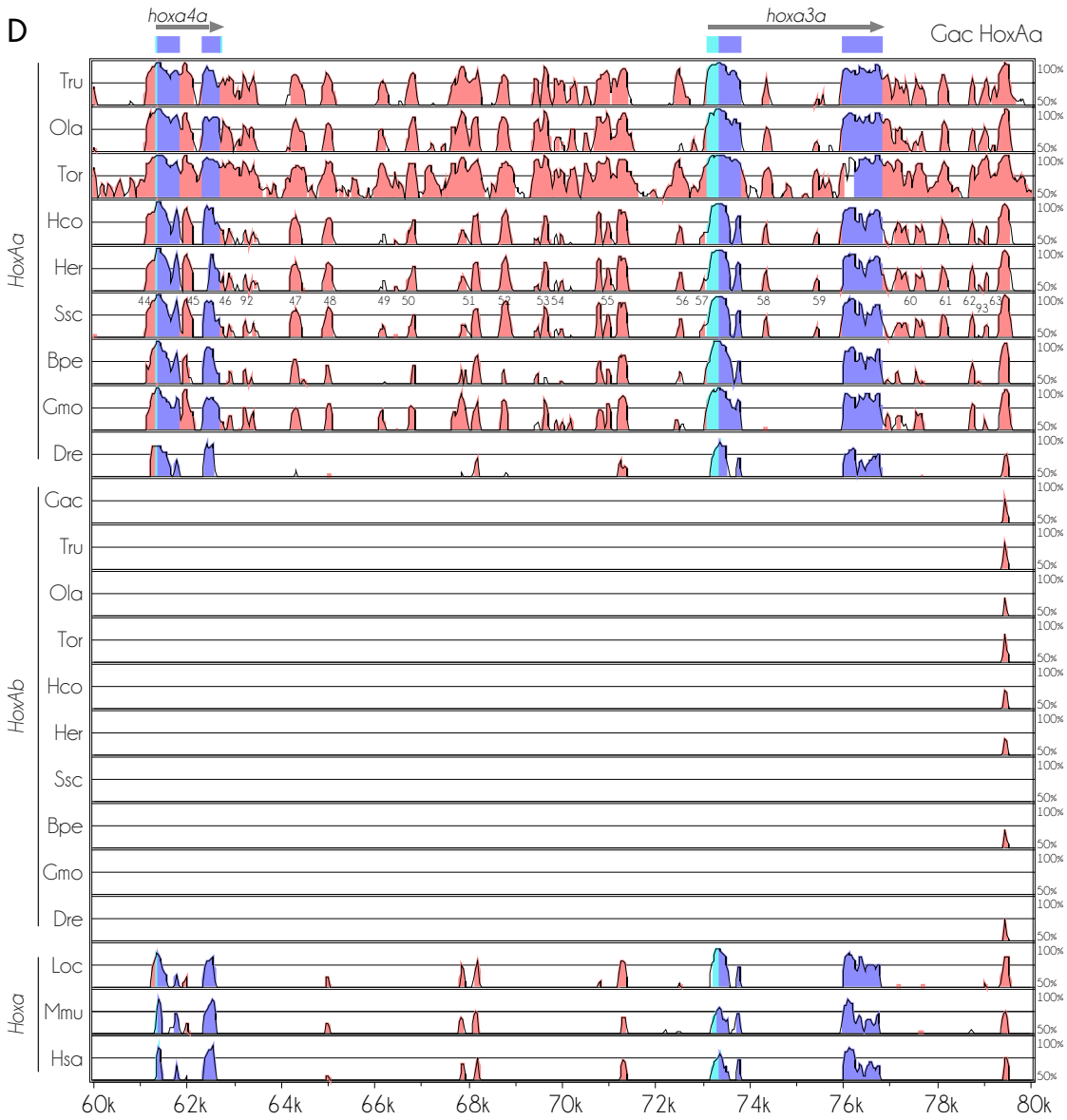


Figure S4.8 continued.

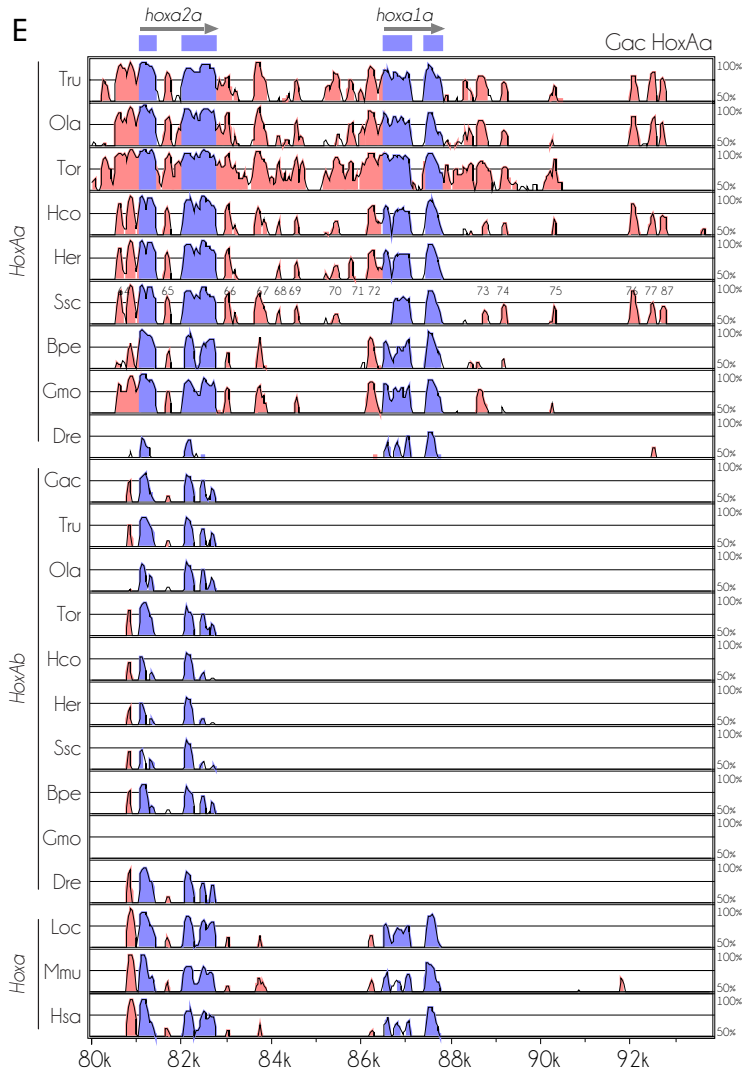


Figure S4.8 continued.

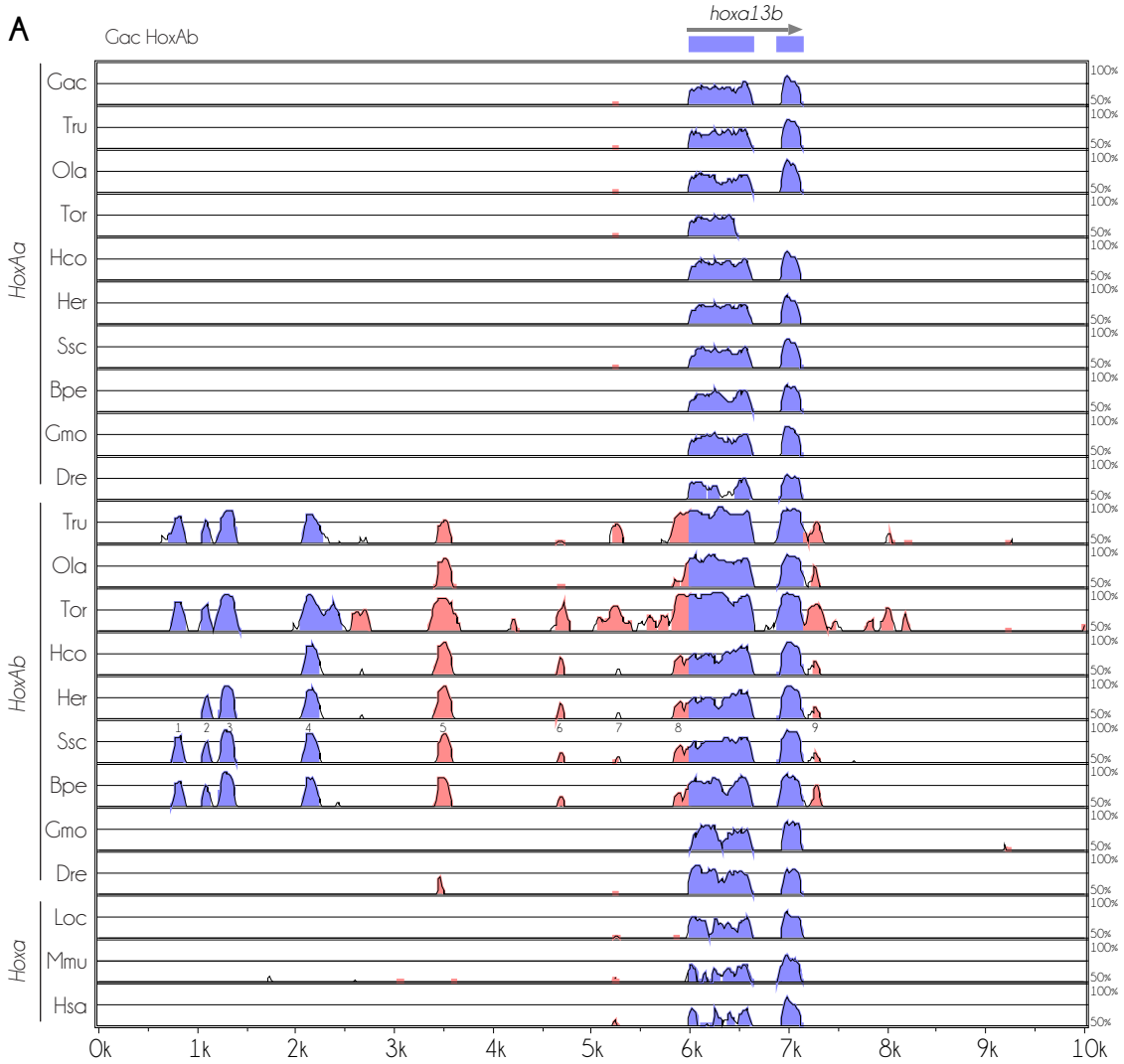


Figure S4.9: VISTA plots for the *HoxA* clusters with threespine stickleback *HoxAb* set as reference sequence. Exons are highlighted in blue, CNEs in pink, UTRs in teal, microRNAs are in the blue boxes. LAGAN alignment was used. The reference, Gac, is the threespine stickleback; Tru, pufferfish; Ola, medaka; Tor, tuna; Ssc, pipefish; Hco, *tiger tail* seahorse; Her, lined seahorse; Bpe, mudskipper; Gmo, cod; Dre, zebrafish; Loc, spotted gar; Mmu, mouse; Hsa, human. Syngnathid CNEs are numbered on the Gulf pipefish. Syngnathid specific losses are in red boxes.

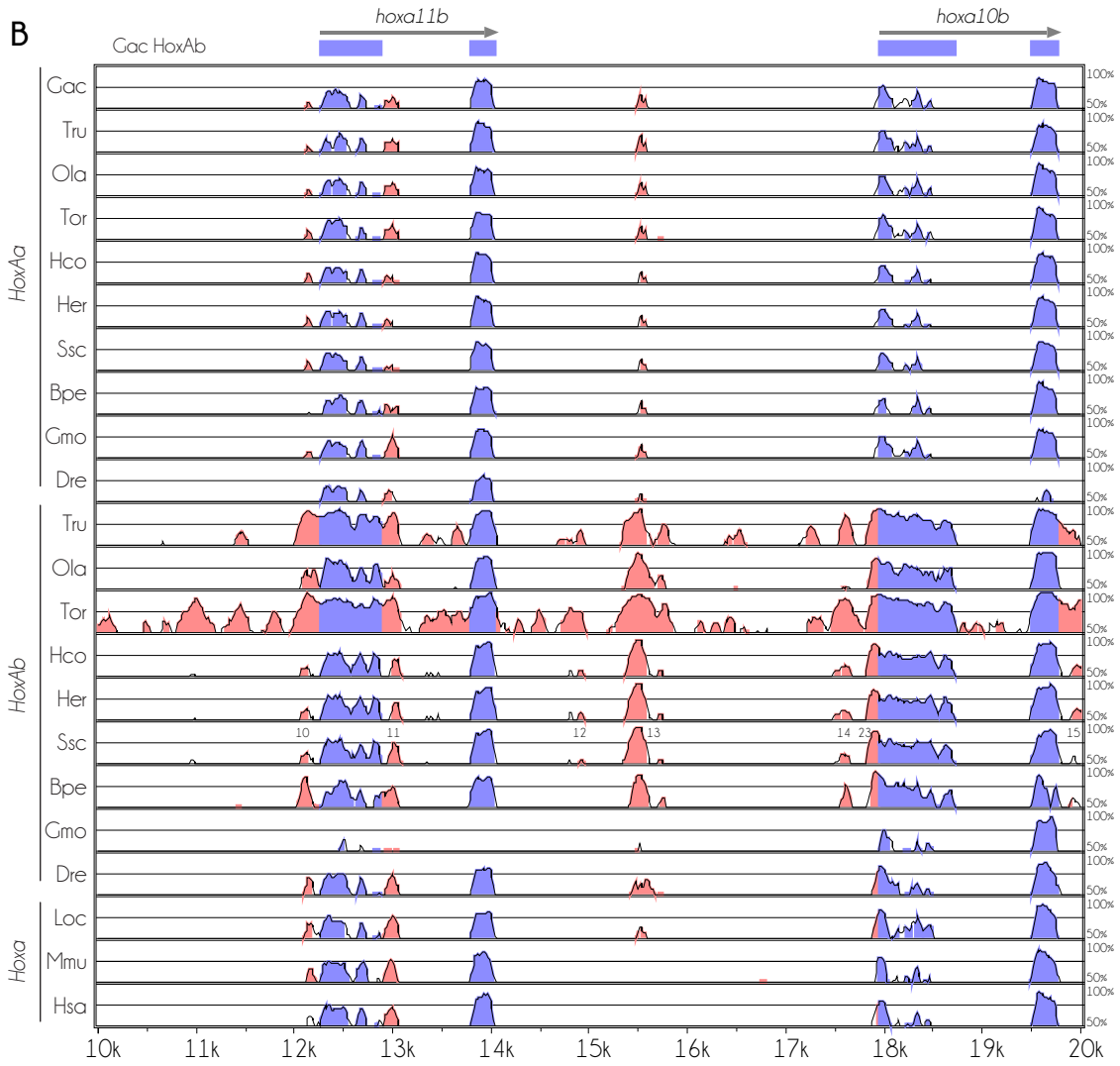


Figure S4.9 continued.

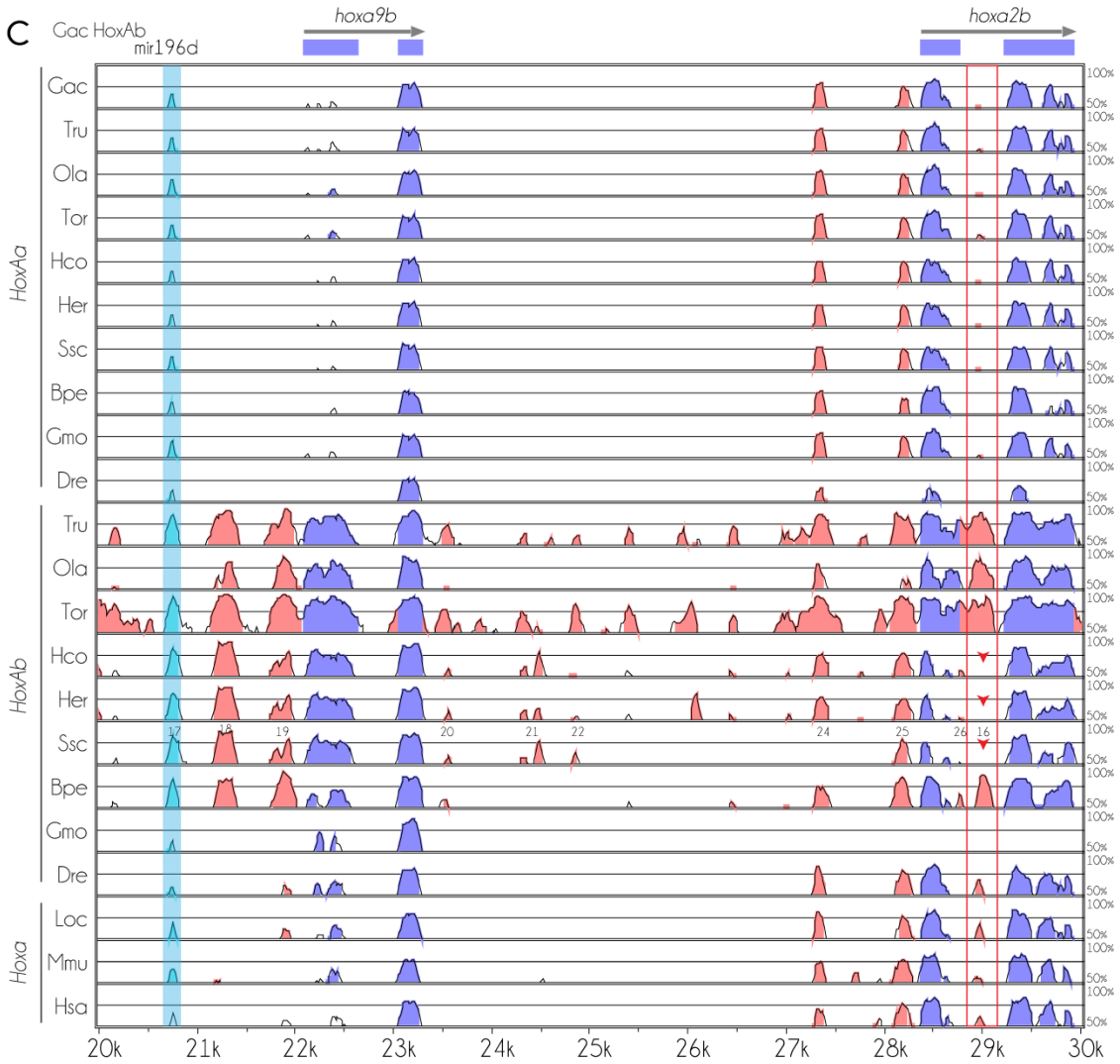


Figure S4.9 continued. CNE24 sequence is missing in the Gulf pipefish genome assembly.

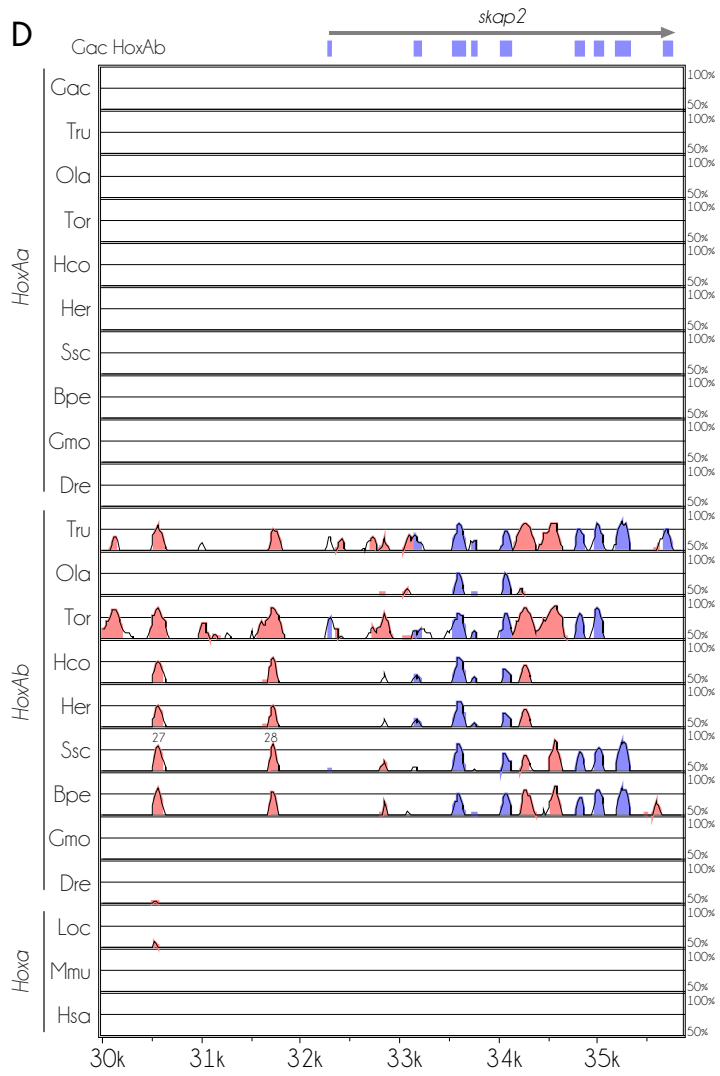


Figure S4.9 continued.

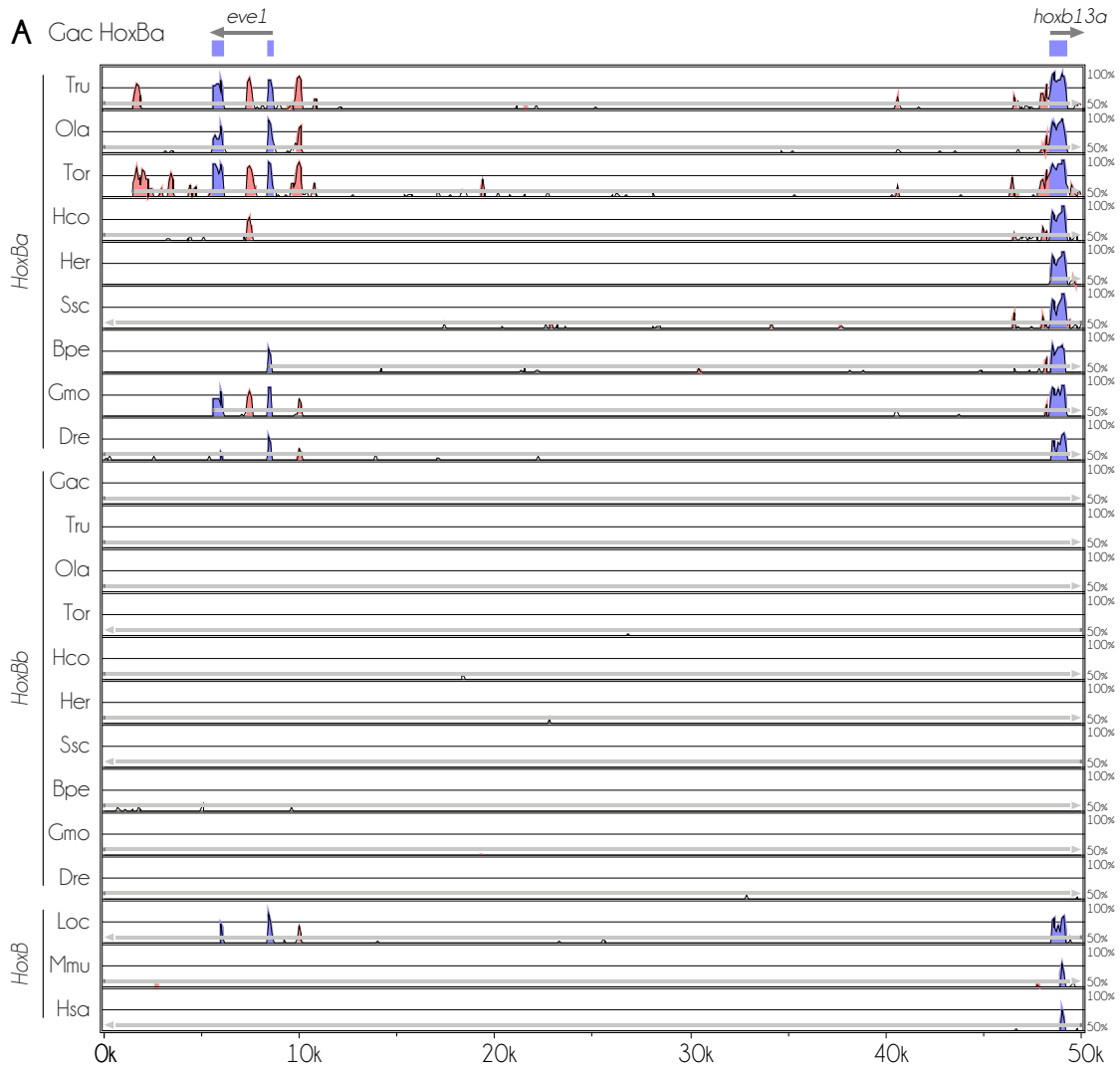


Figure S4.10: VISTA plots for the *HoxB* clusters with threespine stickleback *HoxBa* set as reference sequence. Exons are highlighted in blue, CNEs in pink, UTRs in teal, microRNAs are in the blue boxes. Shuffle LAGAN alignment was used with gray lines indicate stretches of continuous sequence. The reference, Gac, is the threespine stickleback; Tru, pufferfish; Ola, medaka; Tor, tuna; Ssc, pipefish; Hco, *tiger tail* seahorse; Her, lined seahorse; Bpe, mudskipper; Gmo, cod; Dre, zebrafish; Loc, spotted gar; Mmu, mouse; Hsa, human. Syngnathid CNEs are numbered on the Gulf pipefish. Syngnathid specific losses are in red boxes.

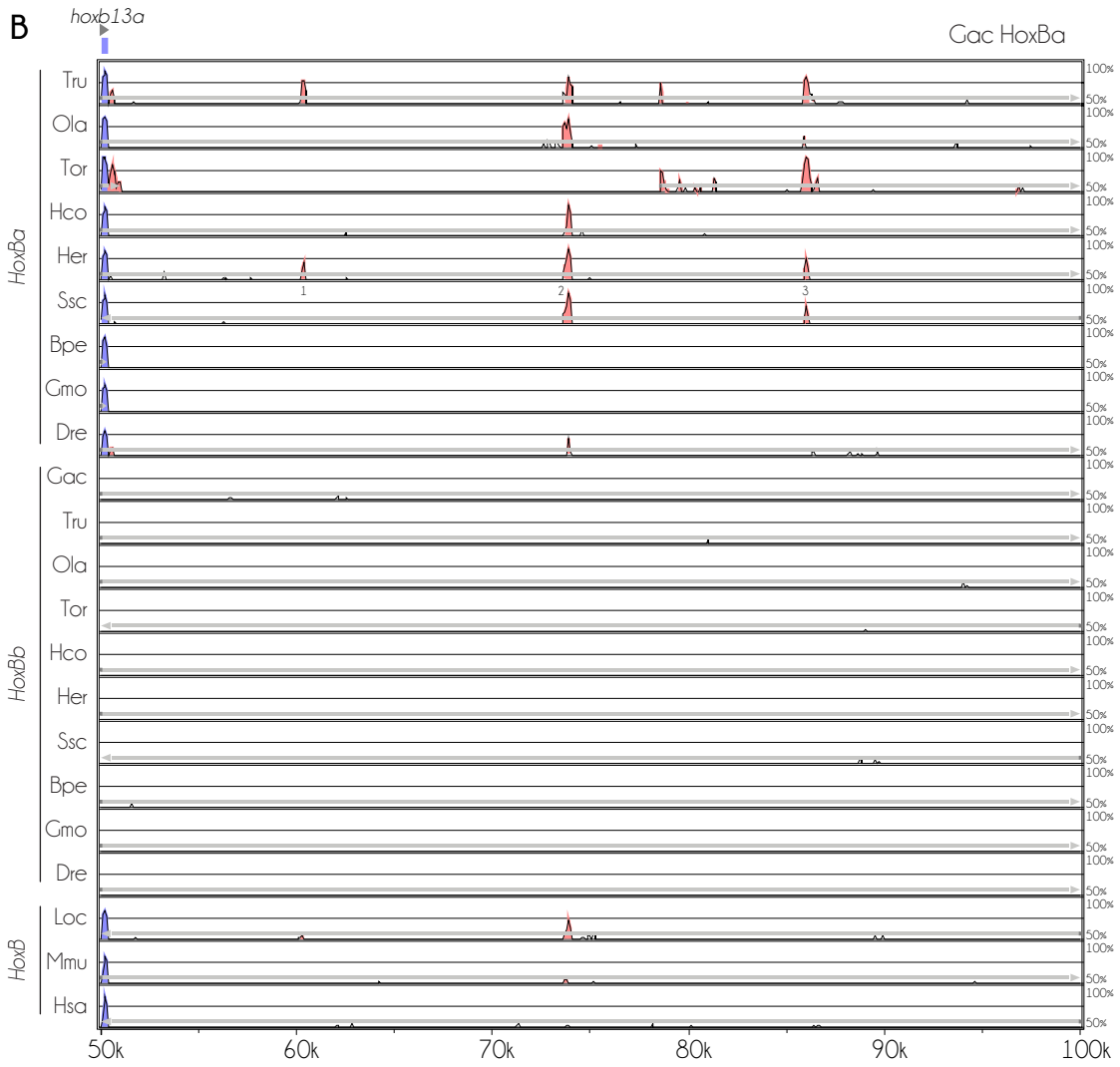


Figure S4.10 continued.

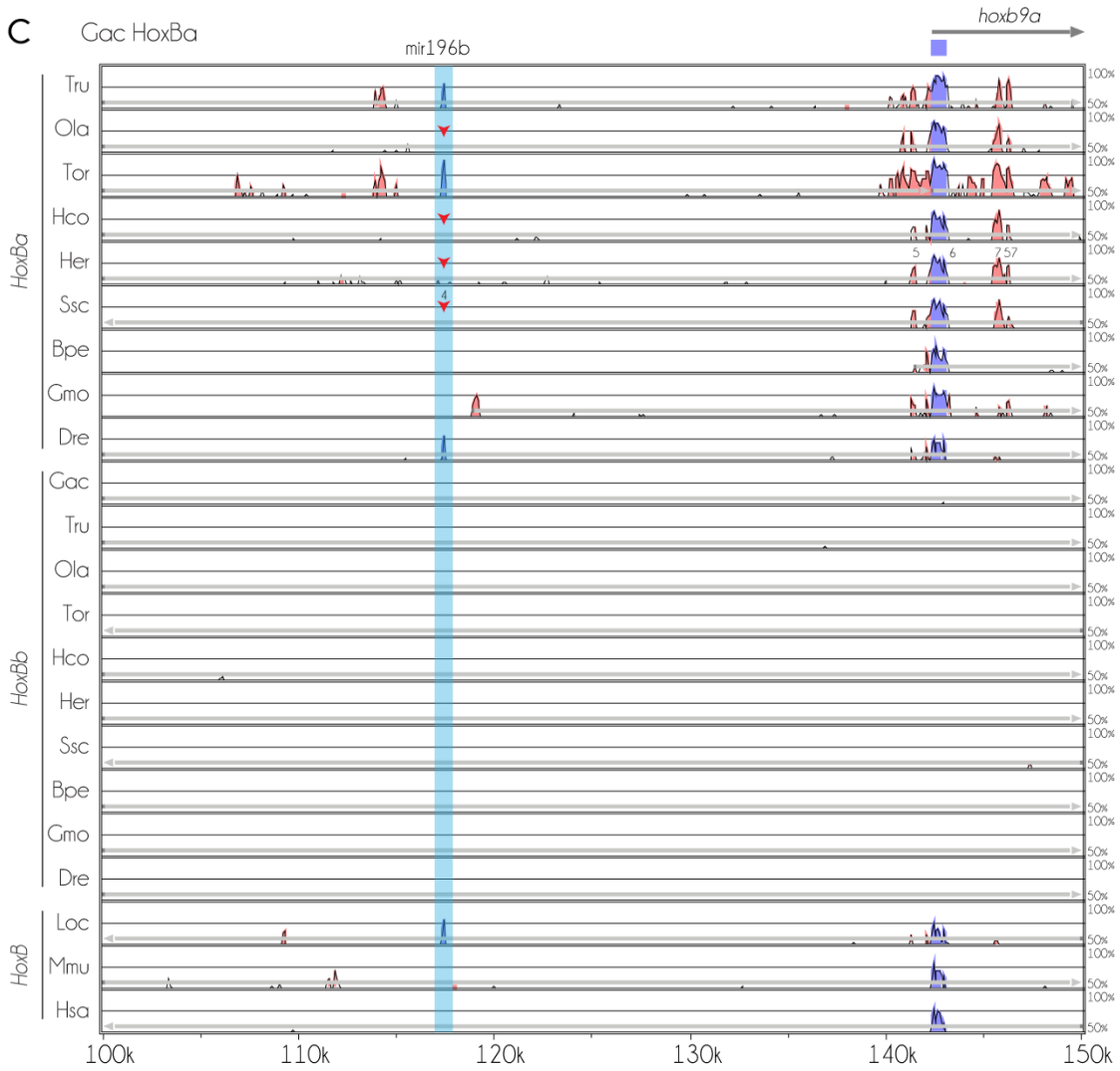


Figure S4.10 continued.

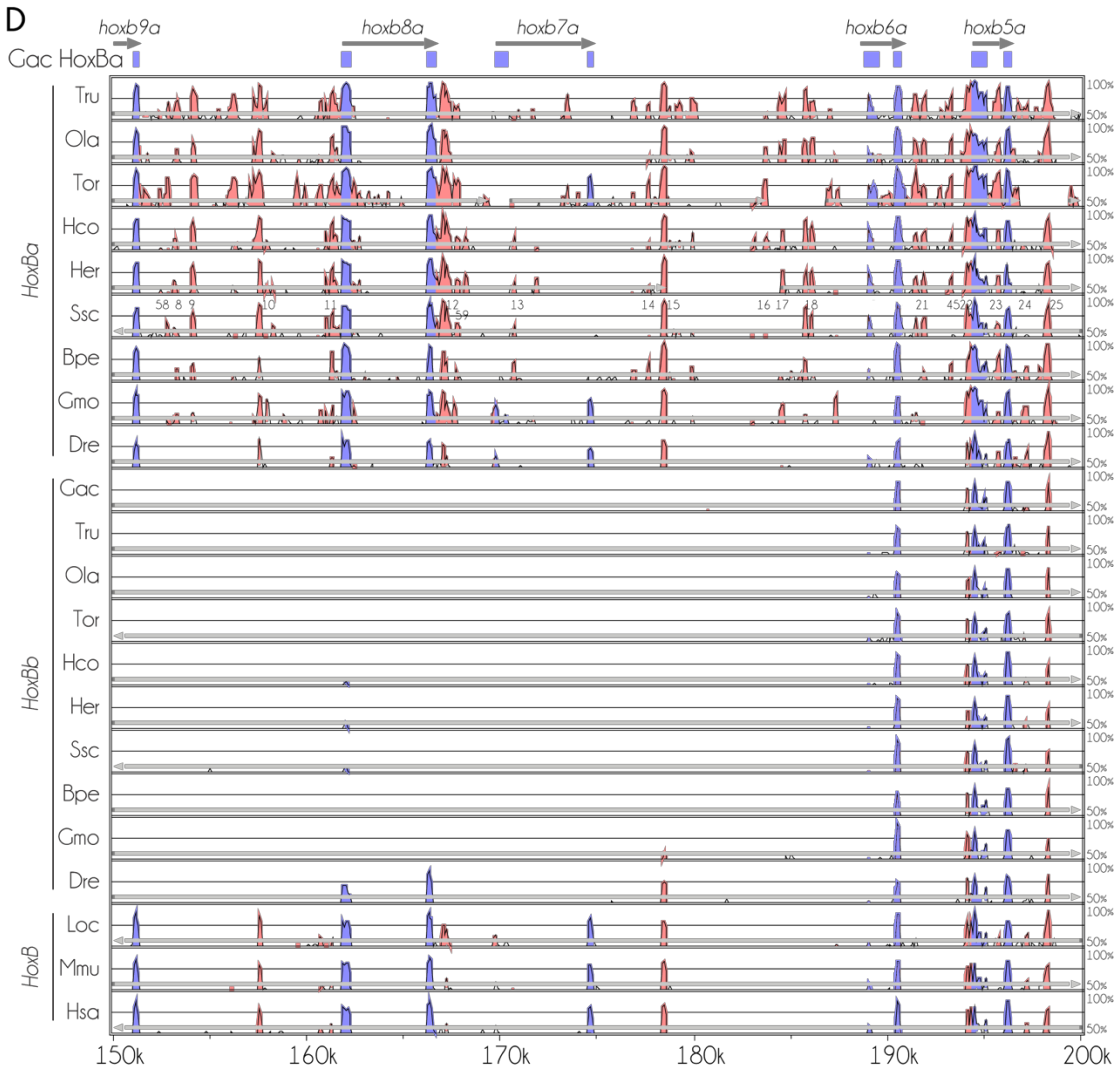


Figure S4.10 continued.

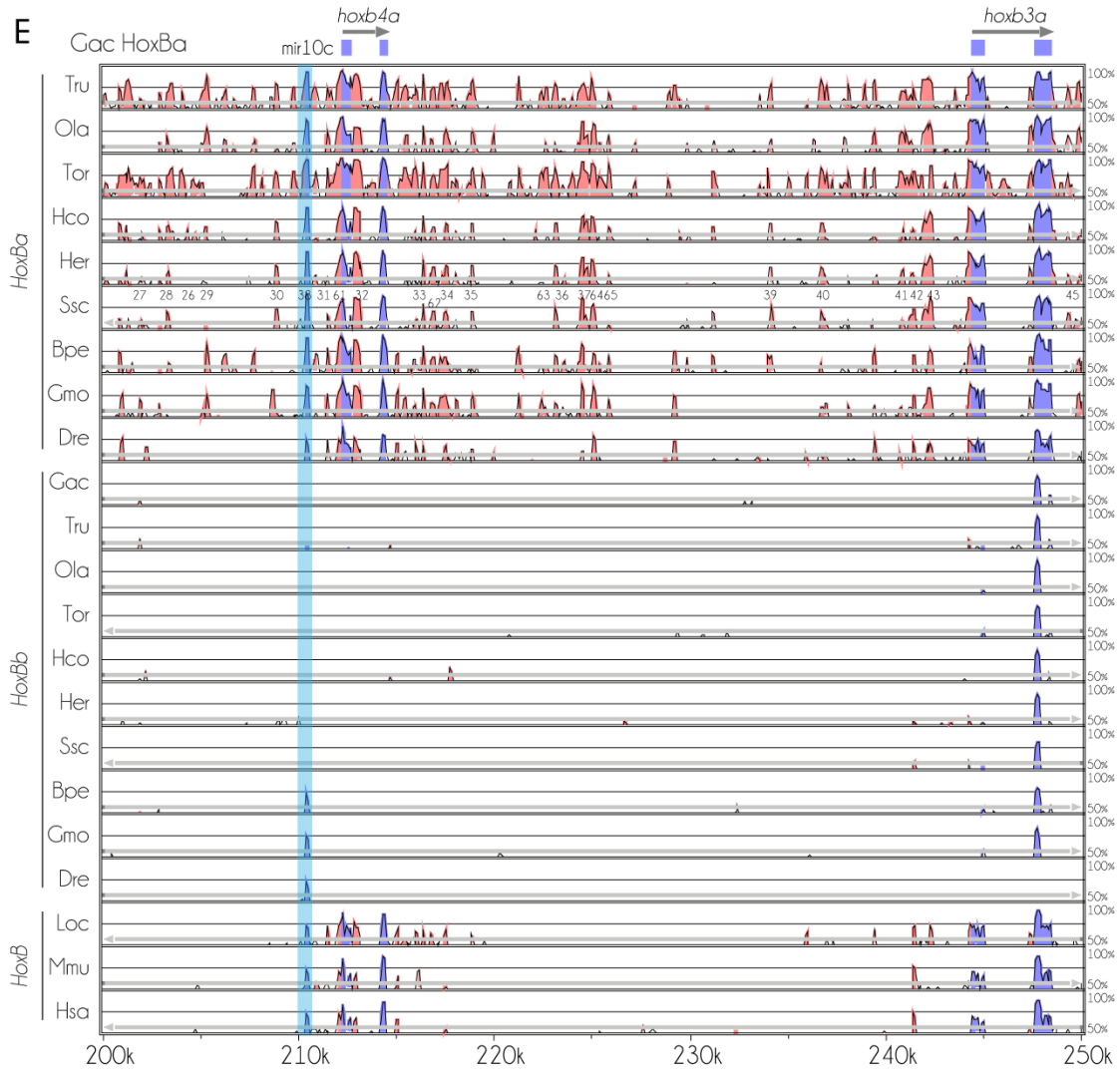


Figure S4.10 continued.

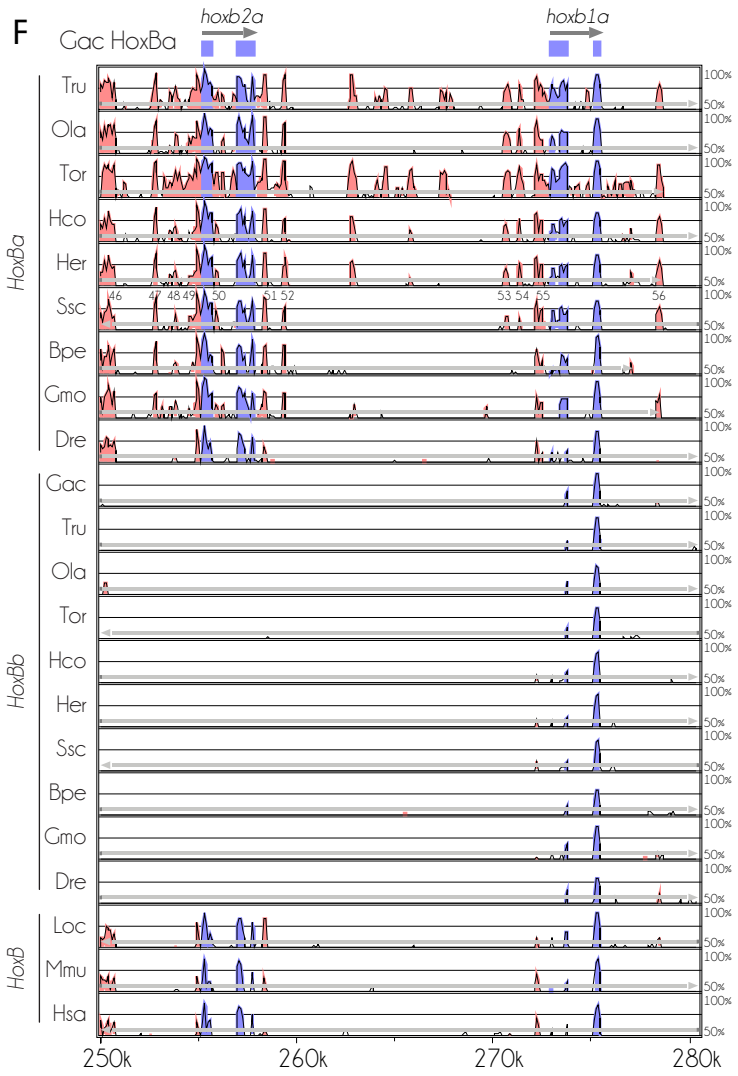


Figure S4.10 continued.

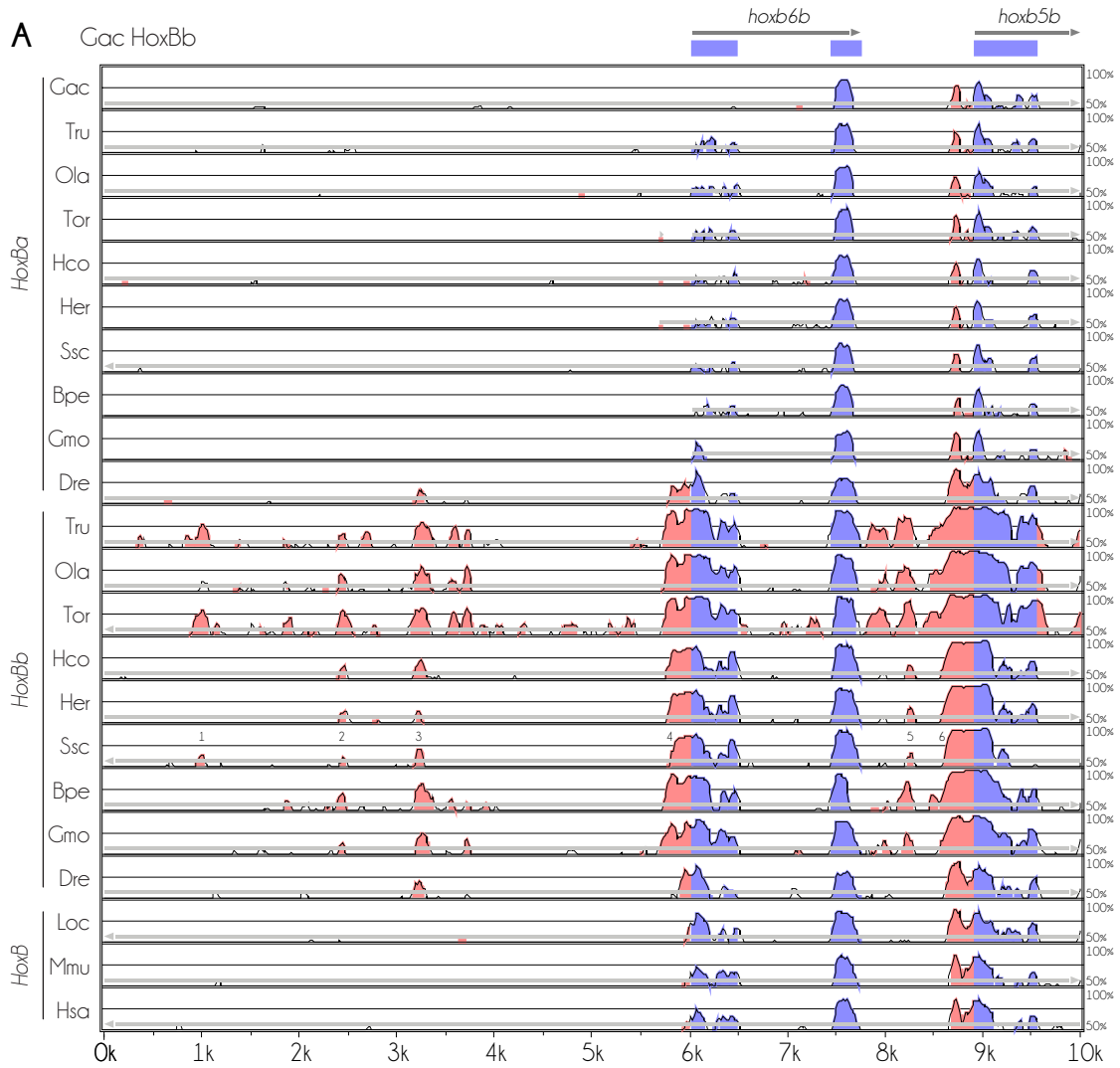


Figure S4.11: VISTA plots for the *HoxB* clusters with threespine stickleback *HoxBb* set as reference sequence. Exons are highlighted in blue, CNEs in pink, UTRs in teal, microRNAs are in the blue boxes. Shuffle LAGAN alignment was used with gray lines indicate stretches of continuous sequence. The reference, Gac, is the threespine stickleback; Tru, pufferfish; Ola, medaka; Tor, tuna; Ssc, pipefish; Hco, *tiger tail* seahorse; Her, lined seahorse; Bpe, mudskipper; Gmo, cod; Dre, zebrafish; Loc, spotted gar; Mmu, mouse; Hsa, human. Syngnathid CNEs are numbered on the Gulf pipefish. Syngnathid specific losses are in red boxes.

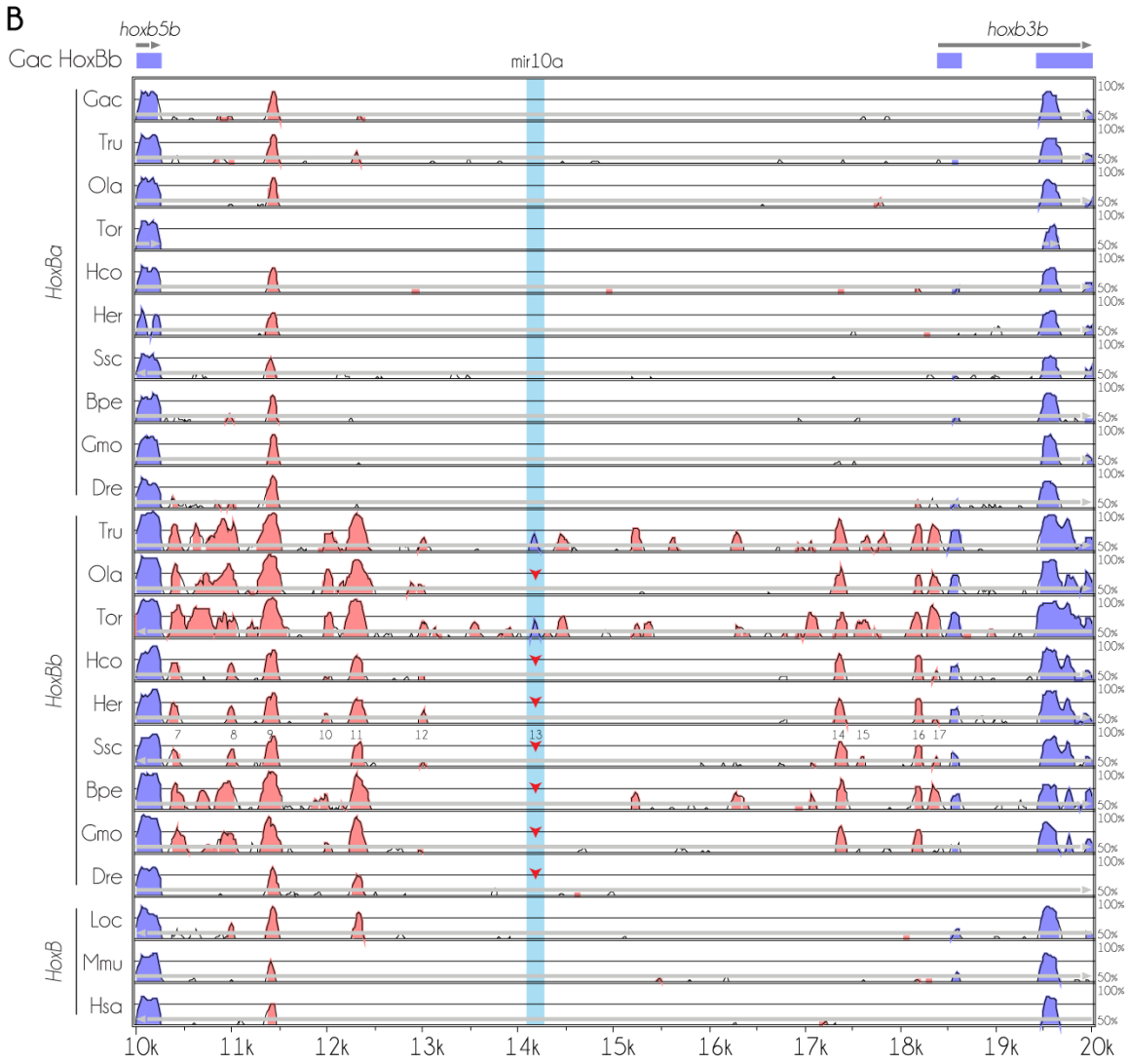


Figure S4.11 continued.

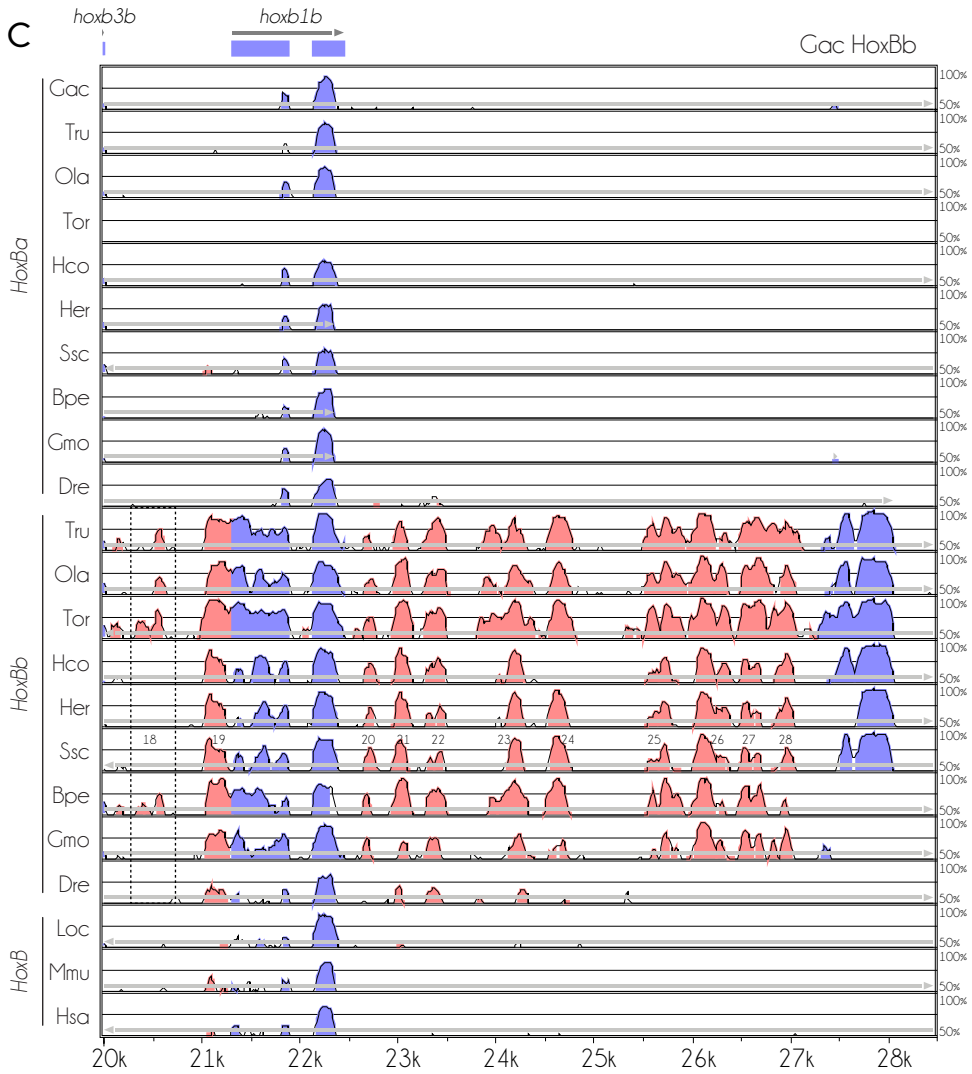


Figure S4.11 continued. CNE18's absence in syngnathid species was sensitive to the species set as the reference and therefore was excluded in the CNE counts.

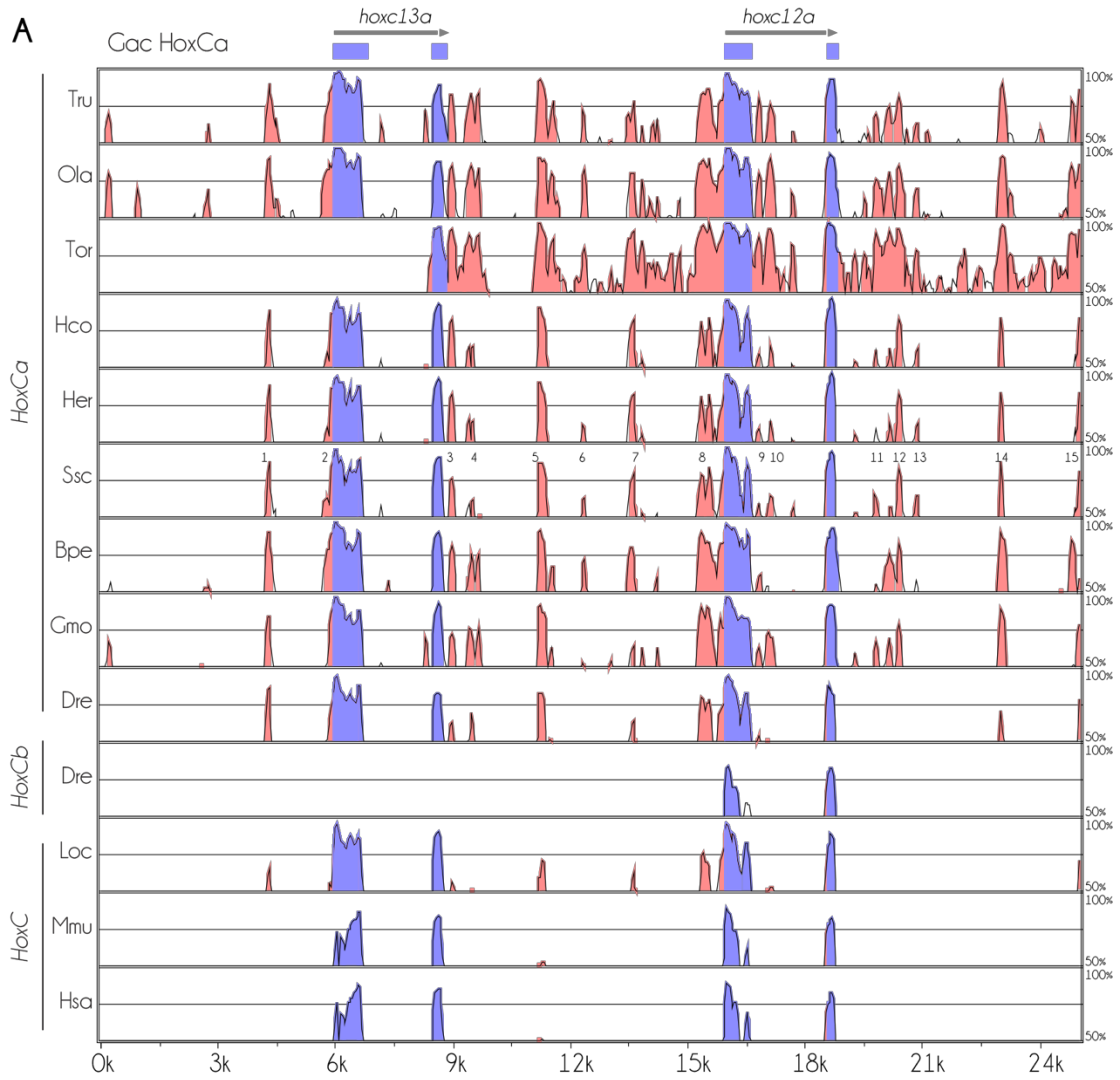


Figure S4.12: VISTA plots for the *HoxC* clusters with threespine stickleback *HoxCa* set as reference sequence. Exons are highlighted in blue, CNEs in pink, UTRs in teal, microRNAs are in the blue boxes. LAGAN alignment was used. The reference, Gac, is the threespine stickleback; Tru, pufferfish; Ola, medaka; Tor, tuna; Ssc, pipefish; Hco, *tiger tail* seahorse; Her, lined seahorse; Bpe, mudskipper; Gmo, cod; Dre, zebrafish; Loc, spotted gar; Mmu, mouse; Hsa, human. Syngnathid CNEs are numbered on the Gulf pipefish. Syngnathid specific losses are in red boxes.

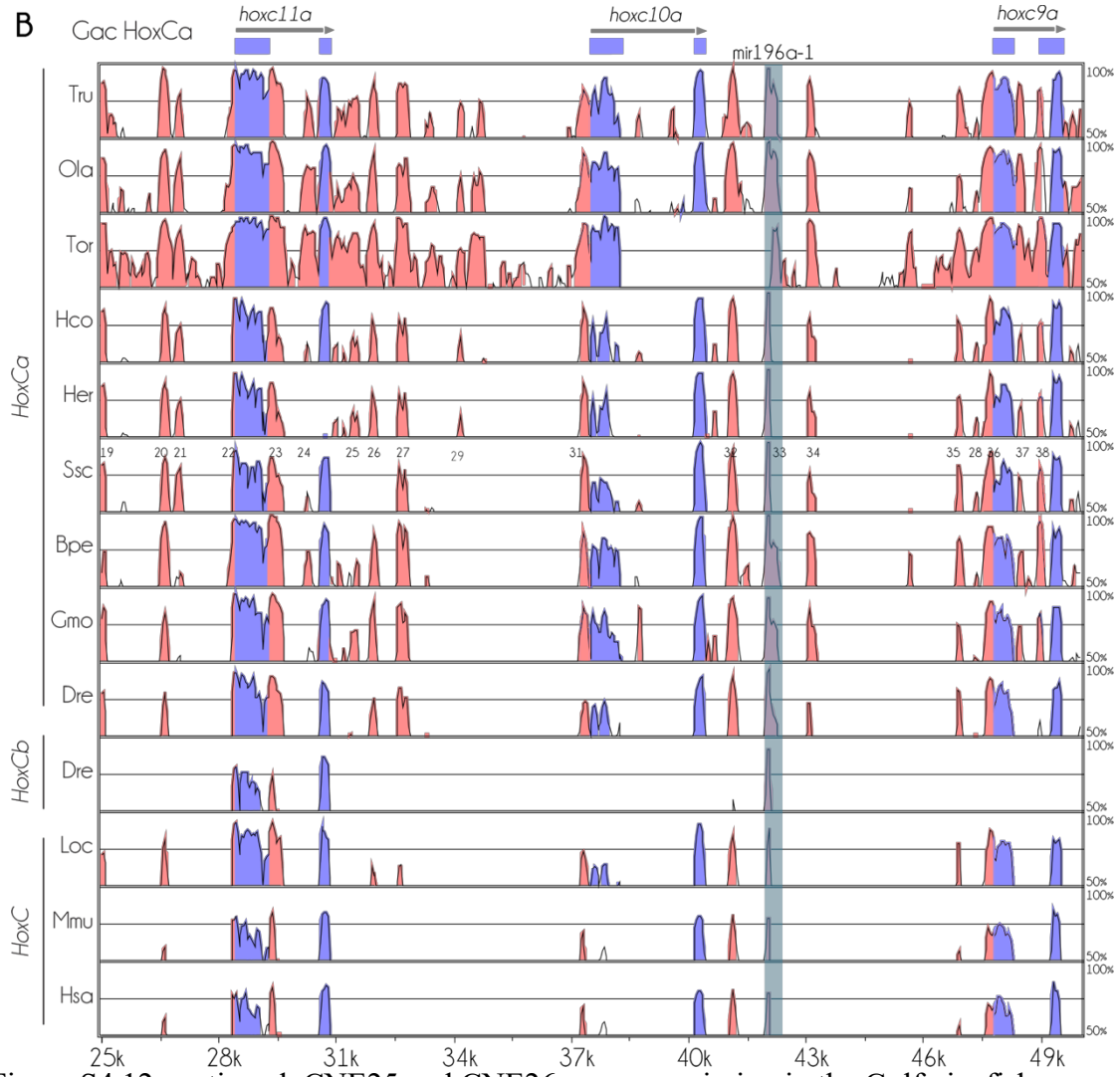


Figure S4.12 continued. CNE25 and CNE26 are gaps missing in the Gulf pipefish genome assembly.

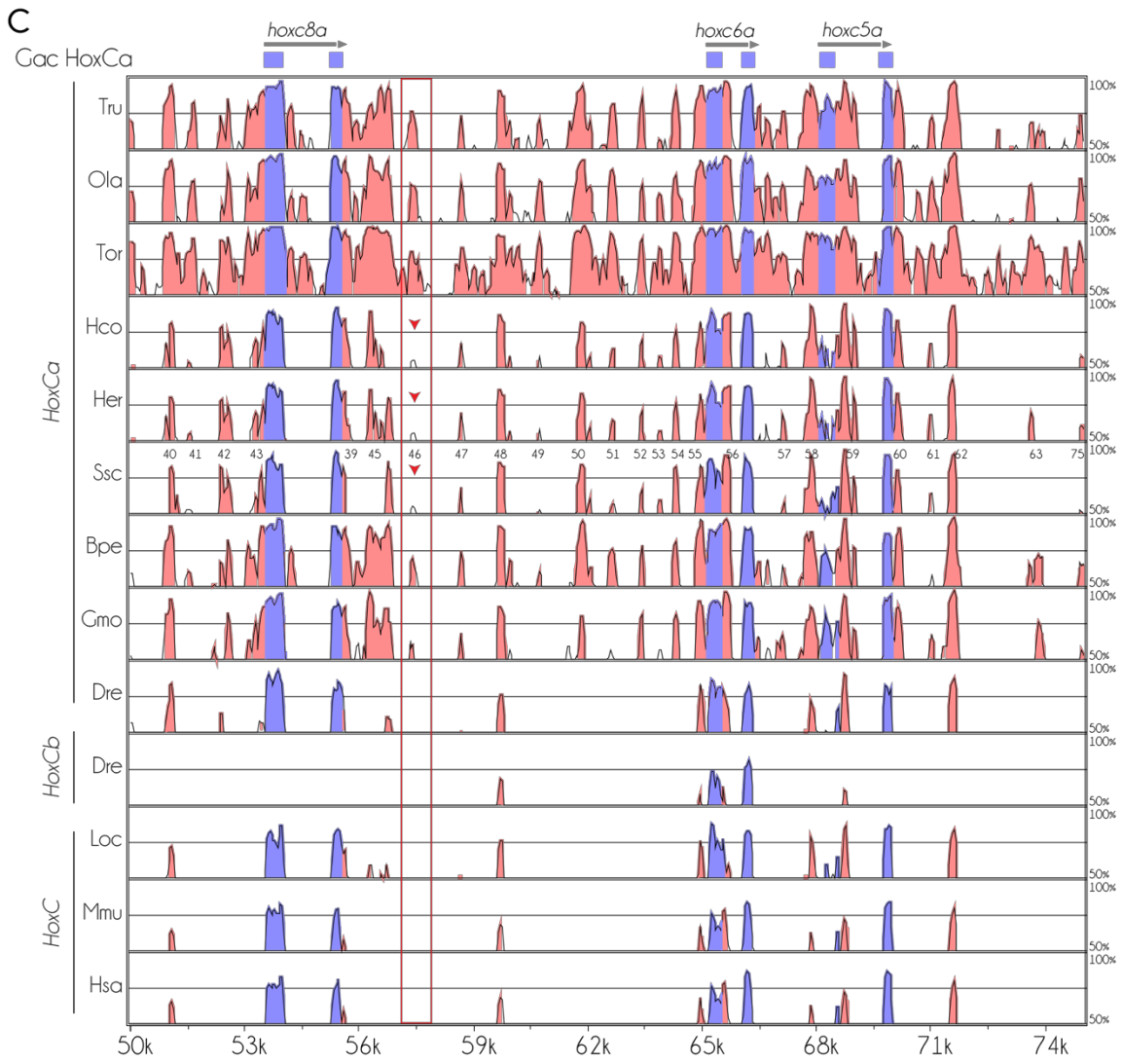


Figure S4.12 continued.

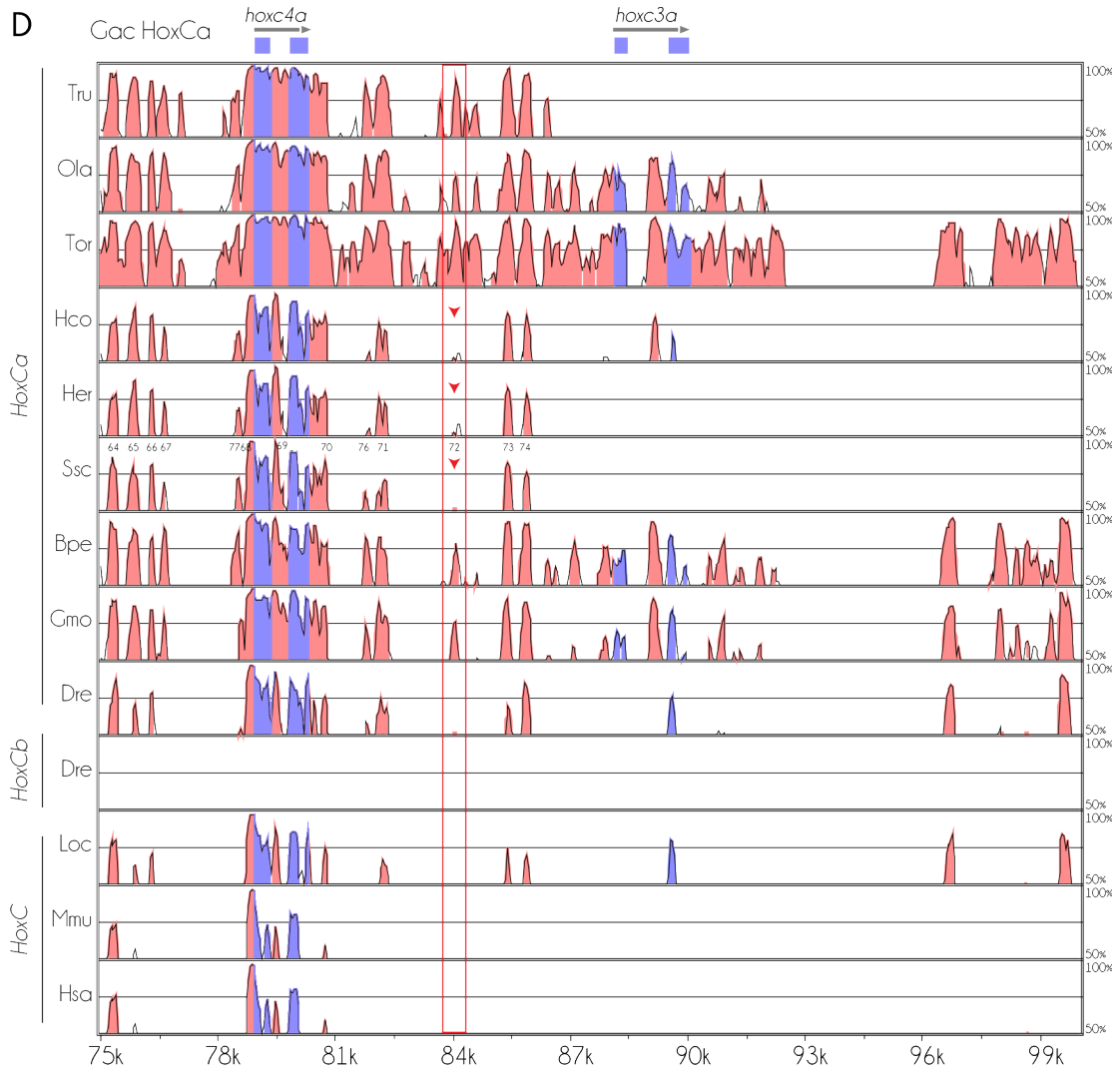


Figure S4.12 continued.

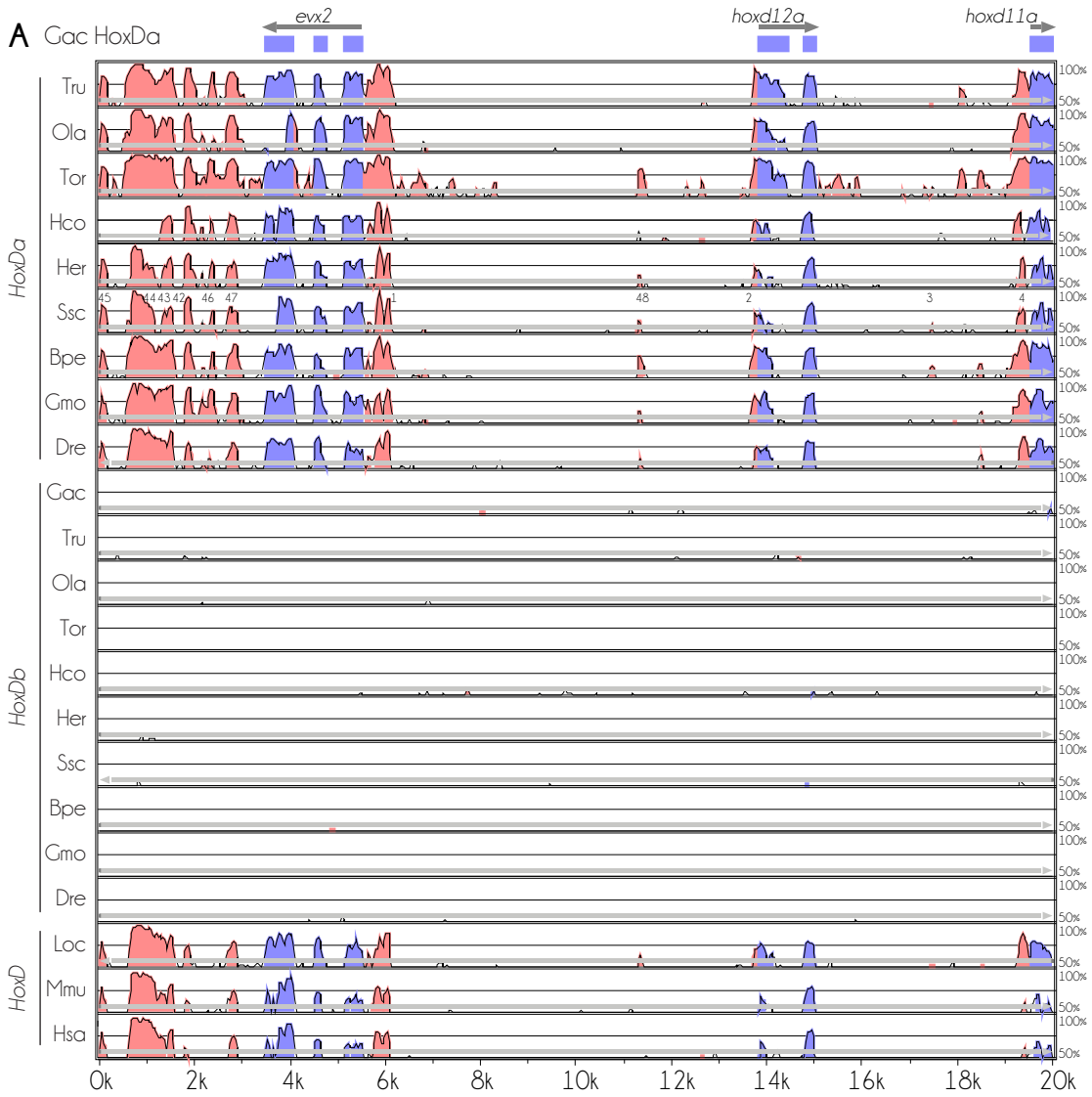


Figure S4.13: VISTA plots for the *HoxD* clusters with threespine stickleback *HoxDa* set as reference sequence. Exons are highlighted in blue, CNEs in pink, UTRs in teal, microRNAs are in the blue boxes. Shuffle LAGAN alignment was used with gray lines indicate stretches of continuous sequence. The reference, Gac, is the threespine stickleback; Tru, pufferfish; Ola, medaka; Tor, tuna; Ssc, pipefish; Hco, *tiger tail* seahorse; Her, lined seahorse; Bpe, mudskipper; Gmo, cod; Dre, zebrafish; Loc, spotted gar; Mmu, mouse; Hsa, human. Syngnathid CNEs are numbered on the Gulf pipefish. Syngnathid specific losses are in red boxes.

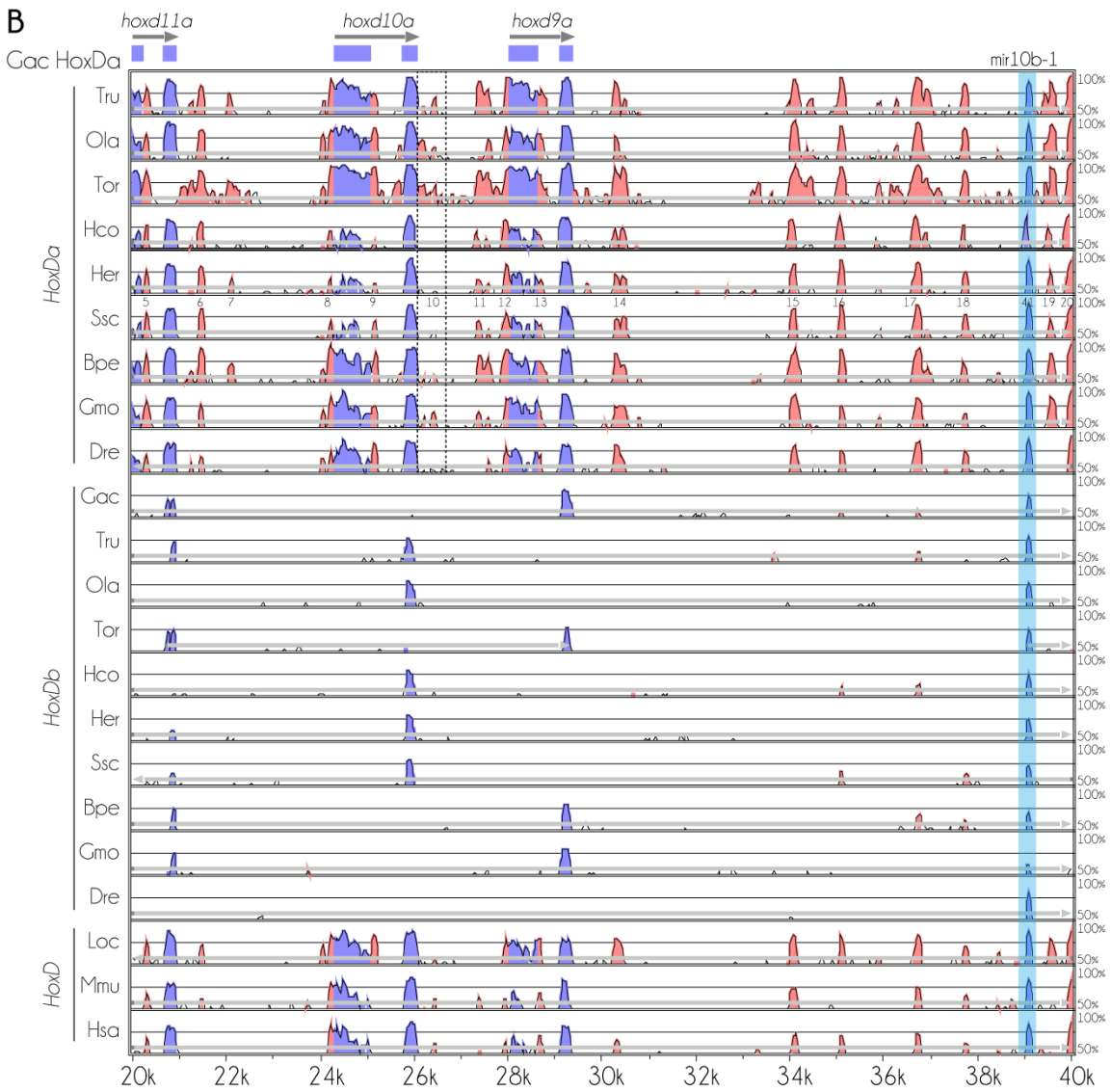


Figure S4.13 continued. CNE10's absence in syngnathid species was sensitive to the species set as the reference and therefore was excluded in the CNE counts.

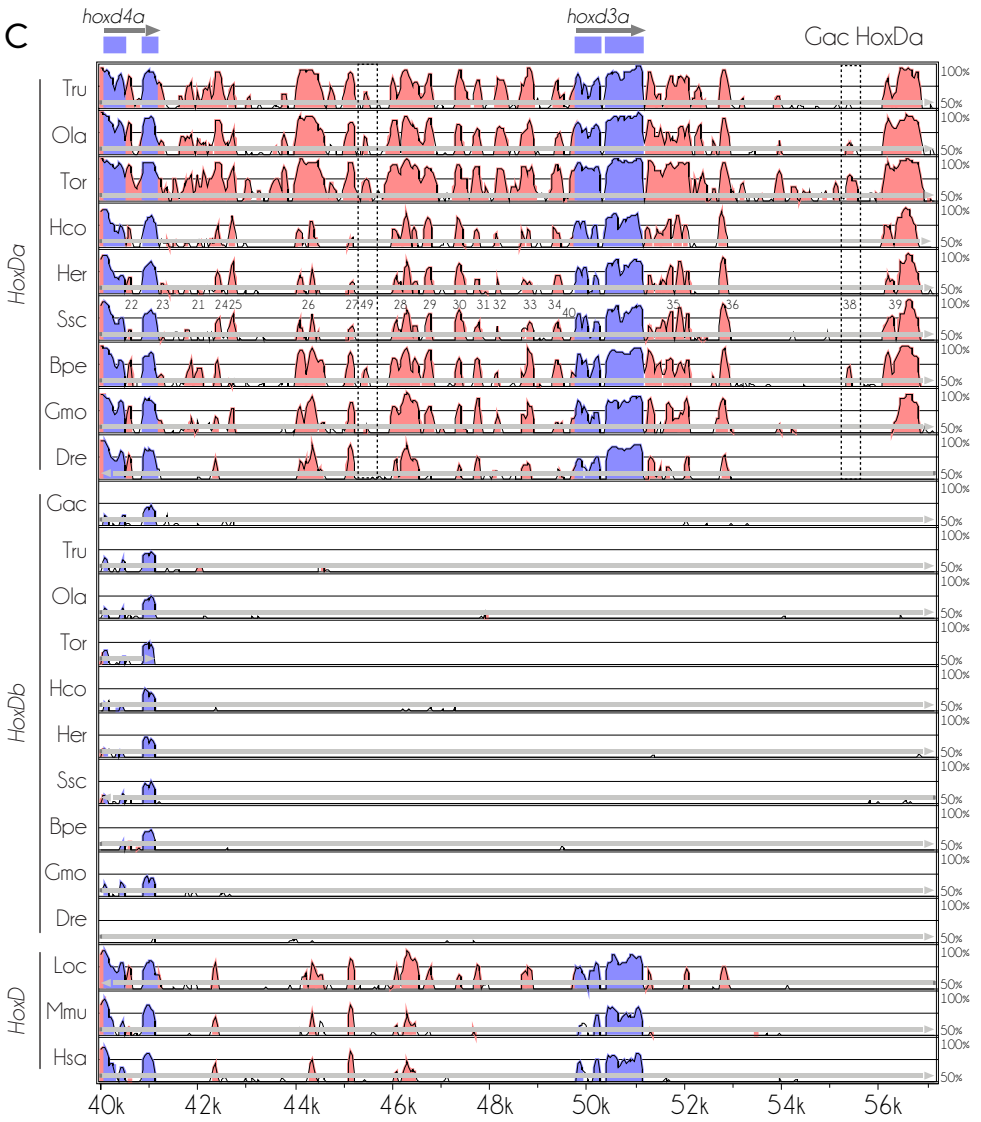


Figure S4.13 continued. CNE49 and CNE38's absences in syngnathid species was sensitive to the species set as the reference and therefore were excluded in the CNE counts.

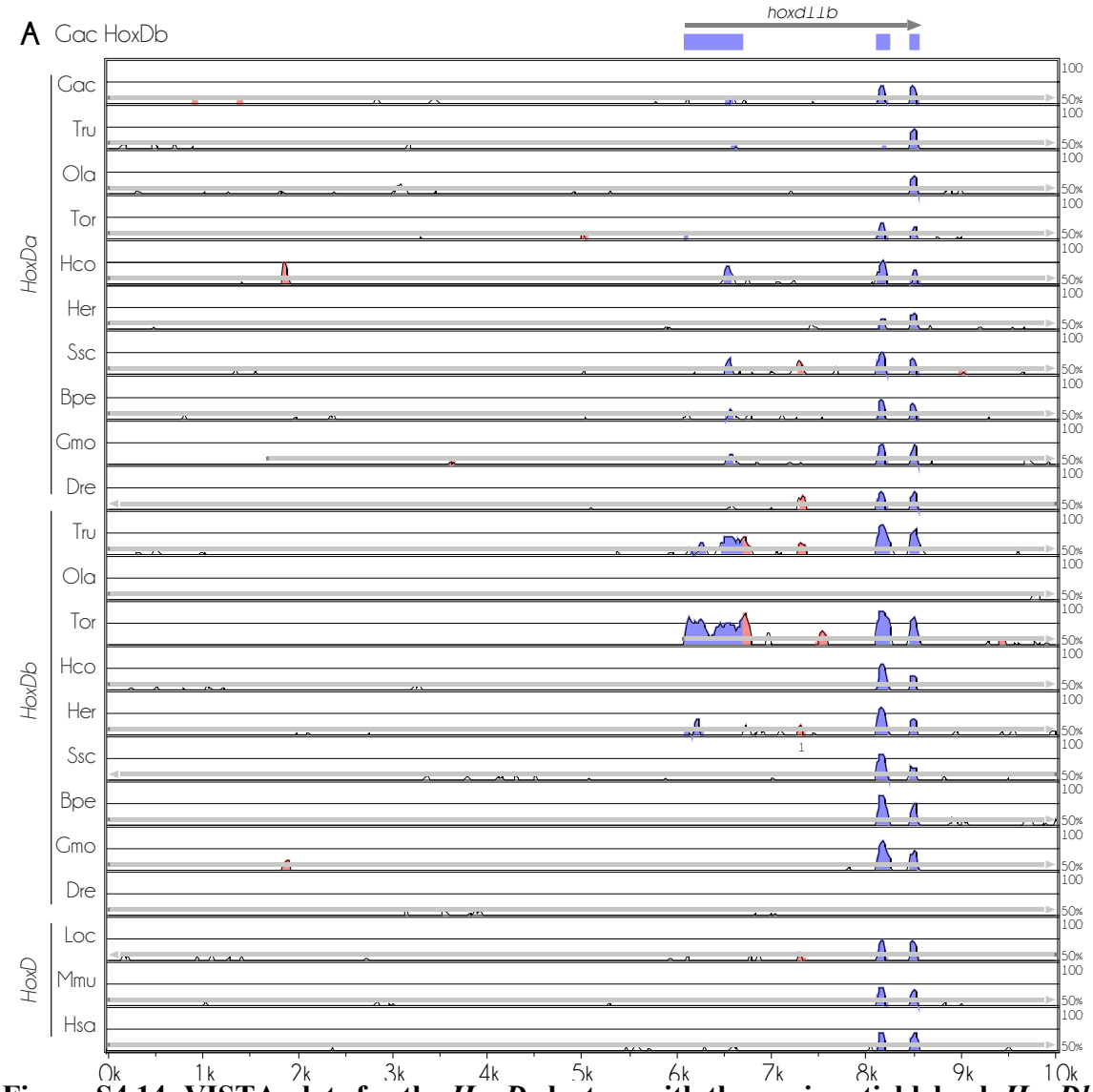


Figure S4.14: VISTA plots for the *HoxD* clusters with threespine stickleback *HoxDb* set as reference sequence. Exons are highlighted in blue, CNEs in pink, UTRs in teal, microRNAs are in the blue boxes. Shuffle LAGAN alignment was used with gray lines indicate stretches of continuous sequence. The reference, Gac, is the threespine stickleback; Tru, pufferfish; Ola, medaka; Tor, tuna; Ssc, pipefish; Hco, *tiger tail* seahorse; Her, lined seahorse; Bpe, mudskipper; Gmo, cod; Dre, zebrafish; Loc, spotted gar; Mmu, mouse; Hsa, human. Syngnathid CNEs are numbered on the Gulf pipefish. Syngnathid specific losses are in red boxes.

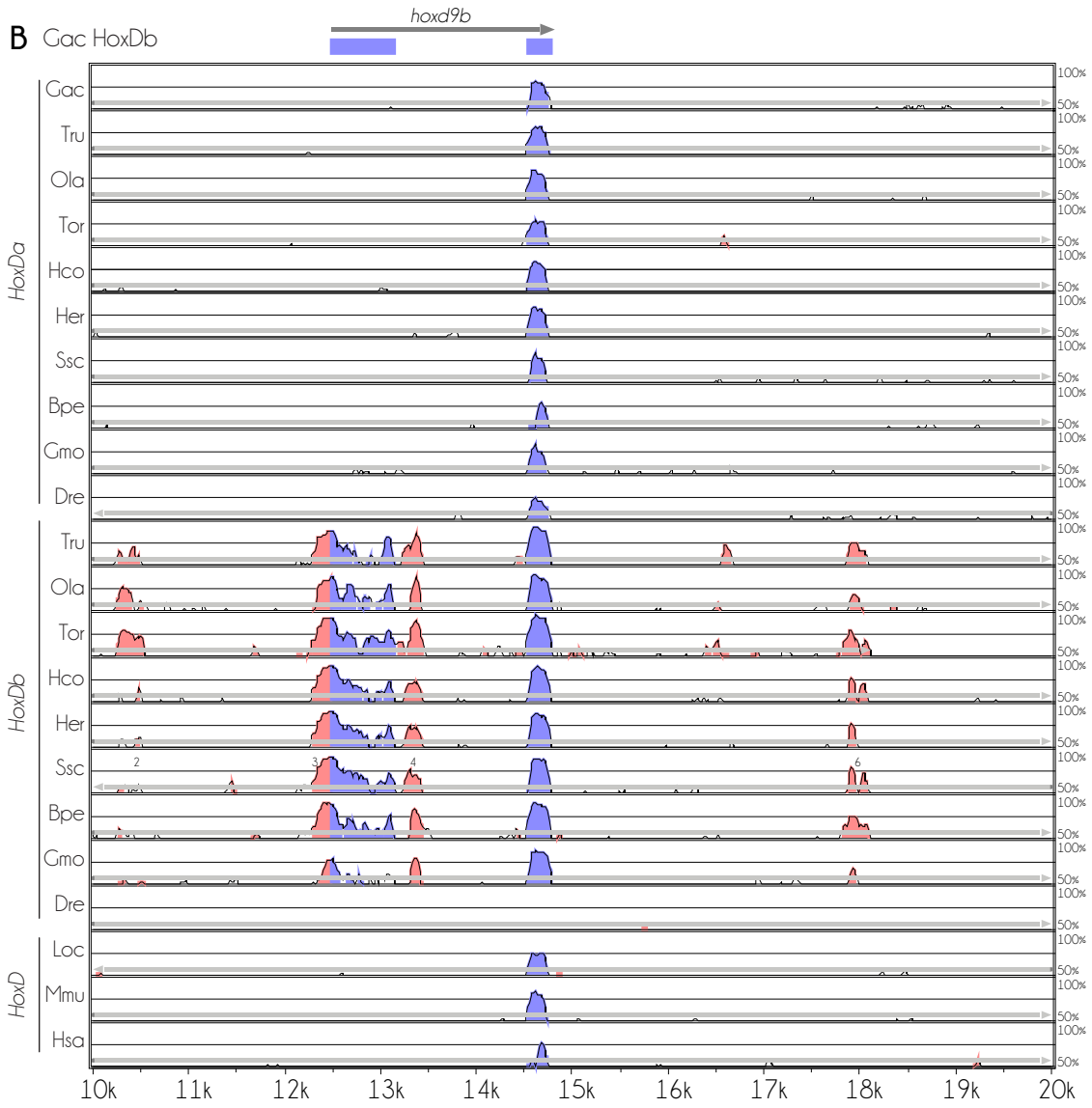


Figure S4.14 continued.

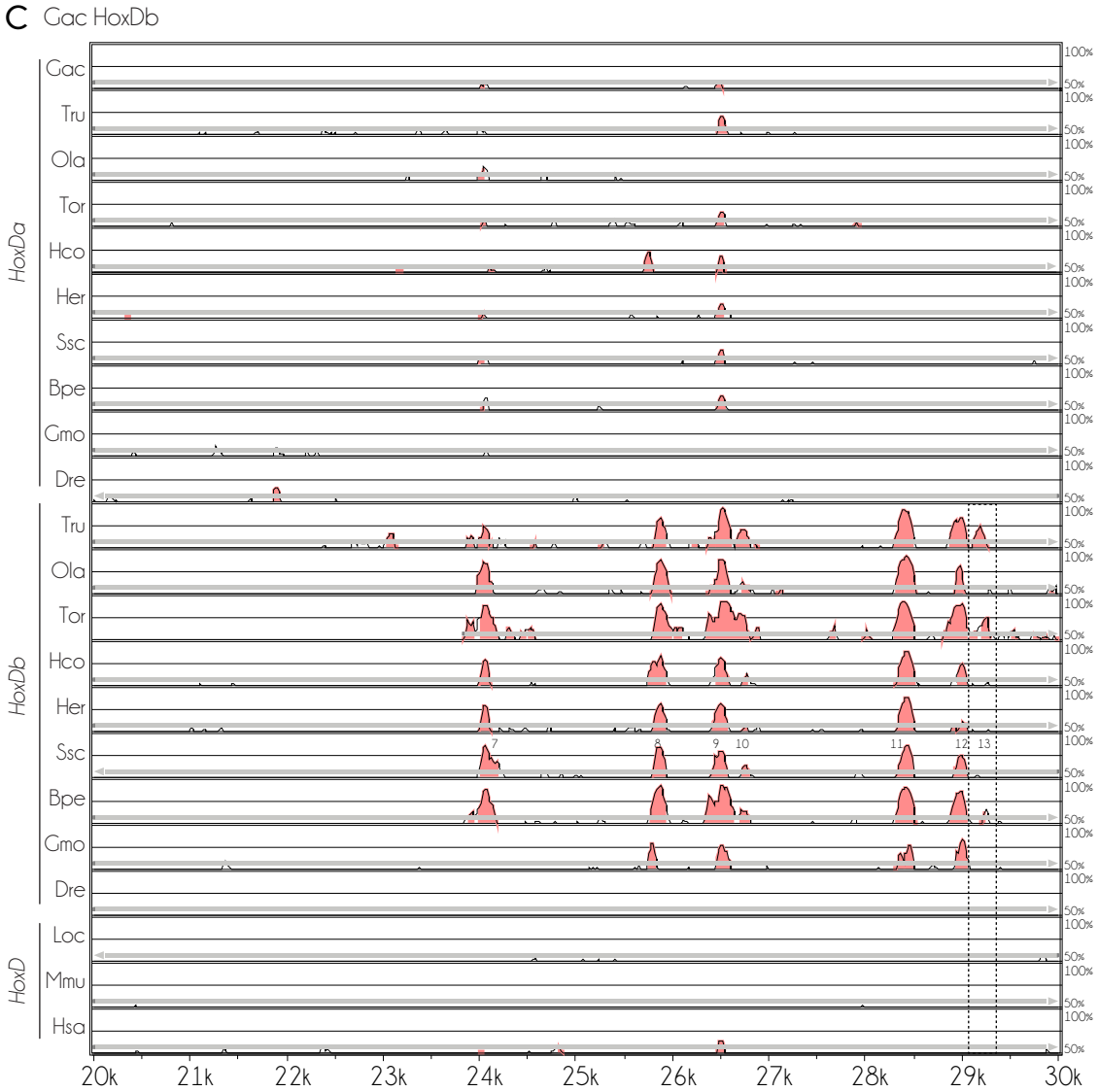


Figure S4.14 continued. CNE13's absence in syngnathid species was sensitive to the species set as the reference and therefore was excluded in the CNE counts.

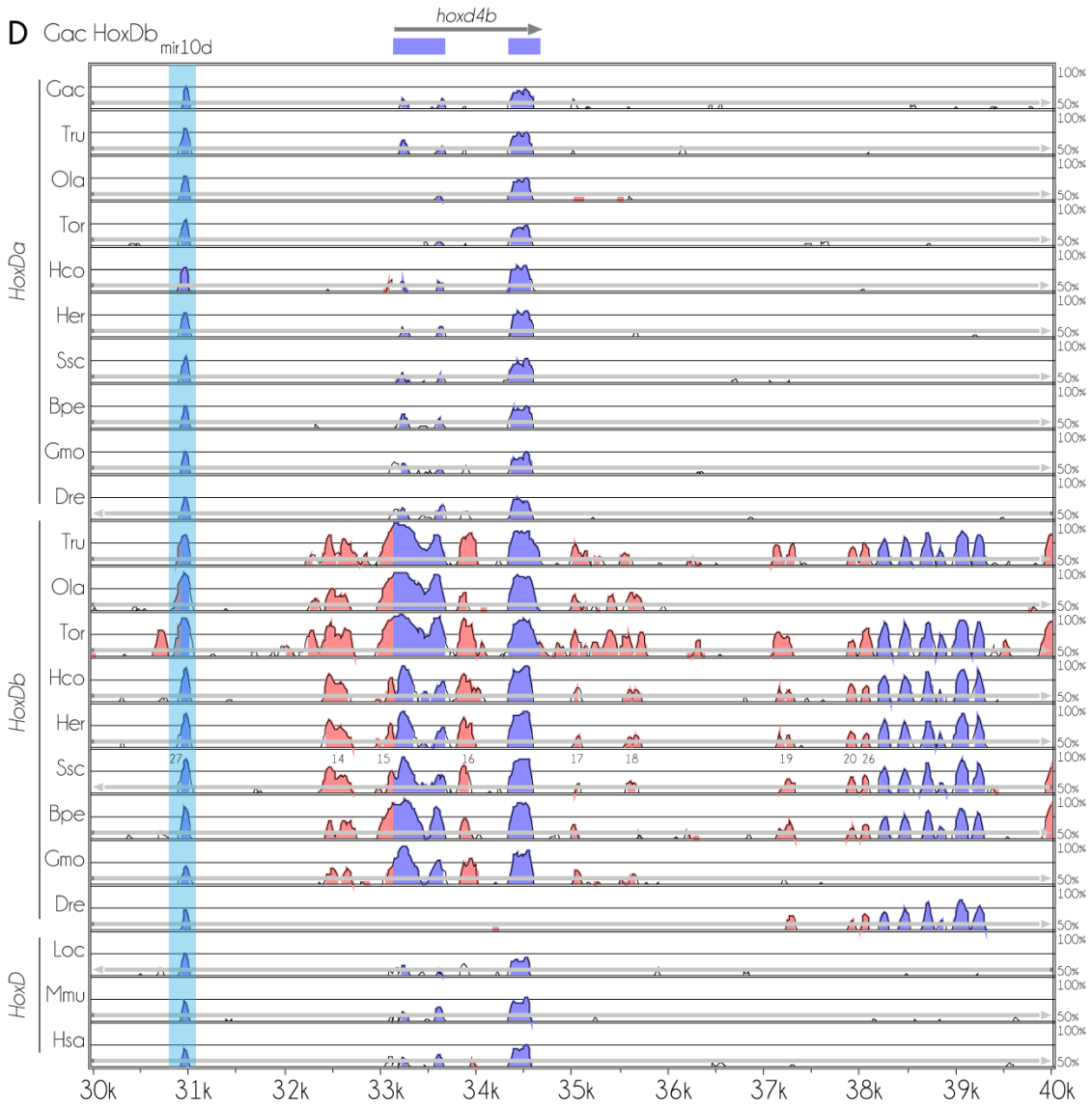


Figure S4.14 continued.

Table S4.1: List of Conserved Noncoding Elements found in the Syngnathid *Hox* clusters. The first column indicates the cluster the CNE is found, the second column indicates the CNE name as labeled in the above VISTA analysis plots with threespine stickleback set as the reference (Figures S4.8–S4.14). Column 3 is the Gulf pipefish sequence that falls within the boundary of that particular CNE peak in the VISTA analysis. Column 4 is the conservation level of the CNE. V, vertebrate; AA, actinopterygiian; T, teleost; AA, acanthomorph; P, percomorph; P2, percomorph without mudskipper.

Cluster	Name	Sequence	Conser. Level
HoxAa	CNE1	CTTAAGTGGAGTCAAGTAAAGTCCTTAAAGGTCAAGTCTCTGGGTCAGTTTA GATAAGA	P2
HoxAa	CNE10	GCCAGCCACGGCGGTCTTAATCCAAGCTCAAGTCTCTCATTCAGACCATATT AGAGTCAGGTTGCACAGGATCCAAAACGAGCT	AA
HoxAa	CNE11	TTAGCCGCTACTTTGAACATGACCCTGATAACCAAATGCGCACTTAGCCAAA TGCTTGCCGTTGTAAAACCACCAAATGCTGATGACTAAGACAATTNCAAACG CAGACTTGGGAGATAAGCTTGACCGTCTCCTTCCTCCAATTCACGT	P
HoxAa	CNE12	AAAATCTGTCAATTTCTTGCCGCGTGGTCACGTGACTCCCGCCTCCGTGGAGT GGATGGCGATGGCTCTCCACGTCAGCTTTACGTCTCCAAATTTCTGTCTCGCG GACCTGCTTGAAAGAGACGAG	V
HoxAa	CNE13	AGCAGCCCGGAGCTGTCTTCCGGCAACAATGAGGAGAAATCCAGCAGCAGT GGCAGCAGCAGTAAGTATTGCATGTATTCAAGTCGCCTCCTTGCTTATGAGA GAAACTTTGCGCGCCGCTGTTAAAATTTTTATATGCTGTTTTATAAGCATATA ATGTTTATAGAGAGCCCCCGTGGACCCCCTCCCCTTCTCTTCCAGCCACAAAA CTTTGGTCAAAACTGCGCCTTGACGT	V
HoxAa	CNE14	TGGTGTGACCAAATAGCGTCCAACGAATAAGAATGTAATTTTTTTTTTCCTTT GTGTGTGTATTTATTTTGTAGCCACTCGTTTACTATCTTTTAAT	P2
HoxAa	CNE15	CTGCAGGCCTAACCACCAGAGACATAAGCTCGCGGGATCCACGGAAGTCGC ACTGANGTACACGCCCAAAAACACAGGTAGAAACATGTAGCAGGTCTCCTG CGTTCT	P2
HoxAa	CNE16	ATTGGCTTTGGAAAGCCTTGTTCCAACAGCTCTTAGCTTTGGGCAGGGCCGC CATATTTGCTGTTGTTATCTCTGTGTTTATGGCAGCTAGCCGGAGCGCATTGG CGTTCATATATGACAGTCAAAAGGGTCATAAACGCCGGTGAGAGCGAGCCT GAAGTTTATGGTTGTCCCTCTCTGCTGTAGTCCGGCCTTGCCTTGCGCGCATC CCTTACTGCGCGCTGGAATTCTGTCCCTAATTACGGGACATCCTCCCCGTTGC CTGGGCAACGCAGTCGTAAAAGCTGTCTGAGGGTCTGGAGCATTTGTACAAT TGGAGTGCACTGCAATAAACCGTCTGAGACCCAAGGTTATTAACTGTGTTAC CAAGGAGGCGAGAGCCACGACAGGAAGAGGGAAGATAGGGGGGAAAAGCA CTTCTTGTTCCCTGCCTGGACTTCATTTTTGG	V
HoxAa	CNE17	CACAGGGGGGCTTAAAAGGCAACACATGTGCGCCATTGCTGAATACATGTGC TGTTCATTTTGCTTTCGCTGAGCTGGAGCTTTTCCAAAGGT	AA
HoxAa	CNE18	GCGTGTGGACCCGCTCTCCTTCCCCCCTTTTTTTGCTCTCTGGTGAAAAATAA ATACTTCCGCTGGATGTAATAAAACGCCGCATTATTGTTCATCGCTCCTCTCT GGACGAACATGATCCGACCCTGAGTGCCTCTTCGGCCACCATTATATGGGTG TGATCAAAGCCAGTGGTCAGAGTAAACTGTTGAGAGGCTAAAATGGCTGGG CAGCTGTGAAAATGAAGATTTAATAGCTACAATTAATAACTGCGGCAGAAA GATATCGATAGCCCACCA	AT
HoxAa	CNE19	AATATGAGTACACATGCATAGCCAAGGACTCTGGAGGTTGGAAGCAAAGTTT TATTGTGGCCCTTATGGCCCTGAGTTGGCGCAACAAGATGA	AA
HoxAa	CNE2	GATGAGATCTTGTGATCCTTCAGAGATAACGCTCGAGAGCCTTTTGCGCAAC CGTCATGACCTTGCATGCTTTACAGTAGTGTCACTTATATTGTGTGTGCT	AA
HoxAa	CNE20	GTGGATCCACCTCAAGCCCTAATGGTGCATTAAGGCAAGAAGGCGCACAGT GATGGCTGCGCTCTCTTCGTCCTCAGCTGCTTGATCGATGTTGTACGTAGGCT CCGAGTGGGCGGGACTTTGGCCCGATGCGAGCTTCCCATTGGCCGCNCGGCG CACACGCTAGAATGAGGGGCGCTCCTAATCATATCCAGCATGTTTTGCACAA GAAATGTCAGACCGAGGGGGCTACCTTCTCCCTTCGCCAAAATACTCGACAA TA	V
HoxAa	CNE21	ATGTTTTTGACTTGTATATTTTTTTTGGTGTACATACCGACGTTGTGTTTTGTG TCATGCTGAATAAATTATCGCTCTGGCGCATTATAGTAAATTGTATTTTTGCT CCGTGTGTGTATACGTGTGTTTT	AA
HoxAa	CNE22	ATCGGAAATAATTAAGGAATGGGGCTCTAAGGCGTCTTAGCTGCATGCAGGC GGCTATTGTGGAGTTTTATGACTGGCTGCTATATGAGGTTTCTCCCTCTCCTC TAAGAGNTGCTTCAGTCGTGCATGGGAGAGGCTTTCAAACGGACTGGTGGCC	V
HoxAa	CNE23	CGAGTGGTTTAGGTAGTTTCATGTTGTTGGGGTGCATTTCTAACTCTGCAACA TGAAACTGTCTTAATTGCCCCA	V
HoxAa	CNE24	TGACGCAGACGGGCCAGACACAAGTCCTCGGTTACGACCACTGCAGTGGAG GGGAAGGGGAAGGTGAATGGGCCAACAGCGACCTCCGTGGGCAATGAGACC	AT

		GGATGCTTCCGCTGCGCAATAGTGGCTCAGACAAAAGGAGGGGAAGGGGCT TTATGGCGAGGTTTAGATCCGCCAGACAGGCTGTTATCTACCGAAGCAAACG AGCTCAATAGCCCCCGTGGTAAAGTTGGTGGAAAAAACACCCTGATTTAACT TTACAGCCCGAGAGTCATCGCAATAGTGGCGTGTTTTCCTCATCCGAGGCGT TTGT	
HoxAa	CNE25	ACATTGTGTTTTGGTTAGCAGGGTGCTTACGTGGCCTATAAAATAGCTCATA AAATGAGCTTCGAAAATACCGTGTCGTTTGCAGCAAGGCTTTTTAACATGCA GCTGCAACTAACGGACTTTGCCCGTCCTAGTATTTTCATTCATATTTTCTCGC CTCTATCCGATTGTATCCATTTATTAGTATTTTCTCAACTATGCTTGTTATTGG GGACTGTGTGCGTATTTAATGAGCAGAGGGGATTGATTTATGCTTTATTGTTC AGGCTGTATGATCACGTGCGCGGAGCGTCCAATAGCCTTGCGGCGTGGCTCC CACATTACAACCCACTGTAGTTCTGTGAGGGGCCAAGTTGCTACTTGATTTCT CCACATTGTTATTT	V
HoxAa	CNE26	TCTGCTGGAGCTTCTGGAAAAACTCGTTTTAGTGTTCCATATATCAGAACCTA TTGGGCTTAATCGTCTTGTACAAAAAAAGCATGTTAGGCTCTCGTTTGAAGC AACTACCTTTTTTTATATTGCACAATTTTAACAGTGCCCCCTTTTTGTTAGTTG TCCTTTTCATATGGTGTTTCGTTTTTTTTAATTTTTATTATTAGTACCTCGTTAA AGTGAATTAATTGGCTTGTGCTTGTTTGGATGTTTCCAGCAATGTGTTCTCTG TTGTCTATTTCATAGTCACCTGAGCTAGTTTGTAAAATGTGTTGGATTTTGTT GTTAATTGTTT	AA
HoxAa	CNE27	AAGTATTAGAATAACACAGAAGCTCTATGTTACATGACTGACTACCTATATG TGTTTATCAACTACCTACAATAACATATTTTTCCTCTCGTGGCAAATTATTGT AGAATTTATTGATTGGTTAGACAAGCCCGAATTTAAGGTTTTGTTTTGTTTTC GGGAAAGCATTGTTTAGTGAGAAATCAAATAACACACTGTAGGTTGAAATTG ATTACCTCTCGGTCAGCCGACCGTCTGAATTTTGGTTCTCAAATGTAATTTGT ACTTTATTTATTAAAATAAAACAAATTAT	AA
HoxAa	CNE28	AGGATGGTGATAGCTCTATAGAGCCGTAAAAGACAATTACCGCTATAACATT TTATGAGGTGCAAAGCGCTGCGAGGCAAGCGGACACAAAACAAAAAAAAAA AGACGTGCAGGGCTTGGGCACATGAA	V
HoxAa	CNE29	AATTTGAAGCTTGGAGACTTGTGTTGTGACGCGCTTGGGGGGCGACACACAG ATGGACCCAAACTTCAAAGACTCGCCAAGAGACAGCGCAATAAACCGCCTG GTCTCCTATACTGTCTGGCATTCCACTTTTAATGGCTTTATGGCCGTCCAGAC ACAATTAGGCTGTTTCCAGAATGGCACCCATTTGTTTTTTCTTCTCTTTGGTTC TGGACAAAAGGCCGAGCGGAAATGATCAGGTTTATTGGACTCTTCCCAACGG GGACGCGCATGCACGACTCACGGTCATTTGGACGCTGCCTTTGTTTCTACCTG GGAACCTTCCGCCCTCTGCTCC	V
HoxAa	CNE3	TGAGGCAGCATGCAGGCAGCCCAACTTGAACTTGACCATGGAGTCTGCG	AA
HoxAa	CNE30	CTCTAGTAACATTTTTAAGAGTCCNGCCAAACAATTTCACACCTTTTTAACAG TCTCTG	P2
HoxAa	CNE31	GTTAATCCCACTCTGCAGCCTGCACCATGATGAATTTCATATCTGCATCCCGA TTGGAGACTACCGGCAGGCCGAAGTCACGTGAGTGCGTTCCTCGGTGCACGG TCAAAATGGTGTTTGGTGTAAATCTAGGCCGTATTGCTGTCATATATCACACA ACCTCGTAAA	V
HoxAa	CNE32	GGATACTTAGCCTAACCATTGGTGATTCAATTAATTTCGTGAAGCCACTGGT GTTGCAAGGGAGGAATCGTAAAAGTTTTATTGCGCAACTAATGAGTTTTGCA GNCATAAATTCATTTGGTCAGCGCCTCTGTGTTCAGCACCAATAGACCTTGCC GTTGGGGGGTTTTCTGTGCACTCCTATTACTCTCATAGCCTCTTATCCCCTTTG TGGTTGTAAAAAAAACTGACCTTTTGTGTACATCGCTCTCCTGCACTCTGAAG CTTTACTTTCTAATTGGGCCT	V
HoxAa	CNE33	TTTGGGAATGCAAACTATGAAGATGCGAGACGAGGCAAGCTGAGGGAGAGA ATTTACTGTTGAACTCTGCGTTCCCTCTTCCCTCACTAAGATGGCGCCTGCTG TTGACCTCCTCCTTTGTACTCCCTCCACCTCCGTCTCCCCCCCCTCTCTCCTCCC TCCCTTCTTCCTCCAGCACTGGGCCATAAATCCGTTGTTGTTTATGAAAATTT ACAACATAGCAATCCGGCTTTACGAGCCACCTCGGGCTCCCATTGGCTGCAT CTCGTCACGTGNTCAGGCGCAGTGAACATGAACTTTTT	V
HoxAa	CNE34	TTTTTTTCCCCCTCTCCATGGCCTCTTTTTAGCTCCTGGGTGTGTCCTTGTCTC GACTTGGGGGAAATGGTTTCGTGGGAACTTAGCTGCCATTGGCCTGTCATAA ATCAGCTGTTG	P
HoxAa	CNE35	GCAGATATTGCATTCCCAGCACGCCTGCTCGTCTTGTATTGTTTCCAAGCAGC GCGTCCGCGTTTTTTTGCCGCAATGCTCGTTAACCTCCCAATGACCCCGCTCG CAGATTCAGTGCAACAGTTGCGCCAGAGGATGGCGCACCACGACAACGCAA AGTTCGCCCTGAGCTAGCGAGTCCTCCTCATCCCCACCCCGCGTGAGTTTACC TCCGGAGGTCACCAAGCAGGATTTACGACTGGTCAACAAAAGCACGTGATA CTCCGCCGTACCCCATATTTGGGTGCCTACGTAAGAGAGAATCAAGTCCATG GCCCACTCATTTCCATAATTCATCATAAATTGTGCAAGG	V
HoxAa	CNE36	TGTACATTTCTGCTTAGATTTAGAGAAAACGGTGAAATAAATGTTGGTTTAA ATTATTTCTTGTTCAGAAAAAA	AA
HoxAa	CNE37	CCGCGCGTATACTCCGTGGTGTCAAACTTTTTTTTGTAAATGTTTTTTAAAAG ACTGGAAGTCAAGAATGTACTGTAAAGACATTGTTGCTGTTCATCATGACAA TTATAATAAACTTTTTACT	AA
HoxAa	CNE38	CAGCTCGATGTTTTCAAGCTGTAAACTTTATTAGGCCCCTTTCAGGCTCTTCG ACATTTGGGTGCTAAATGAATGGGGGGTTTTGTCTATGAATTAGATCGTAAA	V

		AATCATCCGGAGCGCGTCCAGATAGGCTCACTGGCCATAAACGGTCACGTGG TGGCCATTAAAGTAAGTTTTATGGTTTTGGGGAGTTGACAGTATATTGCACAT AACATATAATCGCACTGACGACGAGGCTTGGTCTGACTCTGCCTTTTGCAGC CCTTTGAGGAGTGTCTTGACCAGAACACTGCGCCCACTTCAGCTCAGTGGAC CGGACGGGAGCCGACCGGACCTGATCGGACCTGCTGTGNATGCAAGCATGG CCGGGATGAGCAGCTCCTCTCCGGGTGAGTTTGTGCACATCGAACGGTAGCN GGAGAACAAGCGAGCCTGTGTTAAATAGCCTTATTATTGTGCGCTCTGCCTC GATCTTTCCCAATTAGTCCATTACGCTAATTTGTCAGCGATGCTGGGCTGAAT TCTGGTCTGGGCAGATGTTGCGAGCAGGCAGGGTTGGGTCCGGCTCATGGTC ATTTAATTGCTTCCTTCCCCGTGTAACTTGTTTCCCTCCATCTGGGGTGCGACT CATAATCCATTTTTGCATGTTGGATAATTATCTTTGGGTCCATTTCCACCCAC TAACTTCT	
HoxAa	CNE39	GGAATTCCGTTTTGTGTCAGCCACGTCATTTTGTGTTCGGGTTGAAAAGTTCA	P
HoxAa	CNE4	CCTCGCTTCAGTTTGGCCTTGAAACGGGATGACTCATGCAGATTTAACTCGTT GACCTTGGAGCCAAACTCGTCCCGGTGCTGGAAAGCAATTGTCTCATCCGAG CTGCTCCAACGTTGTTGTGCGTGCTTTTAACACAGTTTGGAAGTAGGAATTGG GGTTGTGGAATAAAAATGGGCCAAGATCAGTAGCTGGGTGACATACCAGGT GCAAAGCTTCTACGTTTCACAAGAGTCCAGCTTCTTTAGACTGGCATCTTTTG AGAAAGCT	AA
HoxAa	CNE40	CAGCTCTAAGATAAATCTGCATCCTCTCCGAGCCACCAGCAGAGCTCGCTTT AGGCCAAGTTCACTGTCAGCCTGCAAATGTGTGCAGA	V
HoxAa	CNE41	TCCAATTTTTTTTTTTGGGGTGCGGGGGGGTGCCTGAATGGGCCACCTCTGAC GAGAGCGAGATAGCAGGAGAGAGAAGAGGAAAAATTGGGGTCAAAAGTTG AGCTGCAGCGAGTCTCCGTCATCAGCTTGATGCCTGCAGCCTCATAATAGCG ATCTTGACTCGTCCACACAAAAGGCAGAATAGCTTTGAATTACATATGTTGC GGGGTGCACTCCAGGTGAACCCTGTATGCACGGTGACCCCGCGTTCGGGAGG TGGGGAGGGGGGACCCCAACCCTCCCCCAGCCGTCCCCGGCAACAAGATTG AGTGGCTGGCGTTTTATTATCGCGCTTGATTGGTGAATTTCCTTTGAATAAAT TGCATTTGATATGTTTGGAGACGGGGGTGCGCCATTTGCATTTTCAGGATTGA TACACTCGACCCGGTCACCGCGTAATGGGAAGACCACCCCCCCCCTCCCAAA CTGAGCCCCCGGTCCCCTTAAAAATCATTTAGCCTCTCGCATCCTGATCGTGT TGAATAGAACGCTTTTAGGGGAGCAGAAATGGAACAGATGTTTT	V
HoxAa	CNE42	AGACAAGAGACCTTGNGTGAGCCACCTGCAATTTGTCACGATCGCCACAGCA CTTCACTCATAATCTAAATTTACAGTAAAGTGTCATCATGATGTATCTTAGCT CAGCAGACCTTCACTTGNCGGTTTATGTTCGCTGAGTTGAGACAGCACG	AA
HoxAa	CNE43	GAAGTCCAAGTCCGGGGTGAAGAGCGGGTCAGCGCGTCTAACTGATATGAA AATGTCGCCCCTTTGGAAGAAGGGGGCACGCCTTGTTGTTTAACAAAGACTG TCAATGGGCAAGATTAATCAGAAACAAAATGGAAACTGTGTCAGTTTGGGGT CAGGCACAAGTTATCTGTGTTTGGGGGGGAGTGAA	V
HoxAa	CNE44	TTTTCCACGCGTAGCCTCGCCCACATAAGTTTGACGCCGTGCAGTCAACAAG AATCTTGATTGGTTTTTCTTTTCTATAATGAGATTTAACAAGGAAAAGGGCTA CGTAAGCCAGTCCGGAGAGGTTCTGCTGGTGCCAGTGCCATGCTTTCCATTG GTTCCTGTTTACATGATGCCCGCAGGACGCGCGCTGATTGGTGGCNCCGCTC ACGTGACCACGCAACTTTGTACATTTGACAGCGAGTAGGA	AT
HoxAa	CNE45	AGGCAGCGCTGGTAGCAGGCAGCCACTTTGCAGCTGCAATTTACGACTATGG AGCAGCGGGGTCGGTAATTAAGGACCCTCGTAAATTTTATTGCCTGTGTTCG TCTGTCGGGGCTCATGCGCCTTTGTTGCTTTTTAACCCAAACTACCTCCTCCA CTGGAATGAAGCCAAAGTGTAATATCCGATTCCCGCAGTTTACTTAATCTTTC TTAAGCGTCAGCTGTCCC	V
HoxAa	CNE46	AACAAGACTGAGATGGCAAAGAACACAAGTGGAAATAGAAGGAAGGACCC AGTGAATATTTTGGTGTATTTTGCCCTTCGTCCCCTTCCACTCTGCGGTCGTCT TTCGCTTATTCACATAAGCTTGATCTTTTTTTTTTTTTAATCCCACAACGGCGAC TGAAGGGTTCATATATTTGCTTTTTTTTTTTTTTTGTATTTTTATTTTTTTACGAA TGAAGGATAACGCACAGTCACATGGATCAAAAAAAAAAAAAAACGTGTTGCAT TTCAGTTAGAAATGACCGGATGGGTGAGGCTTGAATCCTTTCCTTATGGATG GTTTATTTTTATTTTATTTTTTATTTTTCACGAAGAAATATTAATATACTTGAA CGTTATGATTTAAAAAAGCATTTCTTTAGGAGTTCCATTGTAGTCACGGTGTG TGTTCGAGTGCTGAATGTATAGTCCAGGTGACCTGCCCATATTGTAATGTAC ATATATAGACAAATCATGATTGAAAGGCACAACGAGGATTGCATTTCTACAG GTAGAGACTTTTTGGTTGTTAGTTCGTTCGTGCGCTCTTCAATATTTGCTTCGT GCTCGTGCGTCTCACGTGTTGTTCTAAACTGTGAATATACATTCGTGGTTGTG CCGTCTCACTAGTGACCCCTGTGCTGTAGTGTAGACTGAATATGTCCTTGTGA AATAAATGTTCACTTGTAATGCTGGGGATGGCTTTTGATTGTTATTC	AA
HoxAa	CNE47	TCAGTGCACCCGAGTGTGAGTGTTGGTCACTACCTGCTACATCCAAACAATG ACGGGTTCAACTAAGGGACACGGCCGATTCTGACCAATCGCCGGNCCACAG CCCTTCTTATGACCCCTGACCTCTCGACACATACAATGTTCAGACCATACATC AGCGTTGAGACCCCCGAGAGTCTTTTCTTTGCGTTTTAGATTGACCCCGAGA	AA
HoxAa	CNE48	AATTACAATCTTCAAGTCATTTGCCAATAGAGGTCACTTCCAAAGTGCTGAA ATTTTAATTTGCCCATTCGCATGGCTCGTTGAATAATGAATGGGCCGGGGTTA GAAATAGATGTGAAAATGTGCCTTATTTTCATTTCATGTGATCTCTTCAGGGC GTGCGTCA	V

HoxAa	CNE49	AGCTGCCTGTTAAGAGCCGCATGTCAACATTGCGCAAAGATATGCTTCTTGT TCTTCTGGCCATCGTCCAAGG	AA
HoxAa	CNE5	CCCTTCCCTTCTTTGCAGAGTTCCCACCATTGTCACACGTATTTGCCAAAAAA CACCTCTAAAACATCTCTAAATAATAAACACAAAATTCTTCNAATCCTACCA CAGCGTATAGAAAAGTACACTTGGCTCAAATGCCGCTGAGACATTGGACGA GTAGTTGAGAAGGAGCGCTAGCTGGTGTAGAAAACAAAATGATAGTAAAAG TTAATGTCCTCCAATAGCTTTCATTGGGACAGGGATTCTCACCCTTCTTCCGT TCTTTTCAAGCAGCTCAGACGCTCCCTTCAAAGCCTCAATGGCCC	V
HoxAa	CNE50	TGATTAATTGTTGTGGTGGCGTTTTTCCACGCCCGCTTCTCGGTGTTGACTGA AAACAAGCGCCAAACGGTGCCG	AA
HoxAa	CNE51	TCAGTGTGTCCGTTAATGATTCACCGACTTAAATAAAAGGCATTTAAGTCAC GTGACGCTGCTCCATGTTATGGCTGCGGCGAGCAAACCGGAGCCACGATGCG GCCATATGATGATCATTACAACCCCACTTTCCACGAGGGGAGACCTGGCTGG GTAGCTCGCGTTTATTGCGTGGAGAGGAGAAGCAAAAAAAAAAAAAAGGAAA AGAGGCGCAATCAAATAGGAGAAACGCTCCCGTTTTTTTGTCGTGTGTGTGT GTGTATGTATGTAGTACACGGCTGTCACGGACGGATATGTTTTTCCACGCGT AGAGAGCCTGAGCTTTGGGAGGATGGATTTTATTTTGAGGCATTTCCGCCAA TTGTTCCTTTCTGTGGTGAATTATTGCAGCGAGAGGCACAGGTCAGTATAAA GG	V
HoxAa	CNE52	AAAAAGTAGGTGAACGCTGTGGGGTTGTGTCGTTCCGCAGGGGGGGGAACG AGGCTATAATGGAAGACGGAGTGCCACGCAGGACGGCTCGGCCGCGCTGAG GTCACGGCGCTCCGGCGTGATTTATGGGCGCGGGGACTTGGGGAAGATGGTG CTCCGCAGCTAATGACTATTCGATCCGTTTGTGGGGGAAAATGCTTACCGGC CGAGAGGGAAACAACAAACA	AA
HoxAa	CNE53	CTCCGGTAAGTCTATATCGTCTTTATTGATTTATCTTGTGATGGCTGCTTATTC CATTTTTTTTTTTAAATAATAAGTAGCACGCCCACCAAAGGCCTGACAGAAG CAGTGTGTGACTTTTCTTTGGATGTGCAACACCAGCAGGCCGGGAGCAATAA TTCCTTCTCTTCCACTTGTCTGCGAGGTGTAATTTGTCACTATAAACGTGACG AGGAAGGCTCTTGGAGCGCTGCCTCGCTCCTTTCAAAAAGCCATTTCGACTC CAGCCAT	AA
HoxAa	CNE54	GAATCAGCAGTTTGAACACAGCACAAAGAGGTCAATAATTAAGCCGAGGTA GCTGAAAAAAAAAATCCTCACGCGTGATTTTTTTGGCCCACTCACCATCCCA	AA
HoxAa	CNE55	CCTCATTTGGTTGGGGGAGGAGGAGGAGGAGAGGCAGCCACTTGATTTGAA CTGCCTTTCCAAGGCGCTACTTCATTTGCATAATTTTTCTTGTCTATGGTGTGA CGCTGCGGTTCGGGCCAGTTCACTGGTGTCCGCACGATCACGCAAGAAGTCC GGTCCAGCTTCGTATAGCCCCGCGCACAGGCTGCGTGCATGCGTGGTTTTATT ATTACTATGGAGAAGCATTGCTTCCAATGGCATGAAGATAAATTCAGCCCTT GAAGCCAGATAAGAGTTATCTCGCTGACCCGCCTGCTTTCGTAATCTCCATGT ACACATCCACGTCGAGTGGGGAAGACACGAGAAGGGTGGAAGAAAGGGAA GAGGAGAAGAGAAAAAAAGAAGCAGAGAGCTTCTCTCTCTCTCTCTTCCT CGAGTCTTCTTGTTGTGGTTGCCGCCTGCACGCATGGCGCTTGCTGCTTGCTG CTGCAGCCCTGCCCCTGCCTCCTCAACTTTCCTCTGGATTACTACTTTGATCG ATTTCGTCGCTGAATGAGAAAATATTGCTCGGTGCTCTGCATTGGTCGTCAG AGGTTAAGGTAAGCAGGCCAGGAATAGGCTGCCTCGGTTCTAAATGGCCGA GTTTGTGTGTCGCGAGCGGTGATTTATCACCGTATGACTTAGATCTCGGTTCA GGAAGAGTTCACACAC	V
HoxAa	CNE56	GTCAACCCCATCAGCCGCCATTATTATGGGTCAGCACAGCAAAAGTTCACAA TCTGGTCCTGAAGATGNACGAATGTCCACCCGGTTTTTCAGGTTTCC	AT
HoxAa	CNE57	AATTGATTTGGCGATGTGAAATGTTTTTGAGCGCATGTTGCTTAGCACAGAT GGTAATAGAAGGTTGGCCGGTGCAAGCCAGAAGGTAAACGGT	P2
HoxAa	CNE58	TTGCTGCCCCCTTGCCCTGCAAATGACCTTTGCCTACTGTAAACATACACATG CACTGATCACATAGCGTCCCCAATGTGTCAACAAACACACTAACATGAGCCA GTTGT	P2
HoxAa	CNE59	ATGTCATCCCACGGACCCTCTGCTCCATGTTTGGCTCGACAGGTAGTAAAATT TACTATGTGCTGTTGGCCGCTTGCCATTAAACTCCAAGTGTGCC	P2
HoxAa	CNE6	AGTTGCGTCTCCGCATATTCCTCCATTTAATTATCTCAAGTGGGGATGTTTAC CATCCCTCCCTTTGCACAGTGGGTCACGTGTGGTGGTGGCAACCAATGATAG CGCGACGTCGGCACTCCGGTACGCGTCACCGCTTTACGCTCTGGGAACCAGG ATGGAT	AA
HoxAa	CNE60	TAGCGGTGCGCGGCGCAGTTGTATCACTCCGCCCATTGTCTAGAAGCCACCT CCGGACCGCATCACCGTCTTTTCTTTGTGTGCGTGCACACTGCGTCTACCATT TCAAACTCGCGCTCTTTTTTTTTCCCTTTAGGTTTCTTTTAGTTTGATGATGTG TGAAGTAGCGTATAGGATAAATAATCTTTTTTTTTAAAGACATTTCGGAGGG CTCGAGGAGAACGTCGAGAGCCAATGCCTATTTAATCTATTTTTAAATGTTTT TTCAAATTCAAACAGGGTATTATGAAACATGTTATTGTTTTTTTTTTAAACGT GAGCTTATTTATCCGCATTTGAAGTTTTTTTTAAAATGCATTTTGTTGCACTTG TGAAAGGATCCATATGGCATACGTGAAAGAAATGCATTTGTCAGAAATAAG GGTATACGAGGTTAGGATGCAACCCTTTTTGCTTGATTTGTTCTATTTGGACA TTCTGATGTTATTTATTTTTTGTTTAGTGTTAAAAGTAGTCCTGGTATTATTTG ACACCCTAGCGAGAAGCTCACTGGGTTCACCCCCCCCCCCTCCAAACTCCCG TTTTCAACTTATCTCCACTTCCACCCTTCTTCTTCGATTAAATATTTTGTGTTT ATTAATTTTCCTACAGAATGTAAAATCAATCAGTCGACGCAATAAGGGTGGA AGGAAGGATACACTTTAGTCGTTTCGTTATTGAACAGATAAAACACGACAAA	V

		CCCCTTTCATACATTCCATGCTGCTCCAGTTCGAAGAAGTAAATAAAACAAT AGAAATACAATCAAAGCATATTTGTTTNAA	
HoxAa	CNE61	TCTTCCTGATGAGACGTGAAGAAAACTCTCGGGTGGCCTGCTTCTTCNCTCCT CTCATCTCATCCGCGTGGGCACATCCATAATAATAAAAATGTCAGGCGCCGG ACCCTTCAGCAAAGCCCTGCAAGTAGACAACAATATAAATGAGAAATCGTT	P2
HoxAa	CNE62	TGATTGATATAACAGGGCATTAGCAGCCATGACATATGGTGAGCAGGGAAC CCTGGAACAATTGAACCATGGTGTGTCA	AA
HoxAa	CNE63	TATTTATCTTTTAAAACATCCCTCCAGCGACAGCTCGGGCTAGAAGGAGCAG AAAATGGTTCTAATCTGAGCGACCAGTGTTTCTCGCCGCTTCCTGGCTGCATT TGATCCGGGGGAGAGTTAGAAGCCTTAAATGTGTTGTGAGGGCACCGAGCTG TCAGACCTTTTGGCGAGTAAGATTGATCGCGCGCAGGCTTCCAGCACTCTTT GTTTGGTATATAAGCAACAAACTGAGAGAGACCTACACCACTTCTTCCTCTC TCGTATACTCACCAAGTCATATTTTTCTTTCGATGGAAATA	V
HoxAa	CNE64	TGGGGGCGTTCTCGCCTGACATCCTCAAACAATACGAGAAAGAGGAGCGCA CGAATCCGCGCGCCCTAAGCTTGGGGTCCTTCCTTGTAATAGTGAGAGTCAT CCATTCTTAAAAACACACTTTTTTCGGGTGGGGGTCTCTTCTCAGAGTGGCGC GCGGTCGGGAGGATAAAGATGTAAATAAAAGTGGCACGGCTGAGTGAGGCA GCAGCAGCAGCAGCAGACTTTTTGGATCAATCAGGCAGTCAGTGGCTTCTTT TGATTAAAGCCCAAATTGTCATTGGGCAGAGGTAATCATGTGACAGGCAACT CGGTCCAATTTCAACCTTGTCTCCATGAATTCAATAGTTTCATAGTAGCTCGG TCTCCACACGGCCGTAATCACAGAAATAGGAAGAGGAAGCCATCACAGGAG ATTTTTTGAATGATTTTCGTCTTCGCTTGGCCATTGAGTTCGTCGACAATAAC GCGTGCAAACTTTCCTGGGCTCGCAGGGAGAGGAG	V
HoxAa	CNE65	GTTTGGTTAAAGTTTGCTCTGATTTATTCCCCCGGTCACACTTTTGATGTGTG ACAGTAATGAAGAGTGATGGAGTTCCGTTGCCGTGGAAGGCAGTTGCAGCAT TCATTAGTGTCCCACTCCCTGCGTGGAGCTTTTTTTTTTT	V
HoxAa	CNE66	TATGGAAACTGGTCATTTTTATTTTTGTATAGCTTCCAATGTCTTGACTATTTT TGTTAGAATACAATTAACTTGCTACAATCTTTTGTTTCGTGCTTTGAGT	V
HoxAa	CNE67	GTGCATGCTGCAAGCAAACAGCCGAGCCATGTTTGCCTCCAGGACCTCATAA ATCACCGCTGTGGCATGAATGGGACGAACATTAAGGAGCCCAATTGTGACTT AATGTTCACAAATACAAGCCCACTTACATTTTCGTGCGGGTTTCCCTGATATC TTAGCTTAAAGGGGGCTGAGTCTATCCACGCGTTGGGCAATGATTCATTTTTA TGGCCTGTTTAAACAGG	V
HoxAa	CNE68	AATATTGGGATTCGATATGATGATTATGAATAGCTTTGCTCTATTGCTTTGCA CTTACGTTGACTTATGGGAATCGTTTAATAGCCC	P2
HoxAa	CNE69	CATAACCAAATCATTCCGAGACAACGCGTCTTTTCTGTGCACACATGTGCTC ACACAAAGAGCACAGAGAGCTGC	AA
HoxAa	CNE7	GAATATCCACGTGAAACGACGCTATTTCGGATTTGGACCTTGGGAATGTTAA ATATTAATCTCCATGGCTTTTTCTTCTTTTTTGATCTTACTCACATGTTTAGCT TGTAGTGCTTGTTGGAACGCGCGTCATCGTTCTTGTCCATTTTCATGGTGCAC TTCTCGAATGCTCAGATAATCTTGCGGGATTTGTGTGTGTTGATATCGTGCAT ATAGCTTTTCATAATGTGTGACTTTTCATGTGGTTGTGTATATAAAGTAAATT CGAGCATGATGTCAAGCGTATGTGTGTGTCCGTGCCAATAAAGTATGTATTT GGGATATGCCTCCAATATTGTTGAAATCAG	AA
HoxAa	CNE70	ATTTGATCACGCCATAATGAATACATACACCGTGTAAGTTAAGTCA	P2
HoxAa	CNE71	ATATTTGCTGAGTAAATCCCTCGCAGTGTTAAAGCGTGAAACATTTGGCATC GGAATGGGACACTGATAGCCTCTGGCAGTGACACACATTGACACATTCATTC ATCCAGA	P2
HoxAa	CNE72	CATGCAACCCCGTCCGCCGTCCGCCACTGTCTGCCACCCAGCTTCATTGTTTG CTCGGCCAATGCGTAAGTGCATGACGCGTCACCTTGATCGACGAGCGGCTGG AATTTAAATAGCCACTCAAACAGCAATCTGTCAGCGGCTGAGAGGATAAGCT CAAGTGTCTCACTGGCTGACCCGGACCCACGTGACCACACCCGTTCATTGAT GTTGG	V
HoxAa	CNE73	AAAAATAAAGAGTATTGATCCAAACATGGAGGTTAGGATATCGACAAGGAT GCAGCCCCACTGGCCCTGTCAGCGTT	AA
HoxAa	CNE74	TAAAAGGGCAAAAAAACGCCCAAAGAAGCATGATTAATGAGCGTAGTTCAC CTGAACCTGGATCCACCTGCAGGATTGGCTCTCTTGGAGNTGTGGAGAC	P
HoxAa	CNE75	AATTGGAATCGCAGATGACCCCCCAAAAGAAAGCCTCGGGTCATCCCACGGT CACACTCTGAGTTTGTTGTCACAGCACGCGCACGCACAGCACACACACACAC ACACACACACACACATACACTCAGGCCCAAACAACAGACCACACTGGTCAA ACAAGCACCTGAATAGCAAGTCACAGTCCACCTCCGAGGACCTGTTTGACCA AAACACCACAGCAGACT	AA
HoxAa	CNE76	GACAGAAGTCAGTCATCAGGGTTGTCGAGGTATTGGCACATGTGCGCTCATT GACCACTCCCCTGTTGTGATTGGCAGCCAAAATGAATTCCTCCATCGACAGA CGTGCGTCTCAGTATGCCGATAAAGGCGGCCAGCCTATAGAGCCATTGACAG TATGATGGAGGGAGAGACGGGG	P2
HoxAa	CNE77	TGAGGCCCAGAGAGGCGAGAGGATGGCTACAAGGGCTGTGGCCTCCAGCTC CATCCATCACACACGTCCTGACCTTGGCTTGACCCAAGCATGGCAGCCAATG AGGCGCCCTC	T
HoxAa	CNE8	AAAGCATGCGTAAATGGAAAAGCAATTAAAGGAACTTCATCTGGGGCACAG TAGGTCCTTTTGCTCTTTGTGCCTCTTTTCCCTCCCCCGAACTTGAAGTCGACC	AA

		GTTCAGACAAAGCTTTGGCTTTGAGAGCGCTTCCAGTGCCAGACTGCACTTT GCTTGCCAAAAATCACACAAGAATG	
HoxAa	CNE80	not available	P2
HoxAa	CNE81	TTTGCCACCGGGCGGCGCTATTGGAACCAATGTCACGGCCGTGGA	AA
HoxAa	CNE83	TCATCACATGCCATTCTTAACATGCCATTTCCTTA	P
HoxAa	CNE87	not available	P2
HoxAa	CNE88	not available	V
HoxAa	CNE89	not available	V
HoxAa	CNE9	GTTGTTGTGTCTCTCGAGCTGCTGACTTGGGATCCGTGATAGGTGGCGACTTG CTGTCACGCTTCGCCGCTTTTGTGGTGTCTGCCTGGCTGTGTGGG	AA
HoxAa	CNE90	not available	V
HoxAa	CNE91	not available	AA
HoxAa	CNE92	AACAAGACTGAGATGGCAAAGAACACAAGTGGAAATAGAAGGAAGGACCC AGTGAATATTTTGGTGTATTTTGCCCTTCGTCCCCTTCCACTCTGCGGTCGTCT TTCGCTTATTCACATAAGCTTGATCTTTTTTTTTTTTTAATCCCACAACGGCGAC TGAAGGGTTCATATATTTGCTTTTTTTTTTTTTTGTATTTTTATTTTTTTACGAA TGAAGGATAACGCACAGTCACATGGATCAAAAAAAAAAAAAACGTGTTGCAT TTCAGTTAGAAATGACCGGATGGGTGAGGCTTGAATCCTTTCCTTATGGATG GTTTATTTTTATTTTATTTTTTATTTTTCACGAAGAAATATTAATATACTTGAA CGTTATGATTTAAAAAAGCATTTCTTTAGGAGTTCCATTGTAGTCACGGTGTG TGTTCGAGTGCTGAATGTATAGTCCAGGTGACCTGCCCATATTGTAATGTAC ATATATAGACAAATCATGATTGAAAGGCACAACGAGGATTGCATTTCTACAG GTAGAGACTTTTTGGTTGTTAGTTCGTTCGTGCGCTCTTCAATATTTGCTTCGT GCTCGTGCGTCTCACGTGTTGTTCTAAACTGTGAATATACATTCGTGGTTGTG CCGTCTCACTAGTGACCCCTGTGCTGTAGTGTAGACTGAATATGTCCTTGTGA AATAAATGTTCACTTGTAATGCTGGGGATGGCTTTTGATTGTTATTC	AA
HoxAa	CNE93	TGATTGATATAACAGGGCATTAGCAGCCATGACATATGGTGAGCAGGGAAC CCTGGAACAATTGAACCATGGTGTGTCA	AT
HoxAb	CNE10	CCTGCGGGCCCATTGAGGCGTGTGTCATGTAGTGCCAGTGGTCACATGGCT	V
HoxAb	CNE11	AAGGTTTATGGAGGGCCACGAGATTTAGCCTGACAAAAAAGTATATTATAAA CGACAAGATCACGTGCTTGGGCTTGAAATTGGCCGTGAGGGGTTTAAAGTCA AG	V
HoxAb	CNE12	ACAGGATCTGNTGCAGCTTTTAATTTGATATCCATTACAAAGAGCCGG	P2
HoxAb	CNE13	ATGGCTTCCTTATACCGTACATCCTGTGAGACTCCCAGGAATTCCGCTTTGAT TGCTGCACATCCTCCCAGTTGCTCCAGTAACTTGGCCATAAAAAGGCGGACT CGTCTGGAGCATTTGGAGTGGAGTGCAATAAAGCGTCTGAGAACCAAGGTTA TTAACTGCGACTATAAGGAGGCTAGCAAAACAAGAGGNGCTGCACCATTGA AGGCTGCGGCTCTCCTTGGTGCAGCATCGGCACGCCAGGCTGTTGCAGGAAT ATTCTGGATTGNTTTTTTTTTACCCCCAGTTTGAGTGTTGTGGCGCAGTTGAA GGCCTTGCTGCTTTTGTCCACAGGAGGGCGCTCCACCGATCAATGTCAATGC CGTTTGATTCTTTCCG	AT
HoxAb	CNE14	TGCACCTTTACTGCTCCTTTTCTTTACAATGCTGTAAAAGCAGTCATAAAGTA ATACATTGCACAAGACTGCTTG	P
HoxAb	CNE15	CTATAATAGGCTAACTTATTGAATGTTATTATTTC	P
HoxAb	CNE17	CGGAACTGTGGAGTGGTTTAGGTAGTCTCATGTTGTTGGGCTATATGTGACTT GCGTCCCACAACAGGAAACTGCCTTGATTACCTCAGTAAAAA	V
HoxAb	CNE18	AGACAAGTCCTCGGTTATGATCATTGCAGTCTGCCAGTTGCCTCGCCGTAGG AAATGCCGGATGCGTCTGCTTTGCGCCTCAGAAAAGACACAAAGGCCCGGCT TTATTGGGGGGGATATTTAGCTTAGACAGCCATGTTGTAACCTTAAGAAAAA CAACCCCAATTGCCCGTGATAAAAGGGGCCGTTATGGCACCCAAATAACCTC GCCCTACTTTCC	AA
HoxAb	CNE19	GCACACCAATGAGCGTGCTCCATTAATAAATAAAGATCTCTTGACAAAAATC ATAACGTTTATTGCTTATAAAAGTGAATGCTACCAGCGCGAACAGGCCCAAT AAAGGAGAGGGAATTTGCTATTAAAATATGGTGGTGAAAAGGCTGGCTATT GAGTTGAGCAGGAGCAGCAGCAGCAGCAGCACAGCAGCAGCAGC AGGCTGATTTATTATTTATTGTTCTGTCCCACGGTCACGTGTTCGGGGCGCCC TATCCTTGCGTGGACAATGTTCACCGCGTTCAACTTGCATTGGGCGGCCAGG CTGGTGCACGA	AT
HoxAb	CNE20	TGCATGAAAAAAAAATCCTCGACTACCTGAATGTTTTTCAGCTACCTATTTGC AATGTTTA	P
HoxAb	CNE21	ACTGTTTATAAGGTGTGAAAGGAGCAAAGCATGTCGAGACTGTGGGACAGT AGTAAAGGACACATTTCAAATGAGCCTCAGACGAATATGGAAAATAGCTGG AAGATTTGGCCGACGACAGACACACACACACACACACACACGCACGCACAC ACACACACACGCACACACACACACACACACA	P2
HoxAb	CNE22	GAGTCCCTCTGCGCAAGAAACGGCCAGNAGGTGGCGGCAGCGGCTCA	P2
HoxAb	CNE23	CTAAACCAAAATATACGACACTGGAGCAGGCTGCTCGGACTGGGAGGAGTC TGTAAGCTGTCTGCTCTTTTAATTGGCTGCGGGCTACAGTGAGAAGCGTGAG TATCNCTAATTATTCAGCAATGTTTTGCACAAGAA	V
HoxAb	CNE24	between CNE23 & CNE 25, between HoxA9b & *Hoxa2b*	V

HoxAb	CNE25	CGGGATGAGGCGGTACCACCACCAGCACCCCCACTCCACCGATCAATCACGA GGAGGGACAGTGGCCTCTTTTGATAAAAAACGCCAAATTGTCATTGGACAGA TGCAATCATGTGACAAGCCGCA	V
HoxAb	CNE26	ATGATTGATAACTCCCATGTTAATAACTTCCATGTTATCTG	P
HoxAb	CNE27	AAAGAAAGACGATTCGCCTGCTTGGAGTCAAGGAGGGCAAAAGCTACGACT TTTCTCCGCTCTCTAATTAAATTTTATTGCCTATAAAACTCACAGCATTGAGG GTGCTTGTTTATTGGCTTTTTCGCTGTGTAATGCCTTCTTTGA	P
HoxAb	CNE28	CACACACAGCACACGGGTGAAAGGTCAAGCTGTCTCTCCTTTCCTCTCCAAG CAGGGAGGAACTAGTGCCAGGTCAGAAGGGGGTCATGATTTGAATGCACAG AG	P
HoxAb	CNE5	CGGCATGTTTTTGATTCAAAGAGCCAGTGAGGGGCCGAAGTGCCACAGGTCA TACGATCACAAGACGAGCNCATAAAAAGTCCATTAAGCACTGTAGGCTGAG AGCTGCGTTGTCATCCGCTTGATAGGTAC	T
HoxAb	CNE6	ACAGCCTCCAAAAGGTGGGAAAATTCTATTAAAAATA	P
HoxAb	CNE7	GACCTTGGACTCCAGCCGCGTCCT	V
HoxAb	CNE8	CCTGATGTGCCTTATTATCTCCTGTGCAGACCTTGCTCATCTCCTTTGCATATC ACGTGTCTGTCCCGGGCCAATAACCTCGCGATGGTTCCGTGTAGGGTATGGT TCCCCTGCCAAAGAGTCCCTTGTCTGGGTGCATTCGTTCGAGGGGGGTTCGA GGAAAAAGAGCCGCATGCGTGCACGTGTCGGGACATGCTA	AT
HoxAb	CNE9	TCCTTTCCTCGCGTTTTAATTGGAAGCAGTCCACGCAGCGCGTTGACAGAAA GTCGGAACTGTTCGCGACAAGTGTGTGCGATTCTT	P
HoxBa	CNE1	GACACTCTTTACAACGCAGTCTCCGTCCGCCCTTAAAACCAACAATACTGAC GGAAGTTCTTGGCCCATAAAGATCTTTCTGCGCCTTGTCTAAATAGTAGAGA ATGATGAGGAGGACTGGGAAAAAAAA	AT
HoxBa	CNE10	GGTTGGAAGAAGCCGCCGCCGACTGTCGTAAAAAACAGTAAGACGGCAAAG AAGAGCAAGAGAGCAGAGAGAAAGGGAATTCAATGCGGTGTGCCGCAATAA AATGATATGACGGCAATAAAAGTTTATAGCGTATAAATTTCTGAAGGTTAAG AACTAAACGGCTGTAAAGCAAACACGCCGAGAGAAAG	V
HoxBa	CNE11	AATATGAAACAAGTCATTTGCAGAAGAGTAAATCACAGAAAAGCCGTTTGA GAGCAAGACAGCGCTCCAGATTTCAACGCTATAGCCGGAATTGTGATGCTCG CTCGCCTTTCCTCCTCCACTTCCCTTTCTCTGGTTTCTTC	V
HoxBa	CNE12	TCAATCCAACATAGTCTACTTTAGTGAAGAAAAAAAAGAATGACCACAATG ATAGCTTCCTTTCTTTGGTGAACACTTGTCGTGGAAACATTTCTATATTGTTC GAAGTTTGTCATTAACAATTTTATCTTTGGCAGCTGTATAGTTTTTAAGTTAA AATGATGATGGTGCACTTTTTACTGTACTTCTCACTCCTATGTTGTTGTTGTTA TTATGGCAATTGTTATTAATGATGAAGTTGATGGCGTAGCAATTAATGAAAT GTCCAACAACATGAAACTGCCTATTTATGCCGTTTTAGTAGGTCGGGTCTCTT TTTTAAAACACTCTAATCGTTGTTTATGGTTTTAGTTCGTTTCTGATTGATGGT GTTTGTTTATGCTGTCTTCCTTTTCTCTTTTGGGATTAAAAGCATGCGCACAGT GCGGTTAATAAAGAGAAAA	V
HoxBa	CNE13	AAGAATGTTTCTATGGTTCGACCAAATCATTTCATTAAACATTTTGTGTCGCC TTGGTTGCGCAGCGGTTACTGTCAGTAGCTCAAATATGCCTGACTGTCTCTTT CTTTTTTTTATACAACCCCTCCTCTCTCTTGCATTCTCTGGGAAATCTTTTATA CTCTGTGACCTCCATAAAACAACAATTTAAGACTTGGTGAATATTCATTAGTT TGTATTGTGTTTAAGA	V
HoxBa	CNE14	TTACAACGAGTCCTGGTGACTTTATTGCACCCACCNAAGGGGCGGAAAGCCA AATAAAAGCCCAAATCGCACAAAGAAACCAAATGAACAAGGTATCGAGTGG CCTGTTTC	P
HoxBa	CNE15	AGAATTGCTATAAATTCTTTGTTGTTTTATGAAAATTTACAACTTTGTGATAC AAGTTTATGAGTGGCCGCGCGCAGGGATTGGCCAATGGACTGGTCATGTGGA CGCCCTTAACGTGAACATGAACTTTTTATGATTTCCCAAGTGGCTATATTGCT GCGGCACTGCTCTCCGGCCCGAATCAGCCACCTCCGCTGAGCCCAGCTCCGC AAGCCTGCGTTGACG	V
HoxBa	CNE16	CATGCCCCGGGGTCGCTTTGGCGAGCGGACACAAACACGAGCTTATAGTGGC GTGAAGGGCGAGCTTTTGCTCCCTTGAAACGATC	P2
HoxBa	CNE17	not available	AA
HoxBa	CNE18	CCTGCACTGTCCCCTGCCATGCAGCGTTGGTAGACATTAATGATTAAACCGC AGTCAGCTCCAACACAATGCTGCAGCTTCCCGTTCCCACAAAGCAGAGAGGA AACGGAGCGTCCTGCCATTGCAGGGAGATGGAAAGATGGCGAAAGGGGTGG GGGTGGAAATGAAAGTTGAAGAGAGTTCCTCCGTCTGTCTATAGTGATCTTG CACAGAAACTGAGGTTTAAACTTATTAGCTGACCCGTGTTACTTCGCAGAAC TGCGAGAAGGCACGGACTCAGGAATGGTGGGGGGTGGGACACGGTGAAGTA AAAGTAAAAATAGAGTCTTTGTCTCGATTTGTCAGCCATACTGTCTGCAATAT GCCCATTTAGACAAATCG	AA
HoxBa	CNE2	TTTTACCACCTTTTCCCCCACAGCACCCCTCCTCTTTTGTCTGCATTGCAAACT GTCTTTTTGTTTCCCAGGTTGGACTTTGCTGTTGAACTTGAGACAGACAGACA GCTTGTAAACTCTTAGTAGTCTGGCCCATGTCTAGTCTGGAATG	V
HoxBa	CNE21	AAATTGTATATTTTATGAGCTTTTGCATTGTTTGGTTATTATGTCTGTGTTTGT CAAGTACTGACCAATACTACCTATATGTCCAATAAAATGTAAAGAAATGTAG TTTAAGTTTGCAAAGAGGTTGCTTTTTTTTTGTCCGGGAAAATTCATAAAAATA ATAATACTCAGCCAGTCGGCATATCTCTATATAATTTCTTCTAGACAGGGGG CACAATGTTGCTGGTGGTCCGAGCTAATTGATGAGCTGGCCGATGGGGTGAA	

		AGTTGACCTGGCCTCGAGCTGAGCGGCCAGACACGACACCCATAAACACAC ACTTCACCCAGACAAAGTGGACAATGTCCTTGTGACCCTTTGAACCATGCTT GTCCCCTGTGTGACC	
HoxBa	CNE22	AAGGGAGGGAGAAATTGCCCCCGCCCATGCCTTTATCATTAGAAACCTCAGA ACATGAATTACCTATCCAGTCATCGGCGAGAATTTATGACGGGTCAACAAAA GCACGTGACCGCTCCCTTCAGCTCTCTGCCCCCCCCCCCCCTTTTGCGCCCTAC CCATCCCACTCCCTCAAACATACACTCACCACCCCACCCCCATATTTGGACG GCGCATACATAGCTAAAAAACCAAGTACATCGCCATAATTCAAACTGTACAT CATAAATCGATGAAAGNAAGCGCGTTATAAGGACCACGAGAAATCCTCCAA ATTAAGATTCCTATAACATTAACAATACGCAGCTCTGTAAACCAACCTAA	V
HoxBa	CNE23	TAGTTTAGAAAATTCCACAAAAGTGTCATAATTCTGCCAAGATAGCGCGACA AGCAGGGCAGTAAAAGCAACGGGAGAAATGCAGAAAAAACGCAGCTGGGT GGGAAACGGGAGGGGGGATCAGGGGGGAAAGGCCAGCAAGACTTTTGAGG TTGTTGACTTAGAAAATTAAAGCTTTATCTTACGTGTGAGCACTTCA	AT
HoxBa	CNE24	not available	V
HoxBa	CNE25	CTGCTTTGCGTTGCACCATAAAAACGAATTCAGCGTTGACATTTACATGTCG AACAGATGAGTGCTTTTTATCTTCAAGTTGCATCGTAAAGTTCAGGCCTGCGT GGCCTNAAACTGATACTGCTCACTGGCTGTCCACTAGTCACGTGGGGTCCAT AAAGCTAGTTTTATGGTTTTGGGGAGTTGACATTGTACAATATATTCAAGATT CTAGAAATCAAGTGACTGTTTAACTGCTTCTAGGGGATCCTAAAGGGGTCAG TAAAGT	V
HoxBa	CNE26	not available	P2
HoxBa	CNE27	GGTTCAAGGAGTTGAAGGGTCAAAAGTTTACCGTCGTGCATGCGTGGGTTCC TCCCTCTCCTCCTTCTCCAAACCAATTGCTGTAGGCTAAGACTACACAAACAT G	AA
HoxBa	CNE28	GAGGTCAGTAGTCACACCATGACGTTGAAAACATGGGGACGAGCCTCCTCCT TTTAACAAAGCCTTACCGCAGATTAATT	P
HoxBa	CNE29	ACTCTCCATAACAAAAGAACGTGGGATTGAGTCTAGGTTTATTTTTCCATATC CACAATGGAGCAAAACTGCATTTTTCTTATTTGCCGTTGCCACTTCTTTGTCC AGTGATTAATATTTCAAAACAATGACAAGACCAGTGTTTGCAAACAAGATCG ATTCACTCTGTAATATATTCATTTATA	AA
HoxBa	CNE3	TCGTTTTTCATGGACATCTTCTTTTATAGAACCAAGGTCGGCTGATAGGCAGC CAGCGTGTGTTCAGTAGACTTTAATCGCCAGTCTTGGGGCAATAAAAGAGGA TGGTGAATTATACAAGTGTAAAACACCCAAGGCTCAAGCTTGGTGC	P2
HoxBa	CNE30	CTGCTGACAAACTTGACATAGGCAGGGGGAAGTGTCACAGTCACACTCAAC ACACCTTGATCGATATACTACCTG	AA
HoxBa	CNE31	ATCAAATGAAGGTAGCAGTGTAACCTTTTTTGGTTGATGATTTTTTTTATCGA TCCCAAACTGGATATAGAGCTGTGTCTGACAGCTGTTGGGTGGCCGCACAAG AGTCCATTTCTCTCTCCCTCATTAACA	AT
HoxBa	CNE32	AGATTTACGATCGTCTGTTTGCACGGCCAACATAATTACACCCCCCATAAATT TTTATTACACCTCTACGCCGAGGTGCCTTTTTGAAGTCCGATCACAGGCTCGA GTCCTTGCTTTATGACAACGCAGACTGAACTCATAAGTTAGGTTTTATGCTGT GGTAATTTGATTTTAAGTGGTAGGTTTGTATAAACTGTGCGCTCCACACTGTC GAGCCTTCCTTCCCCACTCCATGTTCAAGGGGTGGGGAAGTTTTATGGCATCT GAAATATTCCTGTTTATGGAGCACGCCGTAAAACACTAATGCAAAGGCAAAT ACCGCAATCTATTATATCTTCTCAGGTGACTGTA	V
HoxBa	CNE33	TGAGAAATCAGAGACTCACGATGATTAATTTTGCCTTGCCGCTAGTGATCTT GGTACATTACGTCCTGTAATCTCGTTTTAAATCACAAACCGTATTCTGTTTC	V
HoxBa	CNE34	AAGAAAAATGTGAGAATTATACAAAAAAATAATCATTAATCACNAGTTCTTC TTAAACCCTTTTCTTTTTCTATTGTCATCTCGCAGC	V
HoxBa	CNE35	GGGTTCAAACAGGGGACACGTTTTTCTGACCGATGAGGTCGAGTGACTCTGA AGCAAGAGGCTGACCTGTGGTTA	AT
HoxBa	CNE36	GCATGCTTATCTTAGCTGTTGTTCACACTCCGAGTACTATAATTTATGGCCAT CATTACAGACGCCATCATCATCACTGCAAATGTTTGATTTCAAT	AA
HoxBa	CNE37	GAGAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG AGAGAGAGAGAAGGGGGCTGCAATATAACAAATTTGGACGGGTGTGCGTCT TGTGGGTCGTTGTGGACACAACGTGAGTATTGCGGAAGAAAGAGGACCTGC GTGGGGGGGAACGAAAGCACCCTCCTTCGACGCTGCACTTCTCTCGCCTGCA TCCATCACTGGCTCATATAGGCTTTTTAAACACCTTGGTAAATATCCCCCCCA AAAATAACAATAGTGATATTTGACTCATTGTGCATGTTGCTGTCGTGTTGTTT ACACATCTTCC	T
HoxBa	CNE38	CTATATCTACCCTGTAGATCCGGATTTGTGTCAAAATCTTAAAAGCAATCAC AAATTCGCTTCTAGGGGAGTATATAGT	V
HoxBa	CNE39	TTGAGTTTGAGGTTAAAGATTAGCATCTCTGATAAAGATTAACT	P
HoxBa	CNE40	CTGCAAAGGTGGCCTCCTTTTGTCTTGGCCTGCATGCTAGCTTGCTAAGCTAA TCGGGCCTCCATCATTCTCGGGACAACCTGTGACTTGTCAACTTGTTTGTTTT ATGCCTTC	T
HoxBa	CNE41	CATTGTTATCAACCAGCAGCGGTGCAAGCAGGCAGGCAGCGCTGGTAGTGG AGGGGAGAGGGGGGCGGGGGTCTACAGCAGGCAGCAGGCGGGCTCCCGGTA AGCTCGTTCAGCTTGTGTGCTAGCCATAGCCTGTCGGTAGGCTGTCCGTATA	T
HoxBa	CNE42	GACGGCAAATCTGTGATGCGATCTTGCTGGGTCCTATCGGGATTGTACCGGG ACCGCCGGTGCTAATAAAGGTAAGGATGCCTCACCAGTGGGCAGGCCAGGC	V

		GCTAATGCACGAGCGCGTACAACAAAGCAGAAATATCTCGTATTTGCGCATT GTGCACTTTGGGGTTGGCTCGGTTATTTATGACCCGCATGTGCTCAATGCCGG TTCAACAAGAGTTCACTATGCGTTAAACTGTCTTATAGCAACACCCCCCCCCC TCACCTC	
HoxBa	CNE43	GGTTTCGTATCTCCAAGCTTGCATGGTGCGTCACGGCAGGCGCCATTTGTAAT TTTATATCTCAGGATGAAGCAGTCACTGGACACCAAGGCCGACCTCTCCCTC CTTTCGTTTTTTGTTTGGGTGCTTGTGTCACTATGCTTATTATTCTGGGAGACG TCTAAGAAAGTTGGTGATTGTATAAAGCTGTTGCCCTTATAAAACAGGGGAA GATAAATTACGTTTGCCCGTTTTAATCGTAAAACAGCTATGATTCAATAACAT GGTCCTCCTACCCTCACTGCCATGGTCTCGCTAGCTCCCCTTTCTTTTTTTTTT TTTTTTTACTCCCCCTCTGCTTGCGACTATTCCTCAGAGTAAATAATTAGGAT TCCAGCTTGTCTTGAGGAAGATAAATGGCTGACATATGTGACGTGCTGAGAG CTGGAGAGGGGCTGGACAGTGTCTCTGATAAA	AT
HoxBa	CNE45	CATATANAAACCAGTCTTCCAGATGAAATAGTGGAACATTTAATGGGAAGTT ACTGAGGTGCTAAGATTAATGACAGTGAGGTTTCTTCCCCTCC	AA
HoxBa	CNE46	TTTTCAAAAGTGTGATTTACACAAATGCTTTTTGCATCTGGGTGTCATTGTT GCACTTAGAGTTTACATTTTTTAAGTTTAATGTAAAAAAAAAATGTTATTTTG AGGAAGCCTCTCCAATCAGAGTCCTTCGCAAATGTGCGTTCAAGTTAACGTT CATATAAAATATTTATTTGTTATTGCCGGTTTAAAATAATTCTCTCTCTCCATT ATCATTATTTATCCATCGTGTATTTATATTGATAAAAAAATGTAATTTTCTAG CATTTACCTGTTATGAATGAAAAACAGCATTGTGTTTCTTCTTGTCGATTGTA CTGAAGTTACAATTTTAGTTCTTGTATTTTGATAGGGAATAAGTAATTTTTAA GTTCCTACAGTTCTTTTTACAAGAATGTCTTGACATGTAAAAGGAAA	V
HoxBa	CNE47	TAAATAACGGCTGTACTTCCCTTTACATTTTCGTTGGACTAACAGACAATAAA AACACCAATAAGCAAAGATTGAATGTAATTTATTGCCTCGTAAAAACAGCCA GGGAACTTGGCTTCAGCCTCTGCTGAAAATAGCGGAGCCCCGGAGCGTGC	AA
HoxBa	CNE48	TGTTTAAGAAGGGGCTGCGTCAGGCTCTTTGCAGTCACAGAAAGATTGATCA CCGCCACCCACCCGGGGGTCC	T
HoxBa	CNE49	GTCACGTGACCAGGAATTGGCTGCATTTTCACCCCATAGTTCTTCAGTTTAGC CTACCCAGGTCCCCATACGCTAGTAATATCACCAATATACACAATTATTATAT TGCCACACGCATTTACGCTATCGCAGCCGGAGCGTGCGTGCTCGCGCTTTTTT ATTTTTGTTAAATTTGAAGGTGTTTCGAGGCTCTATGAGTCTGCTTCATTGCG TGCGTGTGTGTGCGTGTGTGTGTGTGCGTGTGTTTTTGCTTTTTTATTTTGGGGA AGGGATTTTCAGGGTTCTCGTCGTTTGATTGATATCCGTTTGAAGGAGCA	V
HoxBa	CNE5	AGGCTGCAACACTTGTTTCGCTATTGCGAGCAAGTTGCTTCGAAAGAAAGAA AGGGGGGCGGGGGGGTAGAGCGGGTGAGGCAGAGAAGGGGGGCAGAGAGA GAGAGAGAGAGCGAGAGAGAAGTAGCAAAGGGCAGATGATGAGAAGAGGG AAAGAGAAAATCTCCTAATGTCTATCTTCTTGTTCTCCATTTAGNGAACTCGT AGTGTTTCACAGATAATGTCCGAGA	AT
HoxBa	CNE50	not available	AA
HoxBa	CNE51	AGTTTTAAACCCTAGTTTATTTATTGAGATTCTTTAATAGTGATATCCAAATT ATTTATTGTATACATATTTTCATAGTGGAATACTAATATGATAATAAAAATAC AAANACATAAAGTGGCATTTGTACTTTTT	V
HoxBa	CNE52	CCTCCTCCTGCAGAGTTAAAATGGCGCTACAGACAAAGCAGGATGCAAACA ATGAGCAATGGCAGCCATTACAGGA	AA
HoxBa	CNE53	CAGACAGTCATGTGGGTGTCATTTGATGGTCATCTTGTACTAGTCTGTTCCTG CATTCTTCCTACCCACTTATTTTAGTCTGTGTCCTTTCATTCACTATGTGATCA TTAACCGCCATCCTTGTTACCCACCCACATCACTCGTCATGACTTACA	P2
HoxBa	CNE54	TTTGGTTAACCAGAAGGTCTGCTGTCTCCCTTAATCTCGTTGCTATAACAGAT CAGCAAAATTATTCCATTAGCAAAC	P2
HoxBa	CNE55	TTATTCCCTCACCCTCTCTCCGCCCTGACAGATTCGTGCTGATATCAGATTGA TGGGTCAGTTTGATTGAAGTCTCTTTGTCGTGCCCATGTCACGGCGAGTGATG GATGAGGCGCGTGCGCAAATACAAAACGCGGGCAC	V
HoxBa	CNE56	CACACACACACAAACACACACACAAGCAAACACACACGCAGACACACGCTG AGCTACAGCGAGGAATGTAGGTTAGCTTTCCGGGTGGTTCGCTAAGCTGTCA GCCTTTGGAGAGCCTCTGGGATATCTAAACATGTAAACACACTNTTGCTCAC CAGCTTCATCACCTCCACAACTCGTTTGAAATACTAGTTGTGGTGGAGCCTG GGTGG	AT
HoxBa	CNE57	CCTGGATGACAGAACGGTTGTGGCACAGCTTACACACTAATGGATTGTTCTA GAAGTGATTATACACGACTGTAAAGCACCCACTTTAGCGAGGGACACAAGT ACTGCACACCNCGGTTGTGTCATCCATCTTCTCCTCACCTCCCTACGAGGCAT TATTTTCCCTGAGAAGAAATCGAT	AA
HoxBa	CNE58	GTTTTATGATATTTGAAGTGTCGTCGTATTTAAATGGCGTGCGCTGGCAGTTT TATGGGCCGACCATAAACAAGACATCG	AA
HoxBa	CNE59	ATTCAAACACACTTTTATTTATGGCACACTCCTCCAATTATATCGTTGCAAGT CTGT	AA
HoxBa	CNE6	GTAAGTTATTAGGGCAGGAAATAGCACTTTGGAGAAAGACACACGATGATT GGGGAGAGCGATTGTAAATGCTTTTAATGCTGCTGTAAAACAAGAACAAAG AGCTG	AA
HoxBa	CNE61	TGACTGCACACAAAATGAGAGCATACAACAAGGCGCGTGCGGGGCAGTGTG TACACTTAATCTCATTTCCTTGTCCACGCGGCGAGTCCCGCTCCAGTTCAATT GGCTCGCTGTAGTCACATGACCGCTTAACTTTCTTCACTTGACAGAAAAGTA	V

		GGAGGGTTCAGCGGAACAGGAATGACCGACGAGTAAAGGGATGCGGTATCA ATTTTAGTTATATATTATGTACAAGGAGTGGAATAAATGGTC	
HoxBa	CNE62	TGACAAGCAAATAGATAATCAACAAGACGAATGACTTCTTTTGTAAGGCTCG GCTCCCGTATTAATTTTCATTTTCTTCTAGAGCAGGTATCGAAAAATAAAGAG CAAGAGAG	AT
HoxBa	CNE63	TTCATTATAACAAAAAAACATTTTCAGTTTGTCTCGTTTAGTACTTAAGACAG ATTAATGATTAACGTGAAGGAAGACTGAACCTTCCCTAATCCATCACGTTCC GCACGAGACCGCAGGATTCTTAAGGTTTGC	AA
HoxBa	CNE64	AGCCGATGGCTGTTTATATCTTAAGAAGCACTTTTATATTCTCTCCATCTCTC CCCCCCATCACTTCCTGTCTGCCTCTGCATTCCTCTCTCTGTCCATCCCTCCCT CTCTCCCTTTCCCCTCGCAGCCTCCCTCTCTCTCTCCTTCTGCTTTTTCTCCC TGTCANCTAGCAGCCCAGCTACTTCTCTTGATTTATAGTGAGACTGGGTGTCA GCGTGCGGCCGCGCCTGCCATTGTGTGTTTGTGGGGGAGTGCTTTTGTTTATG GCATTGCACGCGATAA	T
HoxBa	CNE65	CCGTTATTCATGTCAGTATTCAGTGTCTCTGCTTTTACTCAAACCCTCCANCC T	AA
HoxBa	CNE7	AAGTTTATTAAGTTATCACNGGGTGAACAACCCCATTAATTGGCGTGCTGCT GTTGGTGGTTGGCTGAAATGTACAAAAGGGGAGGGAGCAGGGGGGAAAAAT ACGACAGAGAAATGTAATAACGTGCCTTGGGATATATGGGGGGTCAATTTCT GGAACTATTAGGGTGAAATTGCAAGCGGCGGGGAGAANGAGGCGAGAGATG GGATGCGATATTAATCCGATGGGCGATGCGCTCACTGTTTCCATGCAAGCTG CATTCATCATGGTGGACACGC	AT
HoxBa	CNE8	AATGTTTCTAATTTATTGTACGTAACTTGTCTTTTTTGAGAATGA	AA
HoxBa	CNE9	GCATGTGTTTAAAATATGTGTCTCCTCGGTTGCTCGCTGGAGTTTAAAGGAG AGGGGCGTCTGTGTGGCTGCAACACCAAAAGACTGTCATAAAGCCATAAAC GACGCAACCCCCACACGCAGATGATCGGCACATTCATTCGAGGTATCGTCA	AA
HoxBb	CNE1	GGGGGTCCCCAAAGGACGGAGCTCCTGCGCACAACGCGACCCCGCGGACCC GCGGGGCCTCCTGGTCACCTCCACGCGCCCCCCTTTACTCCAGAGACCTCGC CTCGCCCGATGGCCCCCGCA	P2
HoxBb	CNE10	TTGAGCGAGGTTCGGCAGTGATTGATGCCCACCAGGCGAGTCACTGGGCGGC CAAAG	AA
HoxBb	CNE11	GACATTTTGCACATATTCATAACAACTATATTAAGTGGCAATTTGGCCTTCCT GTAATGACTATTGCATATTGATAGGGCTAATGTGGACCCGGATATTATGCAT TATTCATGAAGATGGGATCATCCGGGGGTCAGGAGAC	AT
HoxBb	CNE12	ACTGACAGCAGAGGGCGCCTTAGACCGGCCTGCAGGATGATGGA	AA
HoxBb	CNE14	AGGTAAGGCGGCCGCATGCAGACGCGCTGATGTGTCTTTATGATTTACGAGC TTGACTCGAGGCTGCTCGGTTCAAGCAGAGTTCATAAAGCTGCATGCCT	AA
HoxBb	CNE15	not available	P2
HoxBb	CNE16	GCTATTCATGTCAGCACACTATTGAGTTGATAAGATGGATCATTTGGTCTCCT CCCT	V
HoxBb	CNE17	AAAGCTGCCATGTGACATTTTGGCTGCCGGAGCCTCGTCACAAGGTCAAAGG GCCTCTCTGTGTGCGTGTCATGGAGAGCAGC	V
HoxBb	CNE19	GTTTGTGTTCTTATCAGATGGATGAGCGCGTTTGATTGAAGCGTTTGTCATGG AAATGCGAGCCGCTTGATGGACGAGCGCGTCGGACGTGACAGAGAAGCCGG TGGGCCGCGCTATTGGCCCTCCGCGGATCACGTGGTGTGTCAGGAAGGTGAT ATGAGGGGTAGAGGGCACTTTTTACAGCTTTGACATACTACCTCGGGGTTTG GGCGATTAATGACTCCTNTGGTGTGAGCA	V
HoxBb	CNE2	GTGTGTCAGGAATTTATGGTTTTATTGCGTGTGTAAATTCCGGAAATCCTC	AA
HoxBb	CNE20	CCAAGCACTGAGTTGCTCAGTGTCACTGTTTGCACTATTTATTTGACAGGGTA TTTT	AA
HoxBb	CNE21	CTTGCATAATGTGACAAGAGGAATCTTTCACAACTCACAGGGTGGTTCCTGG TGACCTATTCACCTCCAAATACCTCCAGGGGCTCTAAACAACAAACTGCAGG GTGCATGCAGGTCTCCATTTTTGTTAT	AT
HoxBb	CNE22	TTTCAATATTTGCACTGGGTGTTGGGTTTGTGCACAAACAGCCACACGCTCAT ACTGGCNCGGCAGCACCGGCCTGTCGGCGGCCTTCTTTGCTAACGCCCGAGG TTCACGCAGAGTTCAGCCACAAATGAATGCAGAGTCAATCTGTCAGGACACC AGGGTGGCT	T
HoxBb	CNE23	TCAATACGGGTCATGGTGGCACATTGGACAAGAAGGAAGCGGCCGGCGCTG AAGACAAGTGTGGGATTTTATTNCGGAGGTGTGATCCCTGCCATCTCTGTGG GTAATACCTCAACGCTCCCAGCGGGCTTTCCATCTGGACGTGTCAGAGTCGG CGGCTGTGGCCGACGCCGCCGAGACAGATGTTCACCTAGGGTGGGGGGA	T
HoxBb	CNE24	GCTGACTTTGAGGTCAACCCAGTGCTGTTTGTACCCCATGCTGCTCGCAGTTG TTGCTGGCTCTTTTACAAGTCTATTGCCGGGCTCACTTCCTNGGGTTTCTACC AGTTCCTCTGTCGCACGTCACCCGGCAGGGACCCCCTGCTTCAAAGGACCAG CCGGGTCCAGTATTCATGGGAGAACCGGGGGGGTTTC	T
HoxBb	CNE25	GCGCCTTCCCTTCTTCCTTTGTTGTGGAACTGAAGCTCCGTTTAACTCCAGAC CCTTCCCCCATCTGTGTCAGGCTCTAATGGATGTGCAAGCCTCAAGGCCCGG CCTCTCGGATCCCCCCAGCGGCCGCATCCTTTCTCCGCACCGTCAGTGCGGTA ACTGGGAGAGACAGGTGACTTGGCCATGGTTAACGTCTTAATACTCGCTGGG ATTGCCTCCAGGATCCGAGCGGCGAGCCC	AA
HoxBb	CNE26	CAGAACTCAATCAGTGTCACGTCACATGGTTCAAAGCGGCCAATTCGCTGGC GCCAGCAGGTTGAGAAATGCAAGGTGACAGTTCATTTTACCACTGACAGTGT	AA

		GTCGGTCTCGCTTGCTGGTTCCCCAGTCAATTCCCAGCGAGGAGTGAAGGTT GCCGGAGACAGACAGCAGCACCCCCCCCACTTCCAACACTCCCCCATTCTCTC TCTCCTTCTTCCTCAACGTCATTCCCTCATCGACCTCGCTCTGCTTTGTATTCC AATTAAGTAGACGGGGAGGGAGGCGGCAGGATGGCAAGTGCGGCCCAGATA CGTTCTCCTGGCGTCCTTCCAGTATCTCAATCCC	
HoxBb	CNE27	ATTCATGGATTCTTCTGGAGCAACCCGCCTCTAGGGACAATGTCTTGGTGGT GAAAAATAGGCTGCTCTCATGCATACATACATCACNTGTAATGGAGGTATGT CATCACTCAGTGAGCGAGCTGCAATCGCTATGCTTAGCGGACACCTCAAATA ATGCTGGATTAATATTTGCACATGGCTCATATATTCTA	AA
HoxBb	CNE28	CCCCTTAGCACTGACGTGTGTGAGCGATGGGTGGGTTGGGAATACAAGTATT GTGAGGAAATACGAAAACAAACTGCTTCTTTTTATCTCCTCTGGCAGCACAA ACATTTGGCGGCGACCCTCNGAAGTGTCGGACGGAGCA	AA
HoxBb	CNE3	AAAAGGGCTGCACTTTATGAAAATTTACAACTTTGCCGTTGCAAGTTTTACG ACTGGGGCGGCTCGGTGATTGGCTAGCGGGGATCATGTGCGCGGCTAGAATG	T
HoxBb	CNE4	GCGCGCAACAAGACCCCCCCACCCCCCCCACCCCAAACGGGCGCAGCCGTG CACGCGCGGACGCCGGTGATTGGCCAGAGGTGGATCAGATGGTACATTTCCC CTCTGTGTTCAGTGGCACCTCACAACCACCCCCCTCCTCTCAGCAAAAGGCA AATCTCCGATAAAGTAATGGCAAAAGCACTTGGACTATAAAAGACAACAAA TCATTTTCCTCCGGAGTGCGCCCCGTTGTCCACCCA	V
HoxBb	CNE5	TGCAATATTTTGTTCAGGAAAAATTGTATATAGAAAATAAAATGCCTTTTGA TGTATAAC	AA
HoxBb	CNE6	CCACAAGATGGCGCCCTCACCCCGCGCGGCTTTGCCGGTGCGCGCGGCGGGG AGGCACCATTGACTGGAGCCTAACGACCATTATTTCGTCATCATTTGTAAGC GGCGAGCATGAATTACCTCTTGAAGTCATCAGTGAGGATTTACGACTGGTCA ACAAAGGCACGTGATTCCCGAGCGCGCTCCCATATTTGGCGGCATACCTGGC AAAGTACAGTAGGGCTTCATTGCTTATAATTCATGATTGCGTCCATAAATCGT GCAAGCACACAAGGATTATAGCGACAAAGATCTACAAATCGAGGCTTTAAA AAAGCAAGCAA	V
HoxBb	CNE7	TGAAGCGCAGCACCTTGCAGCATGCTATTGTTATAGACAACAAAAAGTGTTC TGTGGTTC	AA
HoxBb	CNE8	GATGTCAACTTTGTTTCCTTCTCGCCAACAAGCTTTTTGTATGTAGCTACACG TGAACTCATAGCGCTTGGAAATAAAACTTTCTCTC	AT
HoxBb	CNE9	ATCCCTCTTTGCAAAGATTGTAAAGGGCCTATATGACATACAACAGTACGCG GCAAAGTCACGTGAGCTCCATAAAGTTGCTTTTATGGTTTTGGGGAGTAGAC AATGTTCAATATAATTCACGGGGGCTTGAATGAGAGTGACGGCTTAACGGCG TG	V
HoxCa	CNE1	CTGTAAAGAGCAACTCGTTTGCAGAGTGTGAAGTTCATTCTTGTCCGTCTAG ACTCCGATTACGACCCTGTTTGTCTGCGTTTTTTACCTTGACCTTGGACTTCCG TGCAGCTGCCGCCTATGCGCCCTTCGTGCTTGTGTTTTTTTTTTTCTTAAGACA AAAGAGCTTTAGTCCCTCTCC	AT
HoxCa	CNE10	GTAAACGGAATTAATATCTGTTGCATGTGTACTCAACGCTAATTTAACATGA CCAGAAACAGGCAAATAAAGCAAGCGCGCCTTCATTATCACGTTAATGCAAC TTTAACTGCGACAATAAAACTTTGTGCTGTGCCACAAA	AT
HoxCa	CNE11	TGCGGTGCGCCATCTGTCCCGTCGAACGAGTAAAGTGTAGCAGTGAGCGGTG TATTTGCGGCATGATATTAATGTGAATTTACTTCAAGTCCTTAAATGCGGGCG GCCCACTTTTCAT	AA
HoxCa	CNE12	ATGTTCCGTGTCGTAGTTGTTTGTCCAAATGATGAATTCTTTTGATTTGGTCC CGCCTTTGCATTCCTTTCTTTGCGCTTGTAAGGATTGCTCCATTGTCTTGTATA GTCCTAATGATATGAATGTATCTGCGTACGGTCTTCAATAAACACAATTATTG CAACGT	AA
HoxCa	CNE13	GGTACTCTACGAGCCTATAACGGTGTAGCTAAACATGCAACAGCTCTGTAAA GTCACTAATGGGGTTACACCCCGGTGCTCCA	P2
HoxCa	CNE14	TTTGACTTTGAGACTGACCAAATGTGTTTATGGCGGCCTGTCTCCGAGTCTCA AAGGTCAGCGAGCTTCACGGTGCCCTTCGCAGAGAACTTGTGGTCACGTGG	T
HoxCa	CNE15	GCTTGCAAATGTTCGCAGCCTCATGCAGCCAAGTTTAACTGGCCGACTTACA TCCAATTAAAGTTATCTGGGCCATGAGGGGGCTGTAACGCGATTACACTGGC TTTCC	AT
HoxCa	CNE19	GGATTTAGTGTTACTGGAGCCATTGCGAGTTTTATGGTGGACGTATAGAAGC GAGTTTTAGTGGTCAAATTCAAGTGCTGTGTCTGGACAGCTGTCTGCTTTCTT TGTACTTGCAGGATGCATCACGTGAGCCCACCTTTTGTCACCAAGGGCACTA AAGAAGCTAATGCCATTGCCACCTTTTGC	AT
HoxCa	CNE2	AGGACCTGTTAGCGTATCACGTGCCACATCCTTAATGAAGTGGACGCGCGAG GGGCACCCGTGTGCGCATGCGCCGGGTGTAAAAATGTAGTTGGAAATGAGG TGTCCGTCTTGGAGAAGTCGACTTTTAAAAAGCATCGGCAGCCCCTGAT	AT
HoxCa	CNE20	GGATTTAGTGTTACTGGAGCCATTGCGAGTTTTATGGTGGACGTATAGAAGC GAGTTTTAGTGGTCAAATTCAAGTGCTGTGTCTGGACAGCTGTCTGCTTTCTT TGTACTTGCAGGATGCATCACGTGAGCCCACCTTTTGTCACCAAGGGCACTA AAGAAGCTAATGCCATTG	V
HoxCa	CNE21	TATTGGACAATAATTCAAAAGTCTGCTGCAGGATTTAGGAAGTGTCCAATAA TCATAAACAACGGGCCATTATATTAATTCCACCGTGTGATTGTATCTGTTAGT CTATAAAAAAGTGGAAGACTGCCCTCAACAGTACAATGAAAAGTTTACAGC GACTCAGGAAAGAACAAACAAAAGCATCTTTTCAAACCACTCGTTAATTTAT T	P

HoxCa	CNE22	AACCCCCTGGATGTAGCAAGCCCAACCACTCCTACATGCAACTCTTCCAAAT TGATATATGACAATATCTACTTTCGATCACGTGTCCGCCCGGCGACTTAGAC GGATTGCGCGTCATCTCGCCTTCCCCAATTTTTCTGTAGTGCTGCAGCTCGCA TCCAAAACATCTTTATCGACGGAGCGCGAAAG	V
HoxCa	CNE23	AAGCTGTAAAATGGGGGAAGAGGTGAAGGGAACTAGCCCGCTGTTTTGTCG CTCTCGTTTTATGTGCACTTTTATAAGCATTCAAAAGGATTTTATTATAAGCC CACAAAAGCTCTTTCTTGCAATGAGAACAATTTTTACGATGCTGAGCCTCTTC AAAAAATGACCATGTTATACACACTCCGTGCTGAGAGGCTGTGTGATTTTTTT TTTTATAAAGAGACAGCGCTATTGTGACAAATGGTAACTTTGATCGTGAACT TGCATTGGGCGTTTTATGAAGGGATTTTCTTTCTTCTTGTTGTGCCCACCGTA CATGTGTAAAACCCAATTGGGCAGAACATTGCTGTTGGCGTCTTGTGAAATG TCTTTT	V
HoxCa	CNE24	TTACCTGTTTGATAAGAAATTATCGGTCCTGGGAGGTGCGCAAAGGCAAATG TGGCCAAATGCATTTCAACANGGTGCCAAGAGTGTGTAAATTGTTGAGTTCG GCTCCAGAGGAATGCGAGGGTACATAATGCAGCGAAATGCTTTTTTGCTCTC GTCAATTTAAAAGAGATAAACACGTGCTTAGATGCTGTCACTTGCTCTTATTT ACTCCAAACGGTTGGCATATTTTACGCTCTT	P
HoxCa	CNE25	not available	AA
HoxCa	CNE26	not available	AT
HoxCa	CNE27	TAGAATAAAATCTAGTCTTTGAACTTCAATAATGTCAAGGCCTTCACCTTTAA CCCGCTGCCATAAAGCAGAATCAAGGACCCCCGACTCTCACAAGCGCCCTCC CTCCGTCCCTCCTACCCTCAGCCCGGATGTGAGTGTGTTGTGTGTGCGCGTGC GTGGTGGAGCCGGCCAATGAGAATGAGTGGGAACAGAAGGTGGAATGTGTG AGCAGGCTTCAACTTTGACCCCTCGCCTTCCAAACTACAGCAGGTAACGTAC GTGCCCTACTTT	AT
HoxCa	CNE28	TTTGAGCCAGGCAGAGATTGATGACACCGTCAAATAGTGCCAAGCTCAG	AA
HoxCa	CNE29	TTTTATGGTCCTTTTTATGGTTATAGGGCTGAAAAAGAGAAGGGAGAAGAAA AAAAAAGGAGAGCGAGGTCTCATAAACGGAGCAGGAGTCGCGAGCGCGTCA TCTTTATGGCCACCCTTTCCTTTTGTCAAGTGGCA	P2
HoxCa	CNE3	TTTCATTTGAGAGCCGAGCAAGTCTCTCAAAACTGTGAAAATGCGGAACACG GACTCCATATTTGTTTTATATGACTGTACATATATAAATATATAATATTTAAA ACGATATCT	AT
HoxCa	CNE31	CTAGGTGTCGCTCTTGCTCCACGCAAGACACCTCTCCCTTCCCAACTTGACCG CGCTGACGTCACGGCAGTCTGGAGCATCAAGGCCACTTCCACGCCGCTATTG GTCTTGAAAGTCACATGACCACGTCTCCAAGCATCCATAATTATGTTGCTGAT ATATTTTTTGCGCCCCCCTCCTCCCCAAAAGATGTCAGCATCCCACTGTCCTT ATTGTGTCGG	V
HoxCa	CNE32	ATAACCTTGATTTAAATATTATCCAGGTGACCACAGTAAGTCAAGGTCATAA AAATCTAATGTCAGCCTCTCCCCGGAAAGCGTTGGGGGTGGATTTTATGATC TGCAAATATAATGTGGCGCAGCCGTAAAGATGCGCTTTAAAAGGTGTGTG	V
HoxCa	CNE33	AGCTGAAGCGTGGTTTAGGTAGTTTCATGTTGTTGGGGTTGGCTTCCTGACTC GGCAACAAGAAACTGCCTTGATTACGTCAGTTCGTCTTCATCAAGGGCAC	V
HoxCa	CNE34	CAGTCTGGAATCCAAAGTCACTTACATTCACTCGGCTCAATGGGCGGGCGTC TTCTCGTGGAATGCAAAACTCACAAAGACTTTTACGGCGC	T
HoxCa	CNE35	GTTTTGATGTCAAGCTCGGGACTTGGATGCTGGCTTATTAACGACACGTTTAT CAACAATCAAATGGTCGTCGTTAAAATTTATTGGTAGGCATTAAAACATGAA CGGCCACTATGTGGAACATGGCCGCGCTCGTGGGCTGCCTGTCAACAATGTC GAGACCCTCTTCT	V
HoxCa	CNE36	TTCTTTAGTCCGATTAAAGCGCAAGGCTGTCCCGCGGTAAGTGGCGCATTGA TCCGCCGTGCAAAATTTCCTGGGGTAAATACAGTCACGTGACGCAGGGGCAG CCAATAGGAGGCAGCGGAGCTCGGAGAAATAATTACCTGCCGTGATTGTTCT ATGGGCAGATAAAAAAAAAAAAAGTACGCATACAGGCCATATAATAATCGG ATGCATGTAAACGAGTCCGGGAAGCTTCGAC	V
HoxCa	CNE37	ATAAAAGACAAACGCTACCGGGCTAATAATACTGGGGCTATAAAAACCTGA CGGGGTTGTAAACATGCCTCATGCAACCGCAGGCAGGAGTACAATGAATTT	AA
HoxCa	CNE38	GTTGGCCGAGTTTGGCCGAGGACCTGTACAGGCTTGAACTGGGCCCACTTTA TGGGCCAATAAAGTGACTATACAAGCGCATTGAGTATTTTACGATGTTCTTC ACAAGTTAAGAGGACTGCATGGGAAGCAAATCGGTCGTTTTGG	AA
HoxCa	CNE39	not available	V
HoxCa	CNE4	CTGTGTAGTGCAATGCCTTTTTATTATTTAAATTCATATTATGGCCCATGTGT AAACTCGAGTATAAGAGTATGTTCTCCTGTTAATTTTTATTTTTGGTCCAAGG AGAATTGTGTCACGCATCTGCAATGATTGTAAAGCATTAAAGTCAGTGAGTA CTACACACGGATGCAC	T
HoxCa	CNE40	TATTTGTCTCAACTGCAGAGTAACAAAGCAAACTCGAACGACTGGCTAGACG TCTGGCCTAAATGACTTTATGGTTTTAATGGACGCAGCGGCAGGACTCGTTC AAAG	V
HoxCa	CNE41	TCAACCTGGCAGCCTCTGCTGAATTGCTTCTTGTCCCTTGCCTCATAAAAAAA AAAGAAGGCAAGGCAAAATGACCGCCGTG	P
HoxCa	CNE42	CAGTGATTGGCAATCTGCATATTTATAGAGTGTGAAAGGTCTATAAATTCAC CCAGCGCGCCTTTCATATTTTACGAGGATGTGTATAAAACGTTAGTGGCCCC ACAAAAAGCACGCTGGCATATTTGGAGAGAGCGGGCGAGCGAGCGAGCCAG TGTGAGCGTTGAAGTGGGACTCGCGAGTGCTGATTGCAGCTAAACATCCAGT	T

		GTAAAATTTTATGAACTCTTCCACGCTAGGAGATTTATTCTCGGTGGCTTTTA CGGGGTCATTGAGTGGCACCTACTATTCATAGGTCGTGTTTATATTT	
HoxCa	CNE43	GAGCCATCCGATGAGAGCAAAGGCTCAGAGCGACTGAGCTGCCGTGGATCG TCCGAAGGGGCAGCGTTCCTCCAGTTTCCGGTTTTGTCTGTCTGCGCGTGTGT AACGCGCCGTGTTGTGCGACTTGGGAGCCCCAAACAAAAGGCTGCCATCA	T
HoxCa	CNE45	ACGCGCTTTTTTATTTTTTAACCAGGAAATTGCAATCGTTTGAAATCATTTGA TGCTACCTAAACCTGGGGAGAATTGCAATAGAGATGTCTCAATATAGAGGTT TCATACGTTTATTATTATTATTATTATTATTATTATTATTTGAAAGTGTTATAC ACCCCCCACCCCACCCCCTATCCTCACACCCTCGCGCGCGCGCAACGCACCG ATCTTGTTTTCTTAAGTTATGTACATGTAGCATTTAGTCAATACTTGTTGAAG ACAATTGTTGGAAGCCAAGCAATGAGGAGGTTGTTAGCACGCATGCAGGCA CTACCTAAACTTGCGAGCCGCTCGCTCGCACGCTGCCTAATTTGCAAAAGTC ACACTTGCAGCCACTCGAGTCGAGACGAGACGAGACGCGTGGCACGGCTCG CCACGAACCACAAACGAGTTCCCCAACAACTCAAACTGCCTACAAGAAAAC GCACAGTGACTGCGCGTAGAGAGCTGCAACACTCTTTCCTTTCAATTTCAAT GTGCTGGTGAAATAGATAGACGAGGTTCTTTTGCTTGTTTCCTTGCTTTTTTG ATTTTCTTTTGTTTGTTTGTTTGTTTTGCACTGAATATAACCTTTGTGTATGTC TTGTTATATATTGATGTTAACAATTACGCTCGTTTCATGGTGTTTCTCCTGTGA AAGACAACTTTGTGTCCCGCCATTGTTGATGTCAAATCACAAATAAAGAGGA TGCCTCACAATCATTTCTTCCCTCTTTTCTATT	AT
HoxCa	CNE47	ACCAGAGGGCTGACATTGTCAGCACTAAACTTTAGTGTACCTGTTGTAAACT TTATTGGTGGTGTAAAGG	AA
HoxCa	CNE48	AAAGATGGGTGCAAACATTCCTGGTTGTTAAAGAGAGAAATTTACAGCTGAG TAATAAAAGTTTACGACTCACGGCTAGCCGTGATTGGCTGCGGCGAGCCACG TGGCTCACGCTCTATGAACATGAACTTTATGCTGTTGTCTATTCCCCGCTCCT TCCTTTGAATCGCCAGCAGAGCAGGCTGCAGTCGTGTAAAAGCTCCCC	V
HoxCa	CNE49	GTAGGACCCTACAGTGCCCCCTAGCGACCAGCCGTTTTGCAGGGGGGAAGTG GGCGGGGCAGCGGTTGTTTTGAGCT	P
HoxCa	CNE5	TTAGCCTTGACCATGACTATGCAGTCACTTTGACCTTGACATGCAGGCGGCA GGTTGTGCATAATAAATAGGGGGCTGGACACTGATGCAGGCGGTTTGTTTGT GGGCAAACAAGGTGTTATTTCCCCGTCTAGACAGACGAGCTTAATGACACNC CATAAATACATAAAAAAGGGGAAGTCGGCCGAAATGCTCTGACCGCATTTCT GACCGTCTCTCGGGGCCTCGCATTAACATTT	V
HoxCa	CNE50	ATACAGGTCCATTGAAGTCGTACATCTGTCCCCATCACCACATTTAGAACTCC CGACAGGTCCGAAGTTTTTGCCCACTAATGGTCCTAATTTATGACGTGTGGAT CAACATGACGCTCGCTGGGTACTGGCTTCCTGAGTGCATAACGAGCACAAAG TCGAACACTTCCGGCCGTAATTGGGCCAGCCTGCCACGCTAAAGTGCCTATA GATGCCTTTACATTGTTA	AA
HoxCa	CNE51	GCCTGCTGGGCCGCAATAAAAGCGTACAAAGTGAGGGCAAGGCTGACCTTT CTAATTGTGCTGCATTCCAGTCTCACT	P
HoxCa	CNE52	ACGGGGAAGGGAAGGGAAGGAAGGTTTTATGGCAAGTCACTTCCTCGTCTTT GTTGTCTCTTTTGTGGGC	AA
HoxCa	CNE53	CCATTATTTGCCACGAAGATAATCATCAGTTTGCTTGCTCTCTTG	P2
HoxCa	CNE54	AGGCAGCCTTCTCGCTTTGCTGCATAGCTAGTTTTGGGCGAGCTGGGGGAAC GTCACTTGTTTTATTGCCCTCACACAAGCCAGACCATAAACGCATTCTTACCA CCTCCCCCTGTTCACTGCAAAAAGCTTTGGAAATTGCTCACACGTCCAACAA TTCATCGCGCCGTGGAAAC	AA
HoxCa	CNE55	GTGTTCATCGGCATTTTGCTCAAATGCTCCAGCTCTGGGCTACAAGCTACAGT TTCTGCCGCTCTGCCGTCTCGTCATTGGTGGAGCCGCGTCAGCTGACTTTGTC ATATTGTCTGCGAGCGAGCCAAGGCTCTGCTCTAACGCTCGACTTTGTGCGTC TTAAGCGGAGATTGCAAATTAATAATACCCAATCGTCAAAATAGCGGGGAG GCACCCGAGTGGAGCTGCGATCGGGGGTGGGGGGGTACACGAGGAGAAAGA AAGGGGGGAGAAAAAAAACCCGAGAGAAGACGAGCTCGCAGAAAGGTGTG GTTGGCTGACACGCGCGGGAGGGAGATTTAGTGGGAAGCGAGATTGAAAAG CGAGCGATAACGAAGTGGGGAGAGTTGCCGAGGAGGGTTACGAGGCAAGGT	V
HoxCa	CNE56	GTGAGTTTTACTAGCGGCAAATAAATTCAGGAAATGGGCCCCGTTTTACGAT TTGTCTCACCAAGAGAGTAAGCTACTTTTTATGGCCCCATAAAACATGACTG TATCTCCTTGACTTGTAGATGCGGCGGCCAGGGCCTTTTGCATGCATGTGCAA ACAAACA	V
HoxCa	CNE57	TTTTTTTTTTTTTTTTGAAAGTTCCTGTTTTGTCATAATTGTGTATCATGGAAT AAAATATTTGTAAATAAATGTTGGCATTTCGTGAGGTGTTT	AA
HoxCa	CNE58	GCGCGCGTGACCGTCCGCGGCCGTCCTGTCCCCTTCCCAACAGCAGATGACG CTCTTGTTCTACTGTGCTGCCATTTTTTTGGGCGAAAGGCTGCTCGCCTTGCT CCCCCGCACTCACCAGCCGCAGCAGCTCAAAGTTTACTGGACACGTTTTTGC TATTCATCAATATCCCCATTGAGTATCAACGCCAATTTATGAGTGGCCAACAT GCGCACGTGATCCCATACAAATAGGCCATATTTGGGCAGCGCGCATGGGGAT CAAGGATTTNAAAAAAAAAGATACCAAAGGCTAGTCCACCGATAAATCTTGG ATCTTTTAATTTTTTTTTTAGGAGAG	V
HoxCa	CNE59	TTTGCTTCTTGAGTTTTATAGGCCAAACGCAGGAAATAATAAAAACTCGCGG CCATAAATTTTATGACAAAGGCATCCATTGCTCGTAAAGCTGCTCTATTACG GCGAAGAGGTGATCTGGGGTCTGATTTATGATGCAGTAATTGGGCCGGAGGA TCAAATTGTGACAAAATTATCAAATCATATTTTCCAGATTCATCCGTAAGTAT	V

		TATTGTCTTTATTGCTTCTATTTTAGCAACTCCTCGGCGCAGCTTTGACATGAT CAGCCTGCGTGTGTGTGTTTGAAGCTGTAGCCTTCAATTGAACTCTGGGCAA CACNAGTGACCTGACCCTAGCCGAGCCTGAACTTTTCACCTCGAAAAGATTT GAATAGGCATG	
HoxCa	CNE6	AAAGGGGTCTGATAACCTCTGGTGGTGACCAGAGGAAAATATCAGGAAGTG CTGC	AA
HoxCa	CNE60	CTGACCTCCGGATGAACTGCGGGGAGCAATGACGTCATGACACCGCAGCGT GACTCTTCCATCAGCAATTATGTGTCGGCCGTTGCAGTTCGGAATCCAATGTA TTGACTCTCGCTCACGCTGATCAATCGGTTTGCGTTGCTAAATGAATTAATTT GCTGCC	AA
HoxCa	CNE61	GTTTAAAGCCCTCATAAAACTTTATCACCCTCGTTTCCAGACAGACTGTTTCC TGTTTTGCCAGGAGAAAGAAGGGAAGCCCTCGTCCACCTGAGATAATACT	AA
HoxCa	CNE62	TATTTTCAAGTTTACAAGGTTACATAATTATGTCAATTGGCCGGAGCAGGCC CTGTGAATGGTGCTTGGGAAGCACGTGGTGTCATTTAAGTGGCTTTTATGGC CCGCAAGAGCTGACAAACCTTCGACATATACACATCATATATAATCCTAAGT GTCCGGCAATCGCAGCTGCTGG	V
HoxCa	CNE63	CGCTTTCACTTGCAACGTGTGTGCTGTTTTTGTTCCTTTCTCTGACCTGTGCAG CTGACACGGTCATTATTTAAAACGCAGAGGCCACTTCCGGATGACTTT	AA
HoxCa	CNE64	CTTTAATTAGCCCTCCAATGAATAAATGAACGAGGCAAATAATCCGCCCGAG GGCCTCATCTTTTTNACTTGTACATTATAACAAGTTATTGAACTGGCGGTGCA CAATCTGAAAGCCATTTGTGCGGGGAAGGAATTCATTGGTGTCCCTCCATCA ATAATGCTTGGCAGTGAACTATTGGAACGCAGGCGGGGTGAAGCGCGGGTC AGCCTGTCTCCTCGTATTAAAATG	V
HoxCa	CNE65	TAATCAATGCCATCGATTCCCAATTTGNGTTATATGTTTGCCGTTTTTAATAA CCCCTCCCCAAACAGGGGCCACATGTCCGATATGACATTTGTCATCGCGGCC TTTTCTCATTAGCTTCCCGAGTGTCAGCAAGAAAGGTTAGCGAGCGGTCACT ATGCTAATAGCAGGCCGAAAGAAACGCTCNGGAACAATGCAGCCAACTTAT TTTTCATCCCCGCTGTCACGCCGACCCTTTGTGCATGTTGGCGGCTGAATATT CATA	AT
HoxCa	CNE66	ATTAGTTGTCTATATGTACCCTGTAGAACCGAATTTGCGTGCACTTGAACACA ATCGCAAATACGTCTCTACGGGAATACATGGGC	AT
HoxCa	CNE67	TACCGTTCGACACCTCGTCGAACACGTTGCGAACAGGCGCTCGTTAAAACCC GACTGATTTACCGCTTTTATTGCGCAATAAAAGCCTCCCAAGACGAGCGCTA ATAGAAA	AA
HoxCa	CNE68	GCGCGCCATGGCGACTGGAGGTATGCTCAACAACTCTGGTTCCTCATCCGGG CACCTGCAGCTCGGCTGCGATTGGCCGTCATGGTCACGTGGTAAAAGTAACT TTACAGGGCTGCTTGCAAGTAGGAGGGCTTTATGGAGCAGAAAAACGACAA AGCTAGAAAAATTATTTTGCACTCCAGAAATTA	V
HoxCa	CNE69	TTTGAAATGGCACAATTGAGGCGGATTTACGACTCGGCGTTGGTAATTACAC CCACCATAAATTTTATAGCCGAGGTATACTCAGGGCAGCCGTATCACCATGT GGCTTCTGCTCTTATGTATTGCACCATGGAAAGGGGGAGAGAGAGAGAGAG ACAGCGACGGGACATATATGAATAAGAGGGGAAGGCAGCGAAAGAAAACG AACAAAAAAAGGAATCCAGCTTTGTAATCTTGCATTAATAATTGATCTTGCT TGGCTGCAGCGAGTGTCCCCAC	V
HoxCa	CNE7	GCTAGGCCATAATAACCTGCCTTCACCGAATCAAAGAGCTCGCGAAAATAGA AAGAAAGGCCCTTTGTTTGGGGCTATTGCCTCGTCTGGCCACCACCAATGGA TTCCAGAAGGCTAATTCCATTATCGAGGAGACGCTGGACTAGACAGTGACCT TGAC	AT
HoxCa	CNE70	AGAACTGCGTATTTATAGATGATGGTCTATTTTGTTTTGTTTTTTTTTNGTAGCC ATCGTGTGCGGTTTCCTCCAAGTCCAGCGTTATATAAAATGCATGTTATATAG CATGGATTCCTGCGAATACCGATGCAGTCTTTTTTATACGTGATACAGGGTTT CGTTTATTTGAACGCGTTTCAAATGCTCTTTGAATATGTTTGGAAAGAAGGCC GAGCAAAAAATGGCGATACATTGAATACAAATATTATTGTGGGCCATTCTTT TGCCCTTTGCAGGTGAAATGCACAATCTATTTATTTCAGACAACACCGGAAA AATATCAATATAATTATTTTTGGTATTAGGGGAGAATTAAAAAAAAAAAAAA AGCTGAAACACACCATTCATCTTGAACTTGTATGTTTAGGATAAAAAAAAAC ATGTGACAATTACCTCGTGTATTTCAGAAACAATTATTTATTTATTTAATTAT TTATTTAATTATTTATTTATTGTTTACAGTTTGGTTTCGTGGTGTTGGTGTTGT GTTTAGGACCTTCCAAAAAAATGGATTATTGTGAAATGCAATAAACGGATTT AGTGTACCTGTGCTAATCATATTGATCAATAAAACAGTGAGTGTCTTCTAGTT TTCGANACGTTTGCTGACTCGTTGGCTTATTGTAGGC	V
HoxCa	CNE71	TCGCCTCCGTGTCGCCACGCTGAAAAGCGACTTGATGGAAGTTTTTTTTTTGC ATTGGAGAAGCCCGAGTTCACACCGAGGACACATTTTATGTCCCATTAGGTT TATTTCAAGCTGGATGGGCGAGCCACCC	AT
HoxCa	CNE73	CTCACTTGTGAAGTGTGGTGAGCAAACTGCAGTCCTTTCATTTGGCGCTGCAT TGTTTACAGCAATATAGGGCACCGGAACCATCAATCTTCCCTGCGTGCCAGG ACAGCGCCATTTATATGCAGCGTTTGAACCGGCAGCATCCTCTCTGCCCATTA ACG	AT
HoxCa	CNE74	GCGTGCGTGTGCGCGAGGAAGAAAAATGATTTATTGGGACTAGGGCGGTTC AAGCAGAGTTCAGGGAGAATAGTGGTGCAGGCGCAGCAGCGACATGATGGC CAACTTGCTGAGCTTGTGTATCAGACGAGCTTGCTTGCAATGCAG	AT
HoxCa	CNE75	TGTGGGGGCGGGGGGGGGCGTGTATGTGTTCTGTTAAGTGTTTGATTAAAAG CACGTTATGGCNATTTAATCAGGGATATTGCATGCAGCGCTCGGG	AA

HoxCa	CNE76	GAATTTGATAAATTACTGAATTTGTGACATGAATAGGCTAGTATAAATCGCT TTTCCTCTCGGGAAAGGGAGGGGGNCATAGTGGATTCACAAAAGGCT	T
HoxCa	CNE77	CGTTTGGTTGTGTGTGTGTTTGTTTTTTTTCTCCTTGGTCAGCAAAGCGAGCTA CTAGTTTAGTGGCCAAGATCAATGGCTTCGCTACCAAGACAAGCGGAGAGG GAG	T
HoxCa	CNE8	ACAAATCAAGTAATTAGCCGCATTGGGAAAGTCTTTGCTGTTTGAATGAGTG AAGTGTCCCCTGTTATCCTGTAAAAATCTTATAAAAGAATTGTGAGAGAGCT AACAATCCCACCAGAGTTAGTCAGTGGGGCTTTATTGGACCTATCTGTTCAA CTTGTTCGTGTGAGCCGTACTATAGAATATTACGTGAAATACACCCTCTTAAA GTTACCTTATAGAGTGCCCCCTATCCCGTTTTTTTTTAATTTTTTTTTTACAGA GGCTTCAACCTGCTGACCTCGTCTTACTGTAAATAGGGCAATGAGCATCATT CGTTTCTTTAAAACAAGCTTGTAAAGTTGGGACATCACGACTTTATGTGACA AGGAATTCAACCATAATTTGCTACAAGATGCCACATGCATGCAACCCCTCAT TGAAGGGAGCAGTTTTCATGAGATCGAGCATTTAAGAAAAAAAAATAAAAAT AAAAAATAAACTTTTGCCAGCTTGTTAAAGTATTAAACGAGGCGACAGCCAG CTTTTTTAATGTGTGACATTTAATGTCGGACAAAAAAAAGGCAGACCTGGCT TGTTTTGCTCTCACGATTCAAGTCAAGAAAGAGACAAGATTGCTGAATTTTA TAAAAAGCTTTATGGTCTTAAAAAAAAATAACATATAAAGGTGGGTTATAGA GAGAAGTGGATGTTTTTGTTTTGTGTGAGCCGGAGCGAGTGGGTGGGACCCC GTGGAGCAGAAGTTGGCCTAGCCACGTGGTCTGAGGCTGACCAATCAGACA GCCCCCCAGGTTTAACTATGTTAAATGTCAGATTGCAATAAAGTAGAAGCTG TTGGTCGGGCCCGCGGAA	AT
HoxCa	CNE9	TGCGTGGTGCGCATACGCCACTAGGGGTTGCCCTAAAAATAGTCCAGGCAAG CGCCATGTGCTT	T
HoxDa	CNE1	CAAAGCGTGTGGGGAGGAGGAGGAGGAGGAGGAGGATGAGGAGGAGGGAG AAAAGTGGTCCTGCGTGGGTATGGCTGCGGGAGCTCAATCCGCGTGCAAGTG TCTCTGCGATGTCTCTAAAGCTTTTTATAAGCAACACGCAATAAATGGAAGA AATGGAGGAGAGGGCGCTCAAAATGTTCCTGGCATCCTCTCCGTGCACGTCG ACAATATGTCATTTATGCCCCCCCCCCTAAAAAAAAAAAAAAATGTTCTCTGC GCAGGCTGACGCGGCTCCAATGATCATTTATTGTAACAGGTTTATAAGCAAA TAAATAGGAGAGGGCTGTCAGTTGATGATGGAGGCGGCCTTCGGTCAGACG AATAATATCCAAGAGGAGAGCACGAGAGAAGAGGAGTGAGTGAGGAGGAG GAGTGAGGGCGCCCCTCACAGCTCTTTGGCTCTTTGGCCTCCCGGCTGCTCAC ACTGACAGTCTGATTAATAAAATGAGCGCTGCAACATTTCCCCCCTTCATTTA GGAATATTATTATGCTTTGATTCTCTATTGTTTCCCTC	V
HoxDa	CNE11	TGCTTCCTCGAGTCCGTTCTTTCTTTATAGGTGGCATGTATAATCACAGGCAT ACCATAAAACATTTATTGATACACATAAATATGAATATATATTAAGCCGGTA GCGCTATTATTATTATTCTCTTTCCAATGAATATTTTACTGCCGCTTTTTCCAA CTCTGTNTGAAATAAAACAGAAAGAAAGACAATATGAATCCAGGTTGATTA GGAAAAACAGCGTCTGTTTTTTCTTTCTTTCCTTTTACAACGTGTCATTACATC AATTTTACGAGGATTTGCCAAGTCTATAAAGCTTTTTTTTACTGAGTGCATGA GTA	V
HoxDa	CNE12	AGGCTTGATTTACTGGCAGAATTAGTAAATATGATCACGTGATCTGTGTAAC CAATCCGTGCTGACGCAGGCCAGCAAAAATACTATGATTGTTCATAGAGGGG AGCTTCCCCTTAACTGCGAGTCATTTTAT	V
HoxDa	CNE13	ATGCAAATAGCCCCCAGTGCTGATTTATAACTGTGTGAGCTGACTGCGTAAA ACGGGATGTTTTACTGACCCATAAAAGTGTCCTGG	V
HoxDa	CNE14	GGGCTTATAATGACTTTAGAGCAGGAAGAAATAAAGCCATAGAGATGGCTA GACGTCTGGTCTAAATGAGTTTATTGGCCCTGGCAGTCAGTAATTACAACGA NGCCCTTTAAAAGCTCCTTTTCCCTCTCTGGGCACTTCTTTTCTTTGTGCGTGT GTTGCTGACAATCAAAATGTGGACACAGGGGTGTCTGGCCGATGGATTGGTG CACAGGAAATCACTCCCACATCTTAATGGTCACGTTAATTCATCAGCTGTGCT TCTCACTGATGGG	V
HoxDa	CNE15	ACCAGGACGAGTTACTGCAGTGGAAATTCCTTTTTTTGTGAGCGGACAATTT ACAACTTGGCACTAGACGCCTTTTGTGAGCGTCTATGGCCTGTCATTGGATGC CACTGGTCATGTGTGCGAGGTAGCAAACGTCTTCATGGCTCTTTTGCTCATTT CCCATGC	V
HoxDa	CNE16	TGGAGATGGGGCCGAGTGAACAAAGACAGTATTTCAGTGAGGCCTGACAGG CAGCTGCGAAAGTATTTACAACCTCACTGCAATGCGGCAAAATGGA	V
HoxDa	CNE17	TTCGCATGCTCCGCAGCCTCACAGCCATTACCGAGGAAGTGCAACTTTGGAG GGTTATTTATGGGCCCGCGCGGTCTCGTGAGTGGCTGCCAGGAAACACGTGA CGCCATTAAAGTTTGCTTTATGGCCCGTGGCTTGACAAGACAAAATATAATT CGCATTGTGTATNGGCGAACGGCTCCAGATGGGCCGGAGAGGAGCGTCCGT CCGGCCGGCCGGGTGGCCAGCCGGCCGCGTGGGGCTGCGCGTCTGCCGCGG AGTTGGAGGGAGCTGGCGGTGGTGCTGATGTGTGTGCGCGCNCGCGCGNGT GTGTGTGTTGCTGGCACTGCGGTAAGACGGCGTTTTATCATTTGCTGCTTGGG CGTATGATGTCACGCACAC	V
HoxDa	CNE18	AAAACGCTGACTGTATGATTCTTTTGTGCTGCATTTCGGCTTTGACCCGTCTC AAGTGCTGGCTGGCCTGGGCAAGGTGAAAGACAGGTCAGCCTGTCTCAC	V
HoxDa	CNE19	AGCAGATGAAGGGATGAGATTGGTGTAACTTTTGGATGACCCCCCCCCCCTC CTTGACAAAGACAGTCGCCATCAATTGTCCTCACT	AT

HoxDa	CNE2	CCCCGTGCGCACTTTGCATATCACGTGAGCACCCGTTGACAAATCACCAATT AGCTAGCCGTTAGTCTTGAGAAGTTGCTAAGGCGCTTCGGGAGCAGCGTGCG CGCGGCCAGCTCC	AT
HoxDa	CNE21	between CNE23 & CNE24; between *hoxd4a* & *hoxd3a*	AA
HoxDa	CNE22	GCTTTGGAAATCACCCGGTGTGATGTTTATGGGAGCGTTTGAAAGGCCTTGC CAATTACACGAGTCGTAAATTTGATAGCTGGAGGGGGGGCGAAACAGGCCG CGAAACAAGAGAGCC	V
HoxDa	CNE23	TTTTGCATGGACTGAAGTCGTTTTTTGAGCTTTTTTAATTTAACAA	AT
HoxDa	CNE24	GAGGCGAAGGTCGGTTGGCAGGTTCATCCAGGGGACACGCTCAAGCCGCAA GGGATGACCCCGACACCCCCTGACCCGGCAGGACGAGCTGCAT	V
HoxDa	CNE25	TTATTTTTGTACATGCTTTCTACTTTGTATTTGTGCACCGGGATGTGAGATGT ATATTGGAAACAAATAAAGATTTCTACTGATTATGCATTTAAATGA	AA
HoxDa	CNE26	GTGTGATTGGACTACACGTGTTGGGGTGAGCTTCCGCGCTGACCTACATGTG CGAGCACCAATAAATGGCAGCCATAGGCGCCGTGCCACTTCCAATGACAGCC GCATGCGCCAAGGCTGCTTCCCCGCTGCCGCTTCTCCCCATCCGGCCATATTT GAGTCTTCTCCCCTCTCCCTCCGCCTTCTTTCCTTCTTGACAACAACAGCACG TCGTCTTTTTTTGGGGGGTAGCATTCGCAGAAAAATCCACGCCGACAGCATC TCTCCGGAAGCTTGCTTTCCCCCCCTTTCTGGTCTCAGATCACGTGACCGGCC CGATAATTAATGCAGCTCCCCGTTGCCCCCCTCGAAGCTCGGCGTCGTCATTA ATCGCGAGGACTCTATCTAGACTTGAAAAACTGAAAAGATCTTCCTAAG	V
HoxDa	CNE27	AAATGGAACTTTGACACTTATGCATGCGTTGTAAAACACCCCCAGTAATTCC TGAAAGGTTGCGAGGGAGGGAGGGGAGGAGGGGAGTAGGTGTGGGGGGGA CGCGAAAACTGTAAATCTTTCACTTTTATGACCCTGTGAACATATGCTG	V
HoxDa	CNE28	TTTGTCCAAATGGATCGGTGAGGTGGAGGAAGAACAAATCACGTGGACAAA TATGCTCGCATCTTGCAGGGCAGTGCCTTTATTTGTCATCATAAAGCTCCCCC CTCCTCCCCAAACCAAATATTCTACTGTTTAGCTCCCAAATCTACACGGGGGC CATACACTCGAATTCCTGCTATTTCATTTGCTGATTGATTGCTAGCAAGGTTT TTTTTCCCCTTCTCTCCTCTCCTTGCTGTGCAGAAATAAGGATATACCCTATGT TGGCTTTTTATTGGTAGCTGAGTCCTCGCTGCACTCTTTCGGGAATACTGTCT CCACCGGTGTATGGAAATGTCTGCAAAAGAGCAAGATCGAGTTTAGGACAG CATTGTGTGCACACAAAAGGCGAGCTGAAGGATTTATTCGACAAGGAGTCGC GGAAAGCGAGCTCGCGTTGGCTTTTCTTTTTTTTTTTTTTGTTACTGTAAATGA TGAATTGACGCACAGGTAAGCAGCTGCATGGCGCAAAGCAACCTTTAAGCA GGTTAATTTGATGCAAGCGTGATTTATTCCACGCGGAGGACGCTCGTTAAAG CGCTCGCG	V
HoxDa	CNE29	AAGTAGAAGGCACTTTTCTTCTTTTTGACTCCCGGATCGTAAATCACATTAAT TTGTTGTCTTATCGTCACAATGGCGTTCGGCGTGATTTATGTGGTTTGAGCCG CGCGGCTTTGTTAAAAAAAAAAAAAAAA	AT
HoxDa	CNE3	CATGGGCCCTAAAGACAGTGGGGGTGCCATAACA	AT
HoxDa	CNE30	GAGATAATAAGTGTCTACATGGAGTCGGCTCCATTGTGTGGAGCAAACTTAC CGTAGCAGCGCGTAGGCGAGGAACACTGCGCCTTGAATAGGCGAGCATGTC TGCTACTCGTCGCCCAGTG	AT
HoxDa	CNE31	CTCGTTGTTCATTAAACGTCGACTTAAAAGCCGCATGACCCAAAAGGTCAGG TAAAAACTGCTTATGACTGCTAAATATCGTCTTTGTTTAAAGTGCACATGTAG GACTTTTTT	V
HoxDa	CNE32	TTTTACAGGTGGAGTTGGGTACGATGCTGAATGTTTAAAGGACGTCACACAC ATTTGCGCAGATTTTGGTGAGTAACA	T
HoxDa	CNE33	GCACCATATGAAATTCTTTCGCTTGCCTTCCACCTACGCTTGGTTGTTTTGCA CTTGTAGTGGAGTGCAAAGTGTGTGTGCGAGCGTGTGTGGTGTGTTTTTGCTC TGCGTGTACGACGCGCACATTTGCGCTCATAAATTTTGCCGCGGCGCTCTCAC AGTGACAGGTGCAATTGTTATGCAGCACGACGAGCGGCACCACCAAC	AT
HoxDa	CNE34	ACGACGCGCTTTGTCAGAGCTTGTGATTAGTGAAAAGGCTATTGTGTGCTGC CGTGTCAAATGCTGATTCAGCGTCACACCAAAGCCCCACAGAGAGTCTTTTT GAGTCCAAGTT	T
HoxDa	CNE35	GTGTGCCAATATTACGTCGCGCTCTTTTTATTACCTCATTCTTCGAATAAGTC CGAGCGAGTCGTTGTGCATTATTNTCATCACAGTATTGATATGACGATTATTG TGTATTTTTTTTGTTAAATAATCCTCTCACGTTATTGGACACGTGGGGGTGGA TGTCTGAGAGTTCTTGTTTTTTTTCTAAATATATACATATATAAAGTGCAATG CGCATTTTGGACTGAATATTTGACTTGAAGTAGATGATGAACGTAGAAACTC ATCAGTAGTGAGAAGGACTTTTAGTCCCGTTTTTTGTCTCTTTTGTCAGGATC TCATCCNCATCATCGAATTTATTTATTTTCATGATTATTGAATAATCGTTTTCG GGCTCTTATCTATTTATTATTTAATGATTGCACATTATTTATCGTTTATAAGCG AGTATTTATACGTGTTTTTCTTTTTGTAACATTGGACAGATCAGTGCACTTGA TGGCAAAACAGGGTTGTTGAAAAAAAAATCTAATTGTATATCGGGGACAAA TGGGGCATTTTATAGTGTATAAGCACCATTACGTGTCCATGGTTAATCTTGTG TGGAATCTTAAGCAAAATCGTAGACTATCGAAGTATCAAATATTGTACACAC CAGCATGTCATTTTGGTTTTTCTACTCATTGTCTCCTGTCGTTACTCATCAAAT TACCAGACCCCGACGATATTGAAATGTTCTATTTTATTGTGTTGGACATTTTG TAAATAAAGTGCTCAAACTTG	AT
HoxDa	CNE36	GGAGAGCCATTTCAAAGCTCATTAATCAAAAACTCGTCGCTTTCGACTCCTC GCTGATTCCTCCTCCTCGCCAACAAAGCATCTGCATTCCAAATCCAAGAAAC ACACTTCAAACCGTGCGGGCTAATCTTGAATGGGGGG	AT

HoxDa	CNE39	TGTCTGCGAGCCTGAAGTGTGGGGATTTACGGCAGCTCTTACGTGCGGCTGA TAGATATAATTGTAAGTGAAATAAATGACTCTCTCCCGTGGCTATGATTTAG GAGGGCCTGACCAAGACGCTCCTCTCCTCCTCCCCGGTGAGAGATAACTCGC TGATGGGAGGAAACCTGAAGTAACTCACACATTAGCCACGCCTGTCCTACAA AAGCCTCGCTAGACGCCTTTGCTTAGAATTAGCACTTGCACTCTTTTGGTGAG GCTTATCCAGATTTCCCTTCCGTATGAAATTATGAGTGAATAAATTGCATCTT CATTTATCTTTGGAAATCCCCTCCTGTGCTCGGTGCGATACTCGGCCGGCTCT CCGCGCTTCCCGTAATGACATTAGTCTTCTGCTACACAGCAAAATTGATTGCC CAGGAAAATGAACTTGGCTTCGCTATAAATTACACCTATGGATCATAAGAAA TGTGTTATTTTGTTTAAAATAGTAGCTAATATAATTAGAGCTCTAATTCATGG CAAGCTACAAGTGTGCCTATTACAGTAACAGGCTTTGGAATTTAATCTGTCTT CTGGATAAAGAATGATNTCAGGCCGTTTTCATTTGTAATCTAAGACACATTA TTTCAATAAAGTAGCCTGCAACTTGACGATACGAGGATGATTGCGGAATGAA TAACAATGGTCTTACGTTTCACTCTCAGAGTTAAGAATT	AA
HoxDa	CNE4	GAGAGGAGTTCCCCCCAGCATCCAGGGAGCCAATGGAGGCTCTCCCACGGC CTCACATGAGTTTCTCAGGACGGCTTTTTTTTTTTTTTTTAATGACCGATATTG ATGTATGGTAATTTCTTGGCCGGCGGATCACATGACACAATTACCTCAAGAA TCGATCAAGATGTATAGCGAGCCCGCTCGCGGCTCTTTGCGCGCGTCTGGGC TGCCGAGCCAGATGGGCCAAGTTTAACGGGCACAAGCCGCGGAGAAGCCTC CTCCTCCGCCTCCTGGTCCCGCCGCCGCAGAC	V
HoxDa	CNE40	CCAATCACGTGAAAAAAAAAAAAATCACAAGCACCGCCACTGATGCCATTGG AGAGACGCTGCCTGCTTGATTGTCTTTGAGCGAGCCAACAAAA	V
HoxDa	CNE41	CTATATATACCCTGTAGAACCGAATTTGTGTGATGCAAGCCCAGTCACAGAT TCGATTCTAGGGGAGTATATGGT	V
HoxDa	CNE42	not available	V
HoxDa	CNE43	not available	V
HoxDa	CNE44	not available	V
HoxDa	CNE45	not available	V
HoxDa	CNE46	not available	T
HoxDa	CNE47	not available	V
HoxDa	CNE48	not available	AT
HoxDa	CNE5	CTTTATAAGCATGCAAAAGCGTTTTATATCCCAATAAGATGTTCTTTACGGCT GTAAAGGTATTTACAATGGGAAACTGTTAAG	V
HoxDa	CNE6	CTTTATTTCCCCTCCTTTCACACACTGGATTGTTTACTAAACCTTGAACCGTCT AGACACAATCATAAAAGCAGCTGTAAACGGCTAAATAAGGGGCTATTGGAC TTCGGAGTCCCCCCCCCCC	V
HoxDa	CNE7	between CNE27 & CNE28; between *hoxd4a* & *hoxd3a*	P
HoxDa	CNE8	TGAAGGCCACTTTCAAAGCTCATTGGTGGGCTCGTCATGTGGTCGGCGCGCA CGCTGACT	V
HoxDa	CNE9	CATCGTAAAAGCGAGAATAAAACACAAACAAGCCTCCGCAGTCGTAATGTTT AATGAGAG	AT
HoxDb	CNE1	GGTTGCGCCGGTCTGCGAGTCACCCGAAGCCGGGAGGAAGGACGAAGTGGA CGCGCCTTATTGCGGAGGTTG	AT
HoxDb	CNE10	GTTTTCATATTTGGC	P
HoxDb	CNE11	GGTCGTGCAAGTTAATGTCTAAACCGGCTTAGAGGACCGCTCTCAGCCCGGC TTTGACCCGGCTGCAGGCCCGGGGGCGAAGGGGTTAGCCGGCGGTCAGCCT GTCTAAGATGAAAAAGTGAAACGACCCCCGAGAGAAAGAAAAGCCA	AA
HoxDb	CNE12	CCTGTCCAGCTCGCATAGGACTAATGGTAGGTCCGGTGAAGCCTCGGTCAAC GCTTGTGAGTGACCCTGCAGCGACCCGCTCGGCAGCTAGGCTGCATTCTGCC CGCACAAGAATGCTC	AA
HoxDb	CNE14	TGGTGAACGGGACACCTCGGACTGACCCCTACCCGCAACAAAGCCAACTACT AGTCACTGTCTGGGGGCCTTTTTGTCCCATTCTTGTGTGCGTGCCCTCTGCGG AGCGGTCACGTTTTGTTTTGCGCCCTTTGATTGCGGCCACAAGGCCAAATATG GCCCCCTCGTCCCCTCAATATGGGCCCGATTAATTTTTCAGCTCAGGGGCCCA TCTGTGGGGGTCGATCCCGTACGTCAGCT	AA
HoxDb	CNE15	GAGAGCGGCAAAAAAGCGACGCCTCCCATTGGTTGAGCTAGCGGGTCACAT GGTCCATACGGCCGTCCGTCTATTTGACAGCCGCGGGGTGGCTTG	AA
HoxDb	CNE16	TGTTTTGTCCATTTGGGAGCACCATTTCTTGTGGTGAGTTATTTATGATCGCA GTGAGTGTCAGGCAGAATTACAGCCGCTATAAACTTTTATGGCTCTGCCGCA GCTTGTGCGCGCACGTGAGTGTG	V
HoxDb	CNE17	not available	AA
HoxDb	CNE18	not available	AA
HoxDb	CNE19	ATGTCTCTTGCAGCAGATGCTTTCGTNTCACAAATGGCGGGTAAAGTGTTGC AGCTT	T
HoxDb	CNE2	not available	AA
HoxDb	CNE20	TTTGTTTTTACAGCTGCTGAGCCTTGGCCCGAGAGTGCAACCTTGTACCAGCC GCTTAAGGGTGTG	T
HoxDb	CNE26	TAACAATTATATATGTTTTGTCTTGTGCAGAGGATCAGATTCTGCTTTCGGAC TGCGCCTCATCCCTCGCTGTGCAGGTACATTT	T

HoxDb	CNE27	GCCTATAAATACCCTGTAGAACCGAATGTGTGTGGACTCTGCTCGGTCACAG ATTGGGTTCTAGGGGAGTCTATGGGCGGCG	V
HoxDb	CNE3	ACCATGTGATCGGCGCCATAACCAATAGGCGGCTGAGGAGAAGGTAAGGGC AGGAAAAAAATTACTACCATTTTAGGGGAGGATGGAAGCCACTATACTGAC CCATCTAAGGTTTTACAGTCGCCTTTAAGTGTCTTTTTTATTGAAATAATAAA ATAAACATGTCGCTAACAAATTATTATTCC	AA
HoxDb	CNE4	TGTTTGCGTGTGCACGTGTGTCTGTATGCAAGTGTGTGTGTGTGTGTGCAGGA ATGTGGCCATATGTAAGGAAGTGTGTGAGAGACTTTTATTGGCCTATAAACG TCCTCCAAGCCTCGACGACAATGTATAAAACTTTATTGTCCATTAAAATAATT GGAGCGACTTTTGCTGTTATTTAGCA	AA
HoxDb	CNE6	GGCCGCCATATAAAGCGCATCCTCCTGCAGGGTCATAAAGCAACAGCATGCA TTTGGAGAATAAACGTGACTTTCCCCCCCCCACTCTACCCTCGTACACACACAC ACGCACGCATATATAGAAAAGACGTGGGTTTTCTTTCAAAAGAGGGACAGC GAGAGAGAGAGAACAAGAAAAAGCCATAAAGTAATTAAAGCCAGACAGTCTCG GTGACTATATGGTTTTATTGGACATCCTTTTCCTGCCTCGACTGCACCTCGAA ATGTTAA	AA
HoxDb	CNE7	GATAGAGGAGAAACAAAGACAGTATTTCACTGGTGATGCTCACTGGAGCTC AGCTGGCACTGACAGTATTTACAACCTACTGCAATGCGGCAG	V
HoxDb	CNE8	ACACACACACACNCNCGCACGCNCACACGCACGTGTGTTTTCACAGCGCCCA CTTTATGGCTTTATTGCTCCCCATTTACGGAAATCAAAGCAGGAAACAAATG TTCCAGCATTCCA	AA
HoxDb	CNE9	AATTTATGGATCCGCGTGGGAGACGTGAAAAAAAAAAAAAGCGGGTGGGGGT GTCGGAGGGGGAGAACACGTGATAGCAATAAAGTGTGTTTTATTGCCAGTGG CGTGACAGGCCCCAAAATATAACTCAGACTGCAGCCGACAGGCCACCACAT CGCATCAGCGG	V

REFERENCES CITED

Aguirre, W. E., S. E. Contreras, K. M. Carlson, A. J. Jagla, and L. Arellano. 2016. "Evolutionary diversification of body form and the axial skeleton in the Gasterosteoidei: the sticklebacks and their closest relatives." *Evolutionary Ecology Research* 17 (3):373-393.

Ahn, D. G., and G. Gibson. 1999. "Axial variation in the threespine stickleback: genetic and environmental factors." *Evolution and development* 1 (2):100-112.

Akimenko, M. A., M. Ekker, J. Wegner, W. Lin, and M. Westerfield. 1994. "Combinatorial expression of three zebrafish genes related to distal-less: part of a homeobox gene code for the head." *The Journal of Neuroscience* 14 (6):3475-3486.

Albalat, R., and C. Canestro. 2016. "Evolution by gene loss." *Nature Reviews Genetics* 17 (7):379-391. doi: 10.1038/nrg.2016.39.

Albertson, R. C., W. Cresko, H. W. Detrich, 3rd, and J. H. Postlethwait. 2009. "Evolutionary mutant models for human disease." *Trends in Genetics* 25 (2):74-81. doi: 10.1016/j.tig.2008.11.006.

Alexander, T., C. Nolte, and R. Krumlauf. 2009. "Hox genes and segmentation of the hindbrain and axial skeleton." *Annual Review of Cell and Developmental* 25:431-456. doi: 10.1146/annurev.cellbio.042308.113423.

Allen, L. G. , M. M. Yoklavich, G. M. Cailliet, and M. H. Horn. 2006 "Bays and estuaries." In *The Ecology of Marine Fishes: California and Adjacent Waters*, edited by D. J. Pondella and M. H. Horn, 119–148. Berkeley, California, USA: University of California Press.

Amores, A., J. Catchen, A. Ferrara, Q. Fontenot, and J. H. Postlethwait. 2011. "Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication." *Genetics* 188 (4):799-808.

Amores, A., A. Force, Y. L. Yan, L. Joly, C. Amemiya, and A. Fritz. 1998. "Zebrafish hox clusters and vertebrate genome evolution." *Science* 282 (5394):1711-1714. doi: 10.1126/science.282.5394.1711.

Amores, Angel, Tohru Suzuki, Yi-Lin Yan, Jordan Pomeroy, Amy Singer, Chris Amemiya, and John H Postlethwait. 2004. "Developmental roles of pufferfish Hox clusters and genome evolution in ray-fin fish." *Genome Research* 14 (1):1-10.

Anders, S., P. T. Pyl, and W. Huber. 2015. "HTSeq--a Python framework to work with high-throughput sequencing data." *Bioinformatics* 31 (2):166-169. doi: 10.1093/bioinformatics/btu638.

Aristotle. 2014. "History of Animals." In *Complete Works of Aristotle, Volume 1: The Revised Oxford Translation*, 774. Princeton, New Jersey, USA: Princeton University Press.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, and Z. A. Lewis. 2008. "Rapid SNP discovery and genetic mapping using sequenced RAD markers." *PLoS ONE* 3 (10):e3376. doi: 10.1371/journal.pone.0003376.

Bejerano, G., M. Pheasant, I. Makunin, S. Stephen, W. J. Kent, J. S. Mattick, and D. Haussler. 2004. "Ultraconserved elements in the human genome." *Science* 304 (5675):1321-1325. doi: 10.1126/science.1098119.

Benedetti, I., D. Sassi, and A. Stefanelli. 1991. "Mauthner neurons in syngnathid bony fishes." *Acta Embryologiae et Morphologiae Experimentalis* 12 (1):75-76.

Benjamini, Y., and Y. Hochberg. 1995. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal Statistical Society* 57 (1):289-300.

Berthelsen, J., V. Zappavigna, E. Ferretti, F. Mavilio, and F. Blasi. 1998. "The novel homeoprotein Prep1 modulates Pbx-Hox protein cooperativity." *The EMBO Journal* 17 (5):1434-1445. doi: 10.1093/emboj/17.5.1434.

Betancur-R, Ricardo, Richard E. Broughton, Edward O. Wiley, Kent Carpenter, J. Andrés López, Chenhong Li, Nancy I. Holcroft, Dahiana Arcila, Millicent Sanciangco, James C. Cureton Ii, Feifei Zhang, Thaddaeus Buser, Matthew A. Campbell, Jesus A. Ballesteros, Adela Roa-Varon, Stuart Willis, W. Calvin Borden, Thaine Rowley, Paulette C. Reneau, Daniel J. Hough, Guoqing Lu, Terry Grande, Gloria Arratia, and Guillermo Ortí. 2013. "The tree of life and a new classification of bony fishes." *PLoS currents* 5 (2013). doi: 10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288.

Blanc, G., and K. H. Wolfe. 2004. "Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution." *Plant Cell* 16 (7):1679–1691.

Blomme, T., K. Vandepoele, S. De Bodt, C. Simillion, S. Maere, and Y. Van de Peer. 2006. "The gain and loss of genes during 600 million years of vertebrate evolution." *Genome Biology* 7 (5):R43. doi: 10.1186/gb-2006-7-5-r43.

Boettcher, M., and M. T. McManus. 2015. "Choosing the right tool for the job: RNAi, TALEN, or CRISPR." *Molecular Cell* 58 (4):575-585. doi: 10.1016/j.molcel.2015.04.028.

Boetzer, M., C. V. Henkel, H. J. Jansen, D. Butler, and W. Pirovano. 2011. "Scaffolding pre-assembled contigs using SSPACE." *Bioinformatics* 27. doi: 10.1093/bioinformatics/btq683.

Bowne, P. S. 1994. "Systematics and morphology of the Gasterosteiformes." In *Evolutionary biology of the Threespine Stickleback*, 28-60. Oxford, UK: Oxford University Press.

Braasch, Ingo, Andrew R Gehrke, Jeramiah J Smith, Kazuhiko Kawasaki, Tereza Manousaki, Jeremy Pasquier, Angel Amores, Thomas Desvignes, Peter Batzel, Julian Catchen, Aaron M Berlin, Michael S Campbell, Daniel Barrell, Kyle J Martin, John F Mulley, Vydianathan Ravi, Alison P Lee, Tetsuya Nakamura, Domitille Chalopin, Shaohua Fan, Dustin Wcisel, Cristian Cañestro, Jason Sydes, Felix E G Beaudry, Yi Sun, Jana Hertel, Michael J Beam, Mario Fasold, Mikio Ishiyama, Jeremy Johnson, Steffi Kehr, Marcia Lara, John H Letaw, Gary W Litman, Ronda T Litman, Masato Mikami, Tatsuya Ota, Nil Ratan Saha, Louise Williams, Peter F Stadler, Han Wang, John S Taylor, Quenton Fontenot, Allyse Ferrara, Stephen M J Searle, Bronwen Aken, Mark Yandell, Igor Schneider, Jeffrey A Yoder, Jean-Nicolas Volff, Axel Meyer, Chris T Amemiya, Byrappa Venkatesh, Peter W H Holland, Yann Guiguen, Julien Bobe, Neil H Shubin, Federica Di Palma, Jessica Alföldi, Kerstin Lindblad-Toh, and John H Postlethwait. 2015. "A new model army: Emerging fish models to study the genomics of vertebrate Evo-Devo." *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 324 (4):316-341.

Braasch, Ingo, Andrew R Gehrke, Jeramiah J Smith, Kazuhiko Kawasaki, Tereza Manousaki, Jeremy Pasquier, Angel Amores, Thomas Desvignes, Peter Batzel, Julian Catchen, Aaron M Berlin, Michael S Campbell, Daniel Barrell, Kyle J Martin, John F Mulley, Vydianathan Ravi, Alison P Lee, Tetsuya Nakamura, Domitille Chalopin, Shaohua Fan, Dustin Wcisel, Cristian Cañestro, Jason Sydes, Felix E. G Beaudry, Yi Sun, Jana Hertel, Michael J Beam, Mario Fasold, Mikio Ishiyama, Jeremy Johnson, Steffi Kehr, Marcia Lara, John H Letaw, Gary W Litman, Ronda T Litman, Masato Mikami, Tatsuya Ota, Nil Ratan Saha, Louise Williams, Peter F Stadler, Han Wang, John S Taylor, Quenton Fontenot, Allyse Ferrara, Stephen M. J Searle, Bronwen Aken, Mark Yandell, Igor Schneider, Jeffrey A Yoder, Jean-Nicolas Volff, Axel Meyer, Chris T Amemiya, Byrappa Venkatesh, Peter W. H Holland, Yann Guiguen, Julien Bobe, Neil H Shubin, Federica Di Palma, Jessica Alföldi, Kerstin Lindblad-Toh, and John H Postlethwait. 2016. "The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons." *Nature Genetics* 48 (4):427-437. doi: 10.1038/ng.3526.

Britten, R. J., and E. H. Davidson. 1971. "Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty." *The Quarterly Review of Biology* 46 (2):111-138.

Britz, R., and K. W. Conway. 2009. "Osteology of Paedocypris, a miniature and highly developmentally truncated fish (Teleostei: Ostariophysi: Cyprinidae)." *Journal of Morphology* 270 (4):389-412. doi: 10.1002/jmor.10698.

Brouwer, A., D. Berge, R. Wiegerinck, and F. Meijlink. 2003. "The OAR/aristaless domain of the homeodomain protein Cart1 has an attenuating role in vivo." *Mechanisms of Development* 120 (2):241-252. doi: 10.1016/S0925-4773(02)00416-1.

Brown, L. Y., and S. A. Brown. 2004. "Alanine tracts: the expanding story of human illness and trinucleotide repeats." *Trends in Genetics* 20 (1):51-58. doi: 10.1016/j.tig.2003.11.002.

Brown, Robin. 2010. "Craniofacial development in pipefish: A morphological and molecular analysis." Undergraduate Biology Honors Thesis, University of Oregon.

Brudno, M., C. B. Do, G. M. Cooper, M. F. Kim, and E. Davydov. 2003. "LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA." *Genome Research* 13 (4):721-731. doi: 10.1101/gr.926603.

Brudno, M., S. Malde, A. Poliakov, C. B. Do, O. Couronne, and I. Dubchak. 2003. "Glocal alignment: finding rearrangements during alignment." *Bioinformatics* 19:i54-i62. doi: 10.1093/bioinformatics/btg1005.

Bruner, Emilano, and Valerio Bartolino. 2008. "Morphological Variation in the Seahorse Vertebral System." *International Journal of Morphology* 26 (2).

Brunet, F. G., H. Roest Crollius, M. Paris, J. M. Aury, P. Gibert, O. Jaillon, V. Laudet, and M. Robinson-Rechavi. 2006. "Gene loss and evolutionary rates following whole-genome duplication in teleost fishes." *Molecular Biology and Evolution* 23 (9):1808-1816. doi: 10.1093/molbev/msl049.

Burglin, T. R., and M. Affolter. 2016. "Homeodomain proteins: an update." *Chromosoma* 125 (3):497-521. doi: 10.1007/s00412-015-0543-8.

BBMap version 35.

Cantarel, B. L., I. Korf, S. M. Robb, G. Parra, E. Ross, and B. Moore. 2008. "MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes." *Genome Research* 18 (1):188-196. doi: 10.1101/gr.6743907.

Carcupino, M. 2002. "Functional significance of the male brood pouch in the reproductive strategies of pipefishes and seahorses: a morphological and ultrastructural comparative study on three anatomically different pouches." *Journal of Fish Biology* 61 (6):1465-1480. doi: 10.1111/j.1095-8649.2002.tb02490.x.

Carroll, R. 1988. *Vertebrate Paleontology and Evolution*. New York City, New York, USA: W. H. Freeman and Company.

Carroll, S. B. 1995. "Homeotic genes and the evolution of arthropods and chordates." *Nature* 376 (6540):479-485. doi: 10.1038/376479a0.

Carroll, S. B. 2008. "Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution." *Cell* 134 (1):25-36. doi: 10.1016/j.cell.2008.06.030.

Carroll, Sean B, Jennifer K Grenier, and Scott D Weatherbee. 2013. *From DNA to diversity: molecular genetics and the evolution of animal design*. Malden, Massachusetts, USA: Blackwell Pub.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. "Stacks: an analysis tool set for population genomics." *Molecular Ecology* 22 (11):3124-3140. doi: 10.1111/mec.12354.

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. "Stacks: building and genotyping Loci de novo from short-read sequences." *G3 (Bethesda)* 1 (3):171-82. doi: 10.1534/g3.111.000240.

Catchen, J. M., J. S. Conery, and J. H. Postlethwait. 2009. "Automated identification of conserved synteny after whole-genome duplication." *Genome Research* 19 (8):1497-1505. doi: 10.1101/gr.090480.108.

Chan, S. K., H. D. Ryoo, A. Gould, R. Krumlauf, and R. S. Mann. 1997. "Switching the in vivo specificity of a minimal Hox-responsive element." *Development* 124 (10):2007-2014.

Chan, Y. F., M. E. Marks, F. C. Jones, G. Villarreal, M. D. Shapiro, S. D. Brady, Audrey M Southwick, Devin M Absher, Jane Grimwood, Jeremy Schmutz, Richard M Myers, Dmitri Petrov, Bjarni Jónsson, Dolph Schluter, Michael A Bell, and David M Kingsley. 2010. "Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer." *Science* 327 (5963):302-305. doi: 10.1126/science.1182213.

Chang, C. P., W. F. Shen, S. Rozenfeld, H. J. Lawrence, C. Largman, and M. L. Cleary. 1995. "Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins." *Genes and Development* 9 (6):663-674.

Chen, F., and M. R. Capecchi. 1997. "Targeted mutations in hoxa-9 and hoxb-9 reveal synergistic interactions." *Developmental Biology* 181 (2):186-196. doi: 10.1006/dbio.1996.8440.

Chen, F., and M. R. Capecchi. 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 (2):541-546.

Chen, F., J. Greer, and M. R. Capecchi. 1998. "Analysis of Hoxa7/Hoxb7 mutants suggests periodicity in the generation of the different sets of vertebrae." *Mechanisms of Development* 77 (1):49-57. doi: 10.1016/S0925-4773(98)00126-9.

Chiu, C. H., C. Amemiya, K. Dewar, C. B. Kim, F. H. Ruddle, and G. P. Wagner. 2002. "Molecular evolution of the HoxA cluster in the three major gnathostome lineages." *Proceedings of the National Academy of Sciences* 99 (8):5492-5497.

Cohn, M. J., and C. Tickle. 1999. "Developmental basis of limblessness and axial patterning in snakes." *Nature* 399 (6735):474-479. doi: 10.1038/20944.

Condie, B. G., and M. R. Capecchi. 1994. "Mice with targeted disruptions in the paralogous genes hoxa-3 and hoxd-3 reveal synergistic interactions." *Nature* 370 (6487):304-307. doi: 10.1038/370304a0.

Cresko, W. A., A. Amores, C. Wilson, J. Murphy, M. Currey, P. Phillips, M. A. Bell, C. B. Kimmel, and J. H. Postlethwait. 2004. "Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations." *Proceedings of the National Academy of Sciences of the United States of America* 101 (16):6050-6055. doi: 10.1073/pnas.0308479101.

Cruz, C., S. Maegawa, E. S. Weinberg, S. W. Wilson, I. B. Dawid, and T. Kudoh. 2010. "Induction and patterning of trunk and tail neural ectoderm by the homeobox gene eve1 in zebrafish embryos." *Proceedings of the National Academy of Sciences of the United States of America* 107 (8):3564-3569. doi: 10.1073/pnas.1000389107.

Cunningham, F., M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, Denise Carvalho-Silva, Peter Clapham, Guy Coates, Stephen Fitzgerald, Laurent Gil, Carlos García Girón, Leo Gordon, Thibaut Hourlier, Sarah E Hunt, Sophie H Janacek, Nathan Johnson, Thomas Juettemann, Andreas K Kähäri, Stephen Keenan, Fergal J Martin, Thomas Maurel, William Mclaren, Daniel N Murphy, Rishi Nag, Bert Overduin, Anne Parker, Mateus Patricio, Emily Perry, Miguel Pignatelli, Harpreet Singh Riat, Daniel Sheppard, Kieron Taylor, Anja Thormann, Alessandro Vullo, Steven P Wilder, Amonida Zadissa, Bronwen L Aken, Ewan Birney, Jennifer Harrow, Rhoda Kinsella, Matthieu Muffato, Magali Ruffier, Stephen M J Searle, Giulietta Spudich, Stephen J Trevanion, Andy Yates, Daniel R Zerbino, and Paul Flicek. 2015. "Ensembl 2015." *Nucleic Acids Research* 43 (Database issue):D662-D669. doi: 10.1093/nar/gku1010.

Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2012. "jModelTest 2: more models, new heuristics and parallel computing." *Nature Methods* 9 (8):772. doi: 10.1038/nmeth.2109.

Darwin, Charles. 1859. *On the origin of species by means of natural selection (8th Edition)*. 8th ed. New York City, New York, USA: Mentor Books.

Davis, J. C., and D. A. Petrov. 2004. "Preferential duplication of conserved proteins in eukaryotic genomes." *PLoS Biology* 2 (3):E55. doi: 10.1371/journal.pbio.0020055.

Dawson, Charles E. 1985. *Indo-pacific pipefishes (Red Sea to the Americas)*. Ocean Springs, Mississippi, USA: Gulf Coast Research Laboratory.

De Kumar, B., and R. Krumlauf. 2016. "HOXs and lincRNAs: Two sides of the same coin." *Science advances* 2 (1):e1501402.

de Lussanet, M. H., and M. Muller. 2007. "The smaller your mouth, the longer your snout: predicting the snout length of Syngnathus acus, Centriscus scutatus and other pipette feeders." *Journal of the Royal Society, Interface* 4 (14):561-573. doi: 10.1098/rsif.2006.0201.

Debiais-Thibaud, M., V. Borday-Birraux, I. Germon, F. Bourrat, C. J. Metcalfe, and D. Casane. 2007. "Development of oral and pharyngeal teeth in the medaka (Oryzias latipes): comparison of morphology and expression of eve1 gene." *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution* 308 (6):693-708. doi: 10.1002/jez.b.21183.

Dekker, E. J., M. Pannese, E. Houtzager, E. Boncinelli, and A. Durston. 1993. "Colinearity in the Xenopus laevis Hox-2 complex." *Mechanisms of Development* 40 (1-2):3-12.

Don, E. K., T. A. Jong-Curtain, K. Doggett, T. E. Hall, B. Heng, and A. P. Badrock. 2016. "Genetic basis of hindlimb loss in a naturally occurring vertebrate model." *Biology Open* 5 (3):359-366. doi: 10.1242/bio.016295.

Duboule, Denis, and Pascal Dollé. 1989. "The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes." *The EMBO Journal* 8 (5):1497-1505.

Edgar, R. C. 2004. "MUSCLE: multiple sequence alignment with high accuracy and high throughput." *Nucleic Acids Research* 32 (5):1792-1797. doi: 10.1093/nar/gkh340.

Edwards, S. V., L. Liu, and D. K. Pearl. 2007. "High-resolution species trees without concatenation." *Proceedings of the National Academy of Sciences of the United States of America* 104 (14):5936-5941. doi: 10.1073/pnas.0607004104.

Ekblom, R., and J. B. Wolf. 2014. "A field guide to whole-genome sequencing, assembly and annotation." *Evolutionary Applications* 7 (9):1026-1042. doi: 10.1111/eva.12178.

Elder, Pliny the. 2012. *Natural History*. Edited by H. Rackham. London, UK The Folio Society.

Etter, P. D., S. Bassham, P. A. Hohenlohe, E. A. Johnson, and W. A. Cresko. 2011. "SNP discovery and genotyping for evolutionary genetics using RAD sequencing." *Methods in Molecular Biology* 772:157-178. doi: 10.1007/978-1-61779-228-1_9.

Faircloth, B. C., J. E. McCormack, N. G. Crawford, M. G. Harvey, R. T. Brumfield, and T. C. Glenn. 2012. "Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales." *Systematic Biology* 61 (5):717-726. doi: 10.1093/sysbio/sys004.

Faircloth, B. C., L. Sorenson, F. Santini, and M. E. Alfaro. 2013. "A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs)." *PLoS ONE* 8 (6):e65923. doi: 10.1371/journal.pone.0065923.

Ferretti, E., F. Cambronero, S. Tumpel, E. Longobardi, L. M. Wiedemann, F. Blasi, and R. Krumlauf. 2005. "Hoxb1 enhancer and control of rhombomere 4 expression: complex interplay between PREP1-PBX1-HOXB1 binding sites." *Molecular and Cellular Biology* 25 (19):8541-8552. doi: 10.1128/MCB.25.19.8541-8552.2005.

Ferretti, E., H. Marshall, H. Popperl, M. Maconochie, R. Krumlauf, and F. Blasi. 2000. "Segmental expression of Hoxb2 in r4 requires two separate sites that integrate cooperative interactions between Prep1, Pbx and Hox proteins." *Development* 127 (1):155-166.

Flammang, Brooke E, Lara A Ferry-Graham, Christopher Rinewalt, Daniele Ardizzone, Chante Davis, and Tonatiuh Trejo. 2009. "Prey capture kinematics and four-bar linkages in the bay pipefish, Syngnathus leptorhynchus." *Zoology* 112 (2):86-96.

Flanagan, S. P., J. B. Johnson, E. Rose, and A. G. Jones. 2014. "Sexual selection on female ornaments in the sex-role-reversed Gulf pipefish (Syngnathus scovelli)." *Journal of Evolutionary Biology* 27 (11):2457-2467. doi: 10.1111/jeb.12487.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, and J. Postlethwait. 1999. "Preservation of duplicate genes by complementary, degenerative mutations." *Genetics* 151 (4):1531-1545.

Fraser, G. J., C. D. Hulsey, R. F. Bloomquist, K. Uyesugi, N. R. Manley, and J. T. Streelman. 2009. "An ancient gene network is co-opted for teeth on old and new jaws." *PLoS Biology* 7 (2). doi: 10.1371/journal.pbio.1000031.

Frazer, K. A., L. Pachter, A. Poliakov, E. M. Rubin, and I. Dubchak. 2004. "VISTA: computational tools for comparative genomics." *Nucleic Acids Research* 32:W273-W279. doi: 10.1093/nar/gkh458.

Fricke, R., W. N. Eschmeyer, and R. van der Laan. 2019. "Catalog of Fishes: Genera, Species, References." accessed Janurary 2, 2019. http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp.

Froese, R., and D. Pauly. 2018 "Fishbase." accessed January 2, 2019. http://www.fishbase.org.

Fromental-Ramain, C., X. Warot, N. Messadecq, M. LeMeur, P. Dolle, and P. Chambon. 1996. "Hoxa-13 and Hoxd-13 play a crucial role in the patterning of the limb autopod." *Development* 122 (10):2997-3011.

Gadow, H. F. 1933. *The Evolution of the Vertebral Column*. Cambridge, UK: Cambridge University Press.

Galant, R., and S. B. Carroll. 2002. "Evolution of a transcriptional repression domain in an insect Hox protein." *Nature* 415 (6874):910-913. doi: 10.1038/nature717.

Garcia-Fernandez, J. 2005. "The genesis and evolution of homeobox gene clusters." *Nature Review Genetics* 6 (12):881-892. doi: 10.1038/nrg1723.

Gaunt, S. J. 1988. "Mouse homeobox gene transcripts occupy different but overlapping domains in embryonic germ layers and organs: a comparison of Hox-3.1 and Hox-1.5." *Development* 103 (1):135-144.

Gaunt, Stephen J, Paul T Sharpe, and Denis %J Development Duboule. 1988. "Spatially restricted domains of homeo-gene transcripts in mouse embryos: relation to a segmented body plan." *Development* 104 (Supplement):169-179.

Gavalas, A., M. Davenne, A. Lumsden, P. Chambon, and F. M. Rijli. 1997. "Role of Hoxa-2 in axon pathfinding and rostral hindbrain patterning." *Development* 124 (19):3693-3702.

Gavalas, A., M. Studer, A. Lumsden, F. M. Rijli, R. Krumlauf, and P. Chambon. 1998. "Hoxa1 and Hoxb1 synergize in patterning the hindbrain, cranial nerves and second pharyngeal arch." *Development* 125 (6):1123-1136.

Gehring, W. J., M. Muller, M. Affolter, A. Percival-Smith, M. Billeter, Y. Q. Qian, G. Otting, and K. Wuthrich. 1990. "The structure of the homeodomain and its functional implications." *Trends in Genetics* 6 (10):323-329.

Gehring, Walter J, Markus Affolter, and Thomas Bürglin. 1994. "Homeodomain proteins." *Annual review of biochemistry* 63 (1):487-526.

Gehrke, A. R., and N. H. Shubin. 2016. "Cis-regulatory programs in the development and evolution of vertebrate paired appendages." *Seminars in Cell and Developmental Biology* 57:31-39. doi: 10.1016/j.semcdb.2016.01.015.

Gendron-Maguire, M., M. Mallo, M. Zhang, and T. Gridley. 1993. "Hoxa-2 mutant mice exhibit homeotic transformation of skeletal elements derived from cranial neural crest." *Cell* 75 (7):1317-1331.

Geneious 8.1.9

Ghanem, N., O. Jarinova, A. Amores, Q. Long, G. Hatch, and B. K. Park. 2003. "Regulatory roles of conserved intergenic domains in vertebrate Dlx bigene clusters." *Genome Research* 13 (4):533-543. doi: 10.1101/gr.716103.

Gnerre, S., I. Maccallum, D. Przybylski, F. J. Ribeiro, J. N. Burton, and B. J. Walker. 2011. "High-quality draft assemblies of mammalian genomes from massively parallel sequence data." *Proceedings of the National Academy of Sciences of the United States of America* 108 (4):1513-1518. doi: 10.1073/pnas.1017351108.

Godsave, S., E. J. Dekker, T. Holling, M. Pannese, E. Boncinelli, and A. Durston. 1994. "Expression patterns of Hoxb genes in the Xenopus embryo suggest roles in anteroposterior specification of the hindbrain and in dorsoventral patterning of the mesoderm." *Developmental Biology* 166 (2):465-476. doi: 10.1006/dbio.1994.1330.

Goncalves, I. B., I. Ahnesjo, and C. Kvarnemo. 2015. "Embryo oxygenation in pipefish brood pouches: novel insights." *The Journal of Experimental Biology* 218:1639-1646. doi: 10.1242/jeb.120907.

Goodman, F. R. 2003. "Congenital abnormalities of body patterning: embryology revisited." *Lancet* 362 (9384):651-62. doi: 10.1016/S0140-6736(03)14187-6.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, and I. Amit. 2011. "Full-length transcriptome assembly from RNA-Seq data without a reference genome." *Nature Biotechnology* 29 (7):644-652. doi: 10.1038/nbt.1883.

Graham, Anthony, Nancy Papalopulu, and Robb Krumlauf. 1989. "The murine and Drosophila homeobox gene complexes have common features of organization and expression." *Cell* 57 (3):367-378.

Grammatopoulos, G. A., E. Bell, L. Toole, A. Lumsden, and A. S. Tucker. 2000. "Homeotic transformation of branchial arch identity after Hoxa2 overexpression." *Development* 127 (24):5355-5365.

Guindon, S., J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. 2010. "New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0." *Systematic Biology* 59 (3):307-321. doi: 10.1093/sysbio/syq010.

Guindon, S., and O. Gascuel. 2003. "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood." *Systematic Biology* 52 (5):696-704. doi: 10.1080/10635150390235520.

Haase, D., O. Roth, M. Kalbe, G. Schmiedeskamp, J. P. Scharsack, and P. Rosenstiel. 2013. "Absence of major histocompatibility complex class II mediated immunity in pipefish, Syngnathus typhle: evidence from deep transcriptome sequencing." *Biology Letters* 9 (2):20130044. doi: 10.1098/rsbl.2013.0044.

Haeckel, Ernst. 1866. *Generelle Morphologie der Organismen : allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von Charles Darwin reformirte Descendenz-Theorie*. 2 vols. Berlin, Germany: G. Reimer.

Haeckel, Ernst. 1896. *Die amphorideen und cystoideen; beiträge zur morphologie und phylogenie der echinodermen*. Leipzig, Germany: W. Engelmann.

Hamilton, H., N. Saarman, G. Short, A. B. Sellas, B. Moore, and T. Hoang. 2017. "Molecular phylogeny and patterns of diversification in Syngnathid fishes." *Molecular Phylogenetics and Evolution* 17:388-403.

Handrigan, G. R., and R. J. Wassersug. 2007. "The anuran bauplan: a review of the adaptive, developmental, and genetic underpinnings of frog and tadpole morphology." *Biological Reviews* 82 (1):1-25. doi: 10.1111/j.1469-185X.2006.00001.x.

Hardie, D. C., and P. D. N. Hebert. 2004. "Genome-size evolution in fishes." *Canadian Journal Of Fisheries And Aquatic Sciences* 61 (9):1636-1646. doi: 10.1139/f04-106.

Harlin-Cognato, A., E. A. Hoffman, and A. G. Jones. 2006. "Gene cooption without duplication during the evolution of a male-pregnancy gene in pipefish." *Proceedings of the National Academy of Sciences of the United States of America* 103 (51):19407-19412. doi: 10.1073/pnas.0603000103.

Harmston, N., A. Baresic, and B. Lenhard. 2013. "The mystery of extreme non-coding conservation." *Philosophical Transactions of the Royal Society B: Biological Sciences* 368 (1632):20130021. doi: 10.1098/rstb.2013.0021.

Harris, R. 2007. "Improved pairwise alignment of genomic DNA." *ProQuest*.

He, X., Y. L. Yan, J. K. Eberhart, A. Herpin, T. U. Wagner, M. Schartl, and J. H. Postlethwait. 2011. "miR-196 regulates axial patterning and pectoral appendage initiation." *Developmental Biology* 357 (2):463-77. doi: 10.1016/j.ydbio.2011.07.014.

Herald, E. S. 1959. "From pipefish to seahorse - a study of phylogenetic relationships." *Proceedings of the California Academy of Sciences* 29.

Hoegg, S., J. L. Boore, J. V. Kuehl, and A. Meyer. 2007. "Comparative phylogenomic analyses of teleost fish Hox gene clusters: lessons from the cichlid fish Astatotilapia burtoni." *BMC Genomics* 8:317. doi: 10.1186/1471-2164-8-317.

Hoekstra, H. E., and J. A. Coyne. 2007. "The locus of evolution: evo devo and the genetics of adaptation. ." *Evolution: International Journal of Organic Evolution* 61 (5):995-1016.

Hoffman, E. A., K. B. Mobley, and A. G. Jones. 2006. "Male pregnancy and the evolution of body segmentation in seahorses and pipefishes." *Evolution* 60 (2):404-410. doi: 10.1111/j.0014-3820.2006.tb01117.x.

Hohenlohe, P. A., S. Bassham, M. Currey, and W. A. Cresko. 2012. "Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes." *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences* 367 (1587):395-408. doi: 10.1098/rstb.2011.0245.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. "Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags." *PLoS Genetics* 6 (2):e1000862. doi: 10.1371/journal.pgen.1000862.

Holland, P. W. 2013. "Evolution of homeobox genes." *Wiley Interdisciplinary Reviews: Developmental Biology* 2 (1):31-45. doi: 10.1002/wdev.78.

Holt, C., and M. Yandell. 2011. "MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects." *BMC Bioinformatics* 12:491-491. doi: 10.1186/1471-2105-12-491.

Horan, G. S., R. Ramirez-Solis, M. S. Featherstone, D. J. Wolgemuth, A. Bradley, and R. R. Behringer. 1995. "Compound mutants for the paralogous hoxa-4, hoxb-4, and hoxd-4 genes show more complete homeotic transformations and a dose-dependent increase in the number of vertebrae transformed." *Genes and Development* 9 (13):1667-1677.

Howard, R. K., and J. D. Koehn. 1985. "Population dynamics and feeding ecology of pipefish (Syngnathidae) associated with eelgrass beds of Western Port, Victoria." *Marine and Freshwater Research* 36 (3):361–370.

Humphrey, J. H., and R. R. Dourmashkin. 1969. "The lesions in cell membranes caused by complement." 11:75-115. doi: 10.1016/S0065-2776(08)60478-2.

Hunter, Michael P., and Victoria E. Prince. 2002. "Zebrafish Hox Paralogue Group 2 Genes Function Redundantly as Selector Genes to Pattern the Second Pharyngeal Arch." *Developmental Biology* 247 (2):367-389. doi: 10.1006/dbio.2002.0701.

Huxley, Julian. 1942. *Evolution the modern synthesis*: George Allen and Unwin.

Hwang, W. Y., Y. Fu, D. Reyon, M. L. Maeder, S. Q. Tsai, J. D. Sander, R. T. Peterson, J. R. Yeh, and J. K. Joung. 2013. "Efficient genome editing in zebrafish using a CRISPR-Cas system." *Nature Biotechnology* 31 (3):227-229. doi: 10.1038/nbt.2501.

Iimura, T., N. Denans, and O. Pourquie. 2009. "Establishment of Hox vertebral identities in the embryonic spine precursors." *Current Topics in Developmental Biology* 88:201-234. doi: 10.1016/S0070-2153(09)88007-1.

Infante, C. R., A. G. Mihala, S. Park, J. S. Wang, K. K. Johnson, J. D. Lauderdale, and D. B. Menke. 2015. "Shared Enhancer Activity in the Limbs and Phallus and Functional Divergence of a Limb-Genital cis-Regulatory Element in Snakes." *Developmental Cell* 35 (1):107-119. doi: 10.1016/j.devcel.2015.09.003.

Jackman, W. R., B. W. Draper, and D. W. Stock. 2004. "Fgf signaling is required for zebrafish tooth development." *Developmental Biology* 274 (1):139-157. doi: 10.1016/j.ydbio.2004.07.003.

Jao, L. E., S. R. Wente, and W. Chen. 2013. "Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system." *Proceedings of the National Academy of Sciences of the United States of America* 110 (34):13904-13909. doi: 10.1073/pnas.1308335110.

Johnson, David G., and Colin Patterson. 1993. "Percomorph phylogeny: a survey of acanthomorphs and a new proposal." *Bulletin of Marine Science* 52 (1):554-626.

Johnson, L. S., S. R. Eddy, and E. Portugaly. 2010. "Hidden Markov model speed heuristic and iterative HMM search procedure." *BMC Bioinformatics* 11:431-431. doi: 10.1186/1471-2105-11-431.

JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma, Wageningen, The Netherlands.

Jones, A. G., D. Walker, and J. C. Avise. 2001. "Genetic evidence for extreme polyandry and extraordinary sex-role reversal in a pipefish." *Proceedings: Biological Sciences* 268 (1485):2531-2535. doi: 10.1098/rspb.2001.1841.

Jurka, J., V. V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, and J. Walichiewicz. 2005. "Repbase Update, a database of eukaryotic repetitive elements." *Cytogenetic and Genome Research* 110 (1-4):462-467. doi: 10.1159/000084979.

Kanehisa, M., Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe. 2016. "KEGG as a reference resource for gene and protein annotation." *Nucleic Acids Research* 44 (D1):D457-D462. doi: 10.1093/nar/gkv1070.

Karplus, K., C. Barrett, and R. Hughey. 1998. "Hidden Markov models for detecting remote protein homologies." *Bioinformatics* 14 (10):846-856. doi: 10.1093/bioinformatics/14.10.846.

Katoh, K., and D. M. Standley. 2013. "MAFFT multiple sequence alignment software version 7: improvements in performance and usability." *Molecula Biology and Evolution* 30 (4):772-780. doi: 10.1093/molbev/mst010.

Kawaguchi, M., S. Yasumasu, J. Hiroi, K. Naruse, M. Inoue, and I. Iuchi. 2006. "Evolution of teleostean hatching enzyme genes and their paralogous genes." *Development Genes and Evolution* 216 (12):769-784. doi: 10.1007/s00427-006-0104-5.

Kawahara, R., M. Miya, K. Mabuchi, S. Lavoue, J. G. Inoue, and T. P. Satoh. 2008. "Interrelationships of the 11 gasterosteiform families (sticklebacks, pipefishes, and their relatives): a new perspective based on whole mitogenome sequences from 75 higher teleosts." *Molecular Phylogenetics and Evolution* 46 (1):224-236. doi: 10.1016/j.ympev.2007.07.009.

Kern, Andrew D., and Matthew W. Hahn. 2018. "The neutral theory in light of natural selection." *Molecular biology and evolution* 35 (6):1366-1371.

Kiecker, C., and A. Lumsden. 2005. "Compartments and their boundaries in vertebrate brain development." *Nature Reviews Neuroscience* 6 (7):553-564. doi: 10.1038/nrn1702.

Kijimoto, T., M. Watanabe, K. Fujimura, M. Nakazawa, Y. Murakami, and S. Kuratani. 2005. "cimp1, a novel astacin family metalloproteinase gene from East African cichlids, is differentially expressed between species during growth." *Molecular Biology and Evolution* 22 (8):1649-1660. doi: 10.1093/molbev/msi159.

Kimmel, C. B., C. M. Small, and M. L. Knope. 2017. "A rich diversity of opercle bone shape among teleost fishes." *PLoS ONE* 12 (12):e0188888. doi: 10.1371/journal.pone.0188888.

Kimura, Motoo. 1968. "Evolutionary rate at the molecular level." *Nature* 217 (5129):624-626.

King, M. C., and A. C. Wilson. 1975. "Evolution at two levels in humans and chimpanzees." *Science* 188 (4184):107-116.

Kitazawa, T., K. Fujisawa, N. Narboux-Neme, Y. Arima, Y. Kawamura, T. Inoue, Y. Wada, T. Kohro, H. Aburatani, T. Kodama, K. S. Kim, T. Sato, Y. Uchijima, K. Maeda, S. Miyagawa-Tomita, M. Minoux, F. M. Rijli, G. Levi, Y. Kurihara, and H. Kurihara. 2015. "Distinct effects of Hoxa2 overexpression in cranial neural crest populations reveal that the mammalian hyomandibular-ceratohyal boundary maps within the styloid process." *Developmental Biology* 402 (2):162-174. doi: 10.1016/j.ydbio.2015.04.007.

Knoepfler, P. S., Q. Lu, and M. P. Kamps. 1996. "Pbx-1 Hox heterodimers bind DNA on inseparable half-sites that permit intrinsic DNA binding specificity of the Hox partner at nucleotides 3' to a TAAT motif." *Nucleic Acids Research* 24 (12):2288-2294.

Korf, I. 2004. "Gene finding in novel genomes." *BMC Bioinformatics* 5. doi: 10.1186/1471-2105-5-59.

Kozomara, A., and S. Griffiths-Jones. 2011. "miRBase: integrating microRNA annotation and deep-sequencing data." *Nucleic Acids Research* 39 (Database Issue):D152-D157. doi: 10.1093/nar/gkq1027.

Krumlauf, R. 1993. "Mouse Hox genetic functions." *Current Opinion in Genetics and Development* 3 (4):621-625.

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

Kubatko, L. S., and J. H. Degnan. 2007. "Inconsistency of phylogenetic estimates from concatenated data under coalescence." *Systematic Biology* 56 (1):17-24. doi: 10.1080/10635150601146041.

Kurosawa, G., N. Takamatsu, M. Takahashi, M. Sumitomo, E. Sanaka, K. Yamada, K. Nishii, M. Matsuda, S. Asakawa, H. Ishiguro, K. Miura, Y. Kurosawa, N. Shimizu, Y. Kohara, and H. Hori. 2006. "Organization and structure of hox gene loci in medaka genome and comparison with those of pufferfish and zebrafish genomes." *Gene* 370:75-82. doi: 10.1016/j.gene.2005.11.015.

Laksanawimol, P., P. Damrongphol, and M. Kruatrachue. 2006. "Alteration of the brood pouch morphology during gestation of male seahorses, Hippocampus kuda." *Marine and Freshwater Research* 57 (5):497-502. doi: 10.1071/MF05112.

Lan, X., and J. K. Pritchard. 2016. "Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals." *Science* 352 (6288):1009-1013. doi: 10.1126/science.aad8411.

Laurenti, P., C. Thaeron, F. Allizard, A. Huysseune, and J. Y. Sire. 2004. "Cellular expression of eve1 suggests its requirement for the differentiation of the ameloblasts and for the initiation and morphogenesis of the first tooth in the zebrafish (Danio rerio)." *Developmental Dynamics* 230 (4):727-733. doi: 10.1002/dvdy.20080.

Le Pabic, P., J. L. Scemama, and E. J. Stellwag. 2010. "Role of Hox PG2 genes in Nile tilapia pharyngeal arch specification: implications for gnathostome pharyngeal arch evolution." *Evolution and Development* 12 (1):45-60. doi: 10.1111/j.1525-142X.2009.00390.x.

Le Pabic, P., E. J. Stellwag, S. N. Brothers, and J. L. Scemama. 2007. "Comparative analysis of Hox paralog group 2 gene expression during Nile tilapia (Oreochromis niloticus) embryonic development." *Development Genes and Evolution* 217 (11-12):749-758. doi: 10.1007/s00427-007-0182-z.

Lee, A. P., S. Y. Kerk, Y. Y. Tan, S. Brenner, and B. Venkatesh. 2010. "Ancient vertebrate conserved noncoding elements have been evolving rapidly in teleost fishes." *Molecular biology and evolution* 28 (3):1205-1215.

Lee, A. P., E. G. Koh, A. Tay, S. Brenner, and B. Venkatesh. 2006. "Highly conserved syntenic blocks at the vertebrate Hox loci and conserved regulatory elements within and outside Hox gene clusters." *Proceedings of the National Academy of Sciences of the United States of America* 103 (18):6994-6999. doi: 10.1073/pnas.0601492103.

Lewis, E. B. 1963. "Genes and Developmental Pathways." *American Zoologist* 3 (1):33-56.

Lewis, E. B. 1978. "A gene complex controlling segmentation in Drosophila." *Nature* 276 (5688):565-570.

Leysen, H., J. Christiaens, B. Kegel, M. N. Boone, L. Hoorebeke, and D. Adriaens. 2011. "Musculoskeletal structure of the feeding system and implications of snout elongation in Hippocampus reidi and Dunckerocampus dactyliophorus." *Journal of Fish Biology* 78 (6):1799-1823. doi: 10.1111/j.1095-8649.2011.02957.x.

Leysen, H., P. Jouk, M. Brunain, J. Christiaens, and D. Adriaens. 2010. "Cranial architecture of tube-snouted gasterosteiformes (Syngnathus rostellatus and Hippocampus capensis)." *Journal of Morphology* 271 (3):255-270. doi: 10.1002/jmor.10795.

Lin, Q., S. Fan, Y. Zhang, M. Xu, H. Zhang, Y. Yang, A. P. Lee, J. M. Woltering, V. Ravi, H. M. Gunter, W. Luo, Z. Gao, Z. W. Lim, G. Qin, R. F. Schneider, X. Wang, P. Xiong, G. Li, K. Wang, J. Min, C. Zhang, Y. Qiu, J. Bai, W. He, C. Bian, X. Zhang, D. Shan, H. Qu, Y. Sun, Q. Gao, L. Huang, Q. Shi, A. Meyer, and B. Venkatesh. 2016. "The seahorse genome and the evolution of its specialized morphology." *Nature* 540 (7633):395-399. doi: 10.1038/nature20595.

Lin, Q., Y. Qiu, R. Gu, M. Xu, J. Li, C. Bian, H. Zhang, G. Qin, Y. Zhang, W. Luo, J. Chen, X. You, M. Fan, M. Sun, P. Xu, B. Venkatesh, J. Xu, H. Fu, and Q. Shi. 2017. "Draft genome of the lined seahorse, Hippocampus erectus." *Gigascience* 6 (6):1-6. doi: 10.1093/gigascience/gix030.

Linnaeus (Linné), Carl von. 1758. *Systema naturae per regna tria naturae secundum classes, ordines, genera, species, cum characteribus differentiis, synonymis, locis. Vol. 1: Regnum animale*. Editio 10, reformata ed. 2 vols. Stockholm, Sweden: Laurentii Salvii.

Linnaeus (Linné), Carl von. 1766. *Systema naturae per regna tria naturae secundum classes, ordines, genera, species, cum characteribus differentiis, synonymis, locis. Vol. 1: Regnum animale*. Editio 12 ed. 2 vols. Stockholm, Sweden: Laurentii Salvii.

Logan, M., and C. J. Tabin. 1999. "Role of Pitx1 upstream of Tbx4 in specification of hindlimb identity." *Science* 283 (5408):1736-1739. doi: 10.1126/science.283.5408.1736.

Longo, S. J., B. C. Faircloth, A. Meyer, M. W. Westneat, M. E. Alfaro, and P. C. Wainwright. 2017. "Phylogenomic analysis of a rapid radiation of misfit fishes (Syngnathiformes) using ultraconserved elements." *Molecular phylogenetics and evolution* 113:33-48.

Lumsden, A. 2004. "Segmentation and compartition in the early avian hindbrain." *Mechanisms of Development* 121 (9):1081-1088. doi: 10.1016/j.mod.2004.04.018.

Lumsden, A., and R. Krumlauf. 1996. "Patterning the vertebrate neuraxis." *Science* 274 (5290):1109-1115.

Luo, W., and C. Brouwer. 2013. "Pathview: an R/Bioconductor package for pathway-based data integration and visualization." *Bioinformatics* 29 (14):1830-1831. doi: 10.1093/bioinformatics/btt285.

Luo, W., M. S. Friedman, K. Shedden, K. D. Hankenson, and P. J. Woolf. 2009. "GAGE: generally applicable gene set enrichment for pathway analysis." *BMC Bioinformatics* 10. doi: 10.1186/1471-2105-10-161.

MacDonald, R. B., M. Debiais-Thibaud, J. C. Talbot, and M. Ekker. 2010. "The relationship between dlx and gad1 expression indicates highly conserved genetic pathways in the zebrafish forebrain." *Developmental Dynamics* 239 (8):2298-2306. doi: 10.1002/dvdy.22365.

Maconochie, M. K., S. Nonchev, M. Manzanares, H. Marshall, and R. Krumlauf. 2001. "Differences in Krox20-dependent regulation of Hoxa2 and Hoxb2 during hindbrain development." *Developmental Biology* 233 (2):468-481. doi: 10.1006/dbio.2001.0197.

Maconochie, M. K., S. Nonchev, M. Studer, S. K. Chan, H. Pöpperl, M. H. Sham, R. S. Mann, and R. Krumlauf. 1997. "Cross-regulation in the mouse HoxB complex: the expression of Hoxb2 in rhombomere 4 is regulated by Hoxb1." *Genes and Development* 11 (14):1885-1895.

Maconochie, M., R. Krishnamurthy, S. Nonchev, P. Meier, M. Manzanares, P. J. Mitchell, and R. Krumlauf. 1999. "Regulation of Hoxa2 in cranial neural crest cells involves members of the AP-2 family." *Development* 126 (7):1483-1494.

Maere, S., S. De Bodt, J. Raes, T. Casneuf, M. Van Montagu, M. Kuiper, and Y. Van de Peer. 2005. "Modeling gene and genome duplications in eukaryotes." *Proceedings of the National Academy of Sciences of the United States of America* 102 (15):5454-5459.

Magoc, T., and S. L. Salzberg. 2011. "FLASH: fast length adjustment of short reads to improve genome assemblies." *Bioinformatics* 27 (21):2957-2963. doi: 10.1093/bioinformatics/btr507.

Mainguy, G., P. M. I. der Rieden, E. Berezikov, J. M. Woltering, R. H. Plasterk, and A. J. Durston. 2003. "A position-dependent organisation of retinoid response elements is conserved in the vertebrate Hox clusters." *Trends in Genetics* 19((9):476-479.

Mallo, M, T Vinagre, and M Carapuco. 2009. "The road to the vertebral formula." *International Journal Of Developmental Biology* 53 (8-10):1469-1481.

Mallo, M., and C. R. Alonso. 2013. "The regulation of Hox gene expression during animal development." *Development* 140 (19):3951-3963. doi: 10.1242/dev.068346.

Mallo, M., D. M. Wellik, and J. Deschamps. 2010. "Hox genes and regional patterning of the vertebrate body plan." *Developmental Biology* 344 (1):7-15. doi: 10.1016/j.ydbio.2010.04.024.

Malmstrom, M., R. Britz, M. Matschiner, O. K. Torresen, R. K. Hadiaty, N. Yaakob, H. H. Tan, K. S. Jakobsen, W. Salzburger, and L. Ruber. 2018. "The Most Developmentally Truncated Fishes Show Extensive Hox Gene Loss and Miniaturized Genomes." *Genome Biology and Evolution* 10 (4):1088-1103. doi: 10.1093/gbe/evy058.

Mank, J. E., and J. C. Avise. 2006. "Phylogenetic conservation of chromosome numbers in Actinopterygiian fishes." *Genetica* 127 (1-3):321-327. doi: 10.1007/s10709-005-5248-0.

Manley, N. R., and M. R. Capecchi. 1998. "Hox group 3 paralogs regulate the development and migration of the thymus, thyroid, and parathyroid glands." *Developmental Biology* 195 (1):1-15. doi: 10.1006/dbio.1997.8827.

Manzanares, M., S. Bel-Vialar, L. Ariza-McNaughton, E. Ferretti, H. Marshall, M. M. Maconochie, F. Blasi, and R. Krumlauf. 2001. "Independent regulation of initiation and maintenance phases of Hoxa3 expression in the vertebrate hindbrain involve auto- and cross-regulatory mechanisms." *Development* 128 (18):3595-3607.

Marcil, A. 2003. "Pitx1 and Pitx2 are required for development of hindlimb buds." *Development* 130 (1):45-55. doi: 10.1242/dev.00192.

Matsunami, M., K. Sumiyama, and N. Saitou. 2010. "Evolution of conserved non-coding sequences within the vertebrate Hox clusters through the two-round whole genome duplications revealed by phylogenetic footprinting analysis." *Journal of molecular evolution* 71 (5-6):427-436.

Mayor, C., M. Brudno, J. R. Schwartz, A. Poliakov, E. M. Rubin, and K. A. Frazer. 2000. "VISTA: visualizing global DNA sequence alignments of arbitrary length." *Bioinformatics* 16 (11):1046-1047. doi: 10.1093/bioinformatics/16.11.1046.

Mayr, E. 1960. "The emergence of evolutionary novelties." In *Evolution After Darwin*, edited by S. Tax. Chicago, Illinois, USA: The University of Chicago.

McCormack, R., L. Armas, M. Shiratsuchi, and E. R. Podack. 2013. "Killing machines: three pore-forming proteins of the immune system." *Immunologic Research* 57 (1-3):268-278. doi: 10.1007/s12026-013-8469-9.

McEllin, J. A., T. B. Alexander, S. Tumpel, L. M. Wiedemann, and R. Krumlauf. 2016. "Analyses of fugu hoxa2 genes provide evidence for subfunctionalization of neural crest cell and rhombomere cis-regulatory modules during vertebrate evolution." *Developmental Biology* 409 (2):530-542. doi: 10.1016/j.ydbio.2015.11.006.

McGaugh, S. E., J. B. Gross, B. Aken, M. Blin, R. Borowsky, D. Chalopin, Hélène Hinaux, William R Jeffery, Alex Keene, Li Ma, Patrick Minx, Daniel Murphy, Kelly E O'quin, Sylvie Rétaux, Nicolas Rohner, Steve M. J Searle, Bethany A Stahl, Cliff Tabin, Jean-Nicolas Volff, Masato Yoshizawa, and Wesley C Warren. 2014. "The cavefish genome reveals candidate genes for eye loss." *Nature Communications* 5. doi: 10.1038/ncomms6307.

McGinnis, W., R. L. Garber, J. Wirz, A. Kuroiwa, and W. J. Gehring. 1984. "A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans." *Cell* 37 (2):403-408.

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

McIntyre, D. C., S. Rakshit, A. R. Yallowitz, L. Loken, L. Jeannotte, M. R. Capecchi, and D. M. Wellik. 2007. "Hox patterning of the vertebrate rib cage." *Development* 134 (16):2981-2989. doi: 10.1242/dev.007567.

Mi, H., A. Muruganujan, J. T. Casagrande, and P. D. Thomas. 2013. "Large-scale gene function analysis with the PANTHER classification system." *Nature Protocols* 8 (8):1551-1566. doi: 10.1038/nprot.2013.092.

Mi, H., S. Poudel, A. Muruganujan, J. T. Casagrande, and P. D. Thomas. 2016. "PANTHER version 10: expanded protein families and functions, and analysis tools." *Nucleic Acids Res.* 44. doi: 10.1093/nar/gkv1194.

Minoux, M., and F. M. Rijli. 2010. "Molecular mechanisms of cranial neural crest cell migration and patterning in craniofacial development." *Development* 137 (16):2605-2621. doi: 10.1242/dev.040048.

Mohrlen, F., M. Maniura, G. Plickert, M. Frohme, and U. Frank. 2006. "Evolution of astacin-like metalloproteases in animals and their function in development." *Evolution and Development* 8 (2):223-231. doi: 10.1111/j.1525-142X.2006.00092.x.

Moriya, Y., M. Itoh, S. Okuda, A. C. Yoshizawa, and M. Kanehisa. 2007. "KAAS: an automatic genome annotation and pathway reconstruction server." *Nucleic Acids Research* 35 (Web Server Issue):W182-W185. doi: 10.1093/nar/gkm321.

Muller, G. B., and G. P. Wagner. 1991. "Novelty in evolution - restructuring the concept." *Annual Review Of Ecology And Systematics* 22:229-256. doi: 10.1146/annurev.es.22.110191.001305.

Muller, M. 1987. "Optimization principles applied to the mechanism of neurocranium levation and mouth bottom depression in bony fishes (Halecostomi)." *Journal of Theoretical Biology* 126 (3):343-368.

Muller, M., and J. W. M. Osse. 1984. "Hydrodynamics of suction feeding in fish." *Transactions of the Zoological Society of London* 37 (2):51-135.

Naiche, L. A. 2003. "Loss of Tbx4 blocks hindlimb development and affects vascularization and fusion of the allantois." *Development* 130 (12):2681-2693. doi: 10.1242/dev.00504.

Naiche, L. A., and V. E. Papaioannou. 2007. "Tbx4 is not required for hindlimb identity or post-bud hindlimb outgrowth." *Development* 134 (1):93-103. doi: 10.1242/dev.02712.

Nakamura, Y., K. Mori, K. Saitoh, K. Oshima, M. Mekuchi, T. Sugaya, Yuya Shigenobu, Nobuhiko Ojima, Shigeru Muta, Atushi Fujiwara, Motoshige Yasuike, Ichiro Oohara, Hideki Hirakawa, Vishwajit Sur Chowdhury, Takanori Kobayashi, Kazuhiro Nakajima, Motohiko Sano, Tokio Wada, Kosuke Tashiro, Kazuho Ikeo, Masahira Hattori, Satoru Kuhara, Takashi Gojobori, and Kiyoshi Inouye. 2013. "Evolutionary changes of multiple visual pigment genes in the complete genome of Pacific bluefin tuna." *Proceedings of the National Academy of Sciences of the United States of America* 110 (27):11061-11066. doi: 10.1073/pnas.1302051110.

Naruse, K., M. Tanaka, K. Mita, A. Shima, J. Postlethwait, and H. Mitani. 2004. "A medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping." *Genome Research* 14 (5):820-828. doi: 10.1101/gr.2004004.

Navratilova, P., D. Fredman, T. A. Hawkins, K. Turner, B. Lenhard, and T. S. Becker. 2009. "Systematic human/zebrafish comparative identification of cis-regulatory activity around vertebrate developmental transcription factor genes." *Developmental Biology* 327 (2):526-540. doi: 10.1016/j.ydbio.2008.10.044.

Near, T. J., A. Dornburg, R. I. Eytan, B. P. Keck, W. L. Smith, and K. L. Kuhn. 2013. "Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes." *Proceedings of the National Academy of Sciences of the United States of America* 110 (31):12738-12743. doi: 10.1073/pnas.1304661110.

Nelson, G.J. 1989. "Phylogeny of major fish groups." In *The hierarchy of life : molecules and morphology in phylogenetic analysis : proceedings from Nobel Symposium 70 held at Alfred Nobel's Björkborn, Karlskoga, Sweden, August 29-September 2, 1988*, edited by B. Fernholm, K. Bremer, L. Brundin, H. Jörnvall, L. Rutberg and H.E. Wanntorp, 325-336. Amsterdam, Netherlands, New York City, New York, USA: Elsevier Science Publishing Company.

Nelson, J. S. . 1971. "Comparison of the pectoral and pelvic skeletons and of some other bones and their phylogenetic implications in the Aulorhynchidae and Gasterosteidae (Pisces). ." *Journal of the Fisheries Board of Canada* 28 (3): 427-442.

Nelson, Joseph S. 2006. *Fishes of the world*. 4th ed. ed. Hoboken, New Jersey, USA: John Wiley.

Neutens, C., D. Adriaens, J. Christiaens, B. Kegel, M. Dierick, and R. Boistel. 2014. "Grasping convergent evolution in syngnathids: a unique tale of tails." *Journal of Anatomy* 224 (6):710-723. doi: 10.1111/joa.12181.

Nonchev, S., M. Maconochie, C. Vesque, S. Aparicio, L. Ariza-McNaughton, M. Manzanares, K. Maruthainar, A. Kuroiwa, S. Brenner, P. Charnay, and R. Krumlauf. 1996. "The conserved role of Krox-20 in directing Hox gene expression during vertebrate hindbrain segmentation." *Proceedings of the National Academy of Sciences of the United States of America* 93 (18):9339-9345.

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

Ohno, S. 1970. *Evolution by Gene Duplication*. Heidelberg, Germany: Springer-Verlag.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'hara, G. L. Simpson, P. Solymos, M. H. Stevens, and H. Wagner. 2015. "vegan: Community Ecology Package; R package version 2.3-5."

Oma, Y., Y. Kino, N. Sasagawa, and S. Ishiura. 2004. "Intracellular localization of homopolymeric amino acid-containing proteins expressed in mammalian cells." *The Journal of Biological Chemistry* 279 (20):21217-21222. doi: 10.1074/jbc.M309887200.

Paczolt, K. A., and A. G. Jones. 2010. "Post-copulatory sexual selection and sexual conflict in the evolution of male pregnancy." *Nature* 464 (7287):401-404. doi: 10.1038/nature08861.

Pan, Hailin, Hao Yu, Vydianathan Ravi, Cai Li, Alison P. Lee, Michelle M. Lian, Boon-Hui Tay, Sydney Brenner, Jian Wang, Huanming Yang, Guojie Zhang, and Byrappa %J GigaScience Venkatesh. 2016. "The genome of the largest bony fish, ocean sunfish (*Mola mola)*, provides insights into its fast growth rate." *GigaScience* 5 (1):36. doi: 10.1186/s13742-016-0144-3.

Parker, H. J., M. E. Bronner, and R. Krumlauf. 2014. "A Hox regulatory network of hindbrain segmentation is conserved to the base of vertebrates." *Nature* 514 (7523):490-493. doi: 10.1038/nature13723.

Parker, H. J., M. E. Bronner, and R. Krumlauf. 2016. "The vertebrate Hox gene regulatory network for hindbrain segmentation: Evolution and diversification: Coupling of a Hox gene regulatory network to hindbrain segmentation is an ancient trait originating at the base of vertebrates." *Bioessays* 38 (6):526-538. doi: 10.1002/bies.201600010.

Parra, G., K. Bradnam, Z. Ning, T. Keane, and I. Korf. 2009. "Assessing the gene space in draft genomes." *Nucleic Acids Research* 37 (1):289-297. doi: 10.1093/nar/gkn916.

Pascual-Anaya, J., S. D'Aniello, S. Kuratani, and J. Garcia-Fernandez. 2013. "Evolution of Hox gene clusters in deuterostomes." *BMC Developmental Biology* 13:26. doi: 10.1186/1471-213X-13-26.

Pascual-Anaya, J., I. Sato, F. Sugahara, S. Higuchi, J. Paps, Y. Ren, W. Takagi, A. Ruiz-Villalba, K. G. Ota, W. Wang, and S. Kuratani. 2018. "Hagfish and lamprey Hox genes reveal conservation of temporal colinearity in vertebrates." *Nature Ecology and Evolution* 2 (5):859-866. doi: 10.1038/s41559-018-0526-2.

Pasqualetti, M., M. Ori, I. Nardi, and F. M. Rijli. 2000. "Ectopic Hoxa2 induction after neural crest migration results in homeosis of jaw elements in Xenopus." *Development* 127 (24):5367-5378.

Pennacchio, L. A., N. Ahituv, A. M. Moses, S. Prabhakar, M. A. Nobrega, M. Shoukry, S. Minovitsky, I. Dubchak, A. Holt, K. D. Lewis, I. Plajzer-Frick, J. Akiyama, S. De Val, V. Afzal, B. L. Black, O. Couronne, M. B. Eisen, A. Visel, and E. M. Rubin. 2006. "In vivo enhancer analysis of human conserved non-coding sequences." *Nature* 444 (7118):499-502. doi: 10.1038/nature05295.

Peterson, Ron L., Thomas Papenbrock, Michele M. Davda, and Alexander Awgulewitsch. 1994. "The murine Hoxc cluster contains five neighbouring AbdB-related Hox genes that show unique spatially coordinated expression in posterior embryonic regions." *Mechanisms of Development* 47 (3):253-260.

Playfair, R. L., and A. C. L. G. Günther. 1866. *The fishes of Zanzibar, with a list of the fishes of the whole East coast of Africa*. London, United Kingdom: John van Voorst.

Pollard, D. A. 1984. "A review of ecological studies on seagrass fish communities, with particular reference to recent studies in Australia." *Aquatic Botany* 18 (1-2):3–42.

Polychronopoulos, D., J. W. D. King, A. J. Nash, G. Tan, and B. Lenhard. 2017. "Conserved non-coding elements: developmental gene regulation meets genome organization." *Nucleic Acids Research* 45 (22):12611-12624. doi: 10.1093/nar/gkx1074.

Porter, Michael M, Dominique Adriaens, Ross L Hatton, Marc A Meyers, and Joanna McKittrick. 2015. "Why the seahorse tail is square." *Science* 349 (6243):aaa6683.

Postlethwait, J., A. Amores, W. Cresko, A. Singer, and Y. L. Yan. 2004. "Subfunction partitioning, the teleost radiation and the annotation of the human genome." *Trends in Genetics* 20 (10):481-490.

Prince, V., and A. Lumsden. 1994. "Hoxa-2 expression in normal and transposed rhombomeres: independent regulation in the neural tube and neural crest." *Development* 120 (4):911-923.

Putnam, N. H., T. Butts, D. E. Ferrier, R. F. Furlong, U. Hellsten, T. Kawashima, M. Robinson-Rechavi, E. Shoguchi, A. Terry, J. K. Yu, E. L. Benito-Gutierrez, I. Dubchak, J. Garcia-Fernandez, J. J. Gibson-Brown, I. V. Grigoriev, A. C. Horton, P. J. de Jong, J. Jurka, V. V. Kapitonov, Y. Kohara, Y. Kuroki, E. Lindquist, S. Lucas, K. Osoegawa, L. A. Pennacchio, A. A. Salamov, Y. Satou, T. Sauka-Spengler, J. Schmutz, I. T. Shin, A. Toyoda, M. Bronner-Fraser, A. Fujiyama, L. Z. Holland, P. W. Holland, N. Satoh, and D. S. Rokhsar. 2008. "The amphioxus genome and the evolution of the chordate karyotype." *Nature* 453 (7198):1064-1071. doi: 10.1038/nature06967.

Qiu, M., A. Bulfone, S. Martinez, J. J. Meneses, K. Shimamura, and R. A. Pedersen. 1995. "Null mutation of Dlx-2 results in abnormal morphogenesis of proximal first and second branchial arch derivatives and abnormal differentiation in the forebrain." *Genes and Development* 9 (20):2523-2538. doi: 10.1101/gad.9.20.2523.

Quevillon, E., V. Silventoinen, S. Pillai, N. Harte, N. Mulder, and R. Apweiler. 2005. "InterProScan: protein domains identifier." *Nucleic Acids Research* 33 (Web Server Issue):W116-W120. doi: 10.1093/nar/gki442.

Quinlan, A. R. 2014. "BEDTools: The Swiss-Army tool for genome feature analysis." *Current Protocols in Bioinformatics* 47:11.12.1-34.

Quiring, Rebecca, Uwe Walldorf, Urs Kloter, and Walter J Gehring. 1994. "Homology of the eyeless gene of Drosophila to the Small eye gene in mice and Aniridia in humans." *Science* 265 (5173):785-789.

Raff, Rudolf A. 2012. *The shape of life: genes, development, and the evolution of animal form*. Chicago, Illinois, USA: University of Chicago Press.

Ravi, V., K. Lam, B. H. Tay, A. Tay, S. Brenner, and B. Venkatesh. 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 (38):16327-16332. doi: 10.1073/pnas.0907914106.

Rebeiz, M., and M. Tsiantis. 2017. "Enhancer evolution and the origins of morphological novelty." *Current Opinion in Genetics and Development* 45:115-123. doi: 10.1016/j.gde.2017.04.006.

Renz, A. J., H. M. Gunter, J. M. Fischer, H. Qiu, A. Meyer, and S. Kuraku. 2011. "Ancestral and derived attributes of the dlx gene repertoire, cluster structure and expression patterns in an African cichlid fish." *EvoDevo* 2 (1):1. doi: 10.1186/2041-9139-2-1.

Rice, P., I. Longden, and A. Bleasby. 2000. "EMBOSS: The European Molecular Biology Open Software Suite." *Trends in Genetics* 16 (6):276-277. doi: 10.1016/S0168-9525(00)02024-2.

Rijli, F. M., M. Mark, S. Lakkaraju, A. Dierich, P. Dolle, and P. Chambon. 1993. "A homeotic transformation is generated in the rostral branchial region of the head by disruption of Hoxa-2, which acts as a selector gene." *Cell* 75 (7):1333-1349.

Ripley, J. L. 2009. "Osmoregulatory role of the paternal brood pouch for two Syngnathus species." *Comparative Biochemistry and Physiology* 154 (1):98-104. doi: 10.1016/j.cbpa.2009.05.003.

Ripley, J. L., and C. M. Foran. 2009. "Direct evidence for embryonic uptake of paternally-derived nutrients in two pipefishes (Syngnathidae: Syngnathus spp.)." *Journal of Comparative Physiology B* 179 (3):325-333. doi: 10.1007/s00360-008-0316-2.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." *Bioinformatics* 26 (1):139-140. doi: 10.1093/bioinformatics/btp616.

Robinson, M. D., and G. K. Smyth. 2008. "Small-sample estimation of negative binomial dispersion, with applications to SAGE data." *Biostatistics* 9 (2):321-332. doi: 10.1093/biostatistics/kxm030.

Rochtus, A., B. Izzi, E. Vangeel, S. Louwette, C. Wittevrongel, D. Lambrechts, Y. Moreau, R. Winand, C. Verpoorten, K. Jansen, C. Van Geet, and K. Freson. 2015. "DNA methylation analysis of Homeobox genes implicates HOXB7 hypomethylation as risk factor for neural tube defects." *Epigenetics* 10 (1):92-101. doi: 10.1080/15592294.2014.998531.

Roos, Gert, Sam Van Wassenbergh, Anthony Herrel, and Peter Aerts. 2009. "Kinematics of suction feeding in the seahorse Hippocampus reidi." *The Journal of Experimental Biology* 212:3490-3498.

Roth, O., V. Klein, A. Beemelmanns, J. P. Scharsack, and T. B. H. Reusch. 2012. "Male pregnancy and biparental immune priming." *The American Naturalist* 180 (6):802-814. doi: 10.1086/668081.

Sanciangco, M. D., K. E. Carpenter, and R. R. Betancur. 2016. "Phylogenetic placement of enigmatic percomorph families (Teleostei: Percomorphaceae)." *Molecular Phylogenetics and Evolution* 94:565-576. doi: 10.1016/j.ympev.2015.10.006.

Sandelin, A, P Bailey, S Bruce, P. G. Engstrom, J. M. Klos, W. W. Wasserman, J Ericson, and B Lenhard. 2004. "Arrays of ultraconserved noncoding regions span the loci of key developmental genes in vertebrate genomes." *BMC Genomics* 5 (1):99.

Santagati, F., M. Minoux, S. Y. Ren, and F. M. Rijli. 2005. "Temporal requirement of Hoxa2 in cranial neural crest skeletal morphogenesis." *Development* 132 (22):4927-4936. doi: 10.1242/dev.02078.

Santagati, F., and F. M. Rijli. 2003. "Cranial neural crest and the building of the vertebrate head." *Nature Reviews Neuroscience* 4 (10):806-818. doi: 10.1038/nrn1221.

Santini, S., J. L. Boore, and A. Meyer. 2003. "Evolutionary conservation of regulatory elements in vertebrate Hox gene clusters." *Genome research* 13 (6a):1111-1122.

Scemama, J. L., J. L. Vernon, and E. J. Stellwag. 2006. "Differential expression of hoxa2a and hoxa2b genes during striped bass embryonic development." *Gene Expression Patterns* 6 (8):843-848. doi: 10.1016/j.modgep.2006.02.004.

Schilling, Thomas F., and C. B. Kimmel. 1997. "Musculoskeletal patterning in the pharyngeal segments of the zebrafish embryo." *Development* 124 (124):2945-2960.

Scott, M. P. 1992. "Vertebrate homeobox gene nomenclature." *Cell* 71 (4):551-553.

Scott, Matthew P., and Amy J. Weiner. 1984. "Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax and fushi tarazu loci of Drosophila." *Proceedings of the National Academy of Sciences of the United States of America* 81 (13):4113-4117.

Seebald, J. L., and D. P. Szeto. 2011. "Zebrafish eve1 regulates the lateral and ventral fates of mesodermal progenitor cells at the onset of gastrulation." *Developmental Biology* 349 (1):78-89. doi: 10.1016/j.ydbio.2010.10.005.

Semina, E. V., R. E. Ferrell, H. A. Mintz-Hittner, P. Bitoun, W. L. M. Alward, and R. S. Reiter. 1998. "A novel homeobox gene PITX3 is mutated in families with autosomal-dominant cataracts and ASMD." *Nature Genetics* 19 (2):167-170. doi: 10.1038/527.

Seoighe, C., and C. Gehring. 2004. "Genome duplication led to highly selective expansion of the Arabidopsis thaliana proteome." *Trends in Genetics* 20 (10):461-464. doi: 10.1016/j.tig.2004.07.008.

Seoighe, C., and K. H. Wolfe. 1999. "Yeast genome evolution in the post-genome era." *Current Opinion in Microbiology* 2 (5):548-554.

Shapiro, M. D., M. E. Marks, C. L. Peichel, B. K. Blackman, K. S. Nereng, and B. Jonsson. 2004. "Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks." *Nature* 428 (6984):717-723. doi: 10.1038/nature02415.

Shi, X., D. V. Bosenko, N. S. Zinkevich, S. Foley, D. R. Hyde, and E. V. Semina. 2005. "Zebrafish pitx3 is necessary for normal lens and retinal development." *Mechanisms of Development* 122 (4):513-527. doi: 10.1016/j.mod.2004.11.012.

Shin, J. T., J. R. Priest, I. Ovcharenko, A. Ronco, R. K. Moore, C. G. Burns, and C. A. MacRae. 2005. "Human-zebrafish non-coding conserved elements act in vivo to regulate transcription." *Nucleic Acids Research* 33 (17):5437-5445. doi: 10.1093/nar/gki853.

Shubin, N., C. Tabin, and S. Carroll. 2009. "Deep homology and the origins of evolutionary novelty." *Nature* 457 (7231):818-823. doi: 10.1038/nature07891.

Sidow, A. 1996. "Gen(om)e duplications in the evolution of early vertebrates." *Current Opinion in Genetics and Development* 6 (6):715-722.

Simpson, J. T., and R. Durbin. 2012. "Efficient de novo assembly of large genomes using compressed data structures." *Genome Research* 22 (3):549-556. doi: 10.1101/gr.126953.111.

Small, C. M., S. Bassham, J. Catchen, A. Amores, A. M. Fuiten, R. S. Brown, A. G. Jones, and W. A. Cresko. 2016. "The genome of the Gulf pipefish enables understanding of evolutionary innovations." *Genome Biology* 17 (1):258. doi: 10.1186/s13059-016-1126-6.

Small, C. M., A. D. Harlin-Cognato, and A. G. Jones. 2013. "Functional similarity and molecular divergence of a novel reproductive transcriptome in two male-pregnant Syngnathus pipefish species." *Ecology and Evolution* 3 (12):4092-4108. doi: 10.1002/ece3.763.

RepeatModeler Open-1.0.8.

RepeatMasker Open-4.0.5.

Smith, F. W., T. C. Boothby, I. Giovannini, L. Rebecchi, E. L. Jockusch, and B. Goldstein. 2016. "The compact body plan of tardigrades evolved by the loss of a large body region " *Current Biology* 26 (2):224-229.

Sperber, S. M., V. Saxena, G. Hatch, and M. Ekker. 2008. "Zebrafish dlx2a contributes to hindbrain neural crest survival, is necessary for differentiation of sensory ganglia and functions with dlx1a in maturation of the arch cartilage elements." *Developmental Biology* 314 (1):59-70. doi: 10.1016/j.ydbio.2007.11.005.

Stamatakis, A. 2014. "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies." *Bioinformatics* 30 (9):1312-1313. doi: 10.1093/bioinformatics/btu033.

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. "A rapid bootstrap algorithm for the RAxML Web servers." *Systematic Biology* 57 (5):758-771. doi: 10.1080/10635150802429642.

Stanke, M., and S. Waack. 2003. "Gene prediction with a hidden Markov model and a new intron submodel." *Bioinformatics* 19 (2):215-225. doi: 10.1093/bioinformatics/btg1080.

Stern, D. L. . 2000. "Perspective: evolutionary developmental biology and the problem of variation. ." *Evolution* 54 (4):1079-1091.

Stern, D. L., and N. Frankel. 2013. "The structure and evolution of cis-regulatory regions: the shavenbaby story." *Philosophical Transactions Of The Royal Society B-Biological Sciences* 368 (1632):20130028. doi: 10.1098/rstb.2013.0028.

Sternberg, S. H., and J. A. Doudna. 2015. "Expanding the biologist's toolkit with CRISPR-Cas9." *Molecular Cell* 58 (4):568-574. doi: 10.1016/j.molcel.2015.02.032.

Stock, D. W., W. R. Jackman, and J. Trapani. 2006. "Developmental genetic mechanisms of evolutionary tooth loss in cypriniform fishes." *Development* 133 (16):3127-3137. doi: 10.1242/dev.02459.

Struhl, G. 1981. "A homoeotic mutation transforming leg to antenna in Drosophila." *Nature* 292 (5824)):635.

Studer, M., A. Gavalas, H. Marshall, L. Ariza-McNaughton, F. M. Rijli, P. Chambon, and R. Krumlauf. 1998. "Genetic interactions between Hoxa1 and Hoxb1 reveal new roles in regulation of early hindbrain patterning." *Development* 125 (6):1025-1036.

Swalla, B. J. 2006. "Building divergent body plans with similar genetic pathways." *Heredity* 97 (3):235-243. doi: 10.1038/sj.hdy.6800872.

Tanaka, M., L. A. Hale, A. Amores, Y. L. Yan, W. A. Cresko, and T. Suzuki. 2005. "Developmental genetic basis for the evolution of pelvic fin loss in the pufferfish Takifugu rubripes." *Developmental Biology* 281 (2):227-239. doi: 10.1016/j.ydbio.2005.02.016.

Tanzer, A., C. T. Amemiya, C. B. Kim, and P. F. Stadler. 2005. "Evolution of microRNAs located within Hox gene clusters." *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 304 (1):75-85. doi: 10.1002/jez.b.21021.

Team, R Core. 2015. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Teske, P. R., and L. B. Beheregaray. 2009. "Evolution of seahorses' upright posture was linked to Oligocene expansion of seagrass habitats." *Biology Letters* 5 (4):521-3. doi: 10.1098/rsbl.2009.0152.

Torresen, O. K., B. Star, S. Jentoft, W. B. Reinar, H. Grove, J. R. Miller, B. P. Walenz, J. Knight, J. M. Ekholm, P. Peluso, R. B. Edvardsen, A. Tooming-Klunderud, M. Skage, S. Lien, K. S. Jakobsen, and A. J. Nederbragt. 2017. "An improved genome assembly uncovers prolific tandem repeats in Atlantic cod." *BMC Genomics* 18 (1):95. doi: 10.1186/s12864-016-3448-x.

Trainor, P. A., and R. Krumlauf. 2000. "Patterning the cranial neural crest: hindbrain segmentation and Hox gene plasticity." *Nature Reviews Neuroscience* 1 (2):116-124. doi: 10.1038/35039056.

Trainor, P. A., and R. Krumlauf. 2001. "Hox genes, neural crest cells and branchial arch patterning." *Current Opinion in Cell Biology* 13 (6):698-705.

Tumpel, S., F. Cambronero, E. Ferretti, F. Blasi, L. M. Wiedemann, and R. Krumlauf. 2007. "Expression of Hoxa2 in rhombomere 4 is regulated by a conserved cross-regulatory mechanism dependent upon Hoxb1." *Developmental Biology* 302 (2):646-60. doi: 10.1016/j.ydbio.2006.10.029.

Tumpel, S., F. Cambronero, L. M. Wiedemann, and R. Krumlauf. 2006. "Evolution of cis elements in the differential expression of two Hoxa2 coparalogous genes in pufferfish (Takifugu rubripes)." *Proceedings of the National Academy of Sciences of the United States of America* 103 (14):5419-5424. doi: 10.1073/pnas.0600993103.

Tumpel, S., L. M. Wiedemann, and R. Krumlauf. 2009. "Hox genes and segmentation of the vertebrate hindbrain." *Current Topics in Developmental Biology,* 88:103-137. doi: 10.1016/S0070-2153(09)88004-6.

Tyler, J. C. 1980. "Osteology, phylogeny, and higher classification of the fishes of the order Plectognathi (Tetraodontiformes)." *NOAA Technical Report NMFS Circular* 434.

Valenzano, D. R., B. A. Benayoun, P. P. Singh, E. Zhang, P. D. Etter, and C. K. Hu. 2015. "The African turquoise killifish genome provides insights into evolution and genetic architecture of lifespan." *Cell* 163 (6):1539-1554. doi: 10.1016/j.cell.2015.11.008.

Van de Peer, Y., S. Maere, and A. Meyer. 2009. "The evolutionary significance of ancient genome duplications." *Nature Reviews Genetics* 10 (10):725-32. doi: 10.1038/nrg2600.

Van den Akker, E., C. Fromental-Ramain, W. de Graaff, H. Le Mouellic, P. Brulet, P. Chambon, and J. Deschamps. 2001. "Axial skeletal patterning in mice lacking all paralogous group 8 Hox genes." *Development* 128 (10):1911-1921.

Van Wassenbergh, Sam, Gert Roos, Peter Aerts, Anthony Herrel, and Dominique Adriaens. 2011. "Why the long face? A comparative study of feeding kinematics of two pipefishes with different snout lengths." *Journal of Fish Biology* 78 (6):1786-1798.

Van Wassenbergh, Sam, Gert Roos, and Lara Ferry. 2011. "An adaptive explanation for the horse-like shape of seahorses." *Nature Communications* 2:164.

Venkatesh, B., E. F. Kirkness, Y. H. Loh, A. L. Halpern, A. P. Lee, J. Johnson, N. Dandona, L. D. Viswanathan, A. Tay, J. C. Venter, R. L. Strausberg, and S. Brenner. 2006. "Ancient noncoding elements conserved in the human genome." *Science* 314 (5807):1892. doi: 10.1126/science.1130708.

Vitturi, R., A. Libertini, M. Campolmi, F. Calderazzo, and A. Mazzola. 1998. "Conventional karyotype, nucleolar organizer regions and genome size in five Mediterranean species of Syngnathidae (Pisces, Syngnathiformes)." *Journal of Fish Biology* 52 (4):677-687. doi: 10.1111/j.1095-8649.1998.tb00812.x.

von Baer, Karl Ernst 1828. *Uber Entwicklungsgeschichte der Thiere: Beobachtung und Reflexion*. Borntraeger, Koenigsberg.

Wagner, G. P., and V. J. Lynch. 2010. "Evolutionary novelties." *Current Biology* 20 (2):R48-R52. doi: 10.1016/j.cub.2009.11.010.

Wahba, G. M., S. L. Hostikka, and E. M. Carpenter. 2001. "The paralogous Hox genes Hoxa10 and Hoxd10 interact to pattern the mouse hindlimb peripheral nervous system and skeleton." *Developmental Biology* 231 (1):87-102. doi: 10.1006/dbio.2000.0130.

Ward, Andrea B, and Elizabeth L Brainerd. 2007. "Evolution of axial patterning in elongate fishes." *Biological Journal of the Linnean Society* 90 (1):97-116.

Watanabe, S., T. Kaneko, and Y. Watanabe. 1999. "Immunocytochemical detection of mitochondria-rich cells in the brood pouch epithelium of the pipefish, Syngnathus schlegeli: structural comparison with mitochondria-rich cells in the gills and larval epidermis." *Cell and Tissue Research* 295 (1):141-149. doi: 10.1007/s004410051220.

Wellik, D. M. 2009. "Hox genes and vertebrate axial pattern." *Current Topics in Developmental Biology* 88:257-278. doi: 10.1016/S0070-2153(09)88009-5.

Wellik, D. M., and M. R. Capecchi. 2003. "Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton." *Science* 301 (5631):363-367. doi: 10.1126/science.1085672.

Wellik, D. M., P. J. Hawkes, and M. R. Capecchi. 2002. "Hox11 paralogous genes are essential for metanephric kidney induction." *Genes and Development* 16 (11):1423-1432. doi: 10.1101/gad.993302.

Whitfield, A. K. . 1999. "Ichthyofaunal assemblages in estuaries: a South African study." *Reviews in Fish Biology and Fisheries* 9 (2):151–186.

Whittington, C. M., O. W. Griffith, W. Qi, M. B. Thompson, and A. B. Wilson. 2015. "Seahorse brood pouch transcriptome reveals common genes associated with vertebrate pregnancy." *Molecular Biology and Evolution* 32 (12):3114-3131.

Wilkins, Adam S. 2002. *The evolution of developmental pathways*. Sunderland, Massachusetts, USA: Sinauer Associates Inc.

Wilson, A. B., I. Ahnesjo, A. C. Vincent, and A. Meyer. 2003. "The dynamics of male brooding, mating patterns, and sex roles in pipefishes and seahorses (family Syngnathidae)." *Evolution* 57 (6):1374-1386. doi: 10.1111/j.0014-3820.2003.tb00345.x.

Wilson, N. G., and G. W. Rouse. 2010. "Convergent camouflage and the non-monophyly of 'seadragons' (Syngnathidae: Teleostei): suggestions for a revised taxonomy of syngnathids." *Zoologica Scripta* 39 (6):551-558. doi: 10.1111/j.1463-6409.2010.00449.x.

Wong, S. F. L., V. Agarwal, J. H. Mansfield, N. Denans, M. G. Schwartz, H. M. Prosser, Olivier Pourquié, David P. Bartel, Clifford J. Tabin, and E. McGlinn. 2015. "Independent regulation of vertebral number and vertebral identity by microRNA-196 paralogs." *Proceedings of the National Academy of Sciences* 112 (35):E4884-E4893.

Woolfe, Adam, Martin Goodson, Debbie K. Goode, Phil Snell, Gayle K. McEwen, Tanya Vavouri, Sarah F. Smith, Phil North, Heather Callaway, Krys Kelly, Klaudia Walter, Irina Abnizova, Walter Gilks, Yvonne J. K. Edwards, Julie E. Cooke, and Greg Elgar. 2004. "Highly Conserved Non-Coding Sequences Are Associated with Vertebrate Development." *PLOS Biology* 3 (1):e7. doi: 10.1371/journal.pbio.0030007.

Wray, G. A. 2007. "The evolutionary significance of cis-regulatory mutations." *Nature Reviews Genetics* 8 (3):206-216. doi: 10.1038/nrg2063.

Wu, T. D., and S. Nacu. 2010. "Fast and SNP-tolerant detection of complex variants and splicing in short reads." *Bioinformatics* 26 (7):873-881. doi: 10.1093/bioinformatics/btq057.

Yasuike, M., A. Fujiwara, Y. Nakamura, Y. Iwasaki, I. Nishiki, T. Sugaya, A. Shimizu, M. Sano, T. Kobayashi, and M. Ototake. 2016. "A functional genomics tool for the Pacific bluefin tuna: Development of a 44K oligonucleotide microarray from whole-genome sequencing data for global transcriptome analysis." *Gene* 576 (2):603-609. doi: 10.1016/j.gene.2015.10.023.

York, P. H, D. J. Booth, T. M. Glasby, and B. C. Pease. 2006. "Fish assemblages in habitats dominated by Caulerpa taxifolia and native seagrasses in southeastern Australia. ." *Marine Ecology Progress* 223–224.

You, Xinxin, Chao Bian, Qijie Zan, Xun Xu, Xin Liu, Jieming Chen, Jintu Wang, Ying Qiu, Wujiao Li, Xinhui Zhang, Ying Sun, Shixi Chen, Wanshu Hong, Yuxiang Li, Shifeng Cheng, Guangyi Fan, Chengcheng Shi, Jie Liang, Y. Tom Tang, Chengye Yang, Zhiqiang Ruan, Jie Bai, Chao Peng, Qian Mu, Jun Lu, Mingjun Fan, Shuang Yang, Zhiyong Huang, Xuanting Jiang, Xiaodong Fang, Guojie Zhang, Yong Zhang, Gianluca Polgar, Hui Yu, Jia Li, Zhongjian Liu, Guoqiang Zhang, Vydianathan Ravi, Steven L. Coon, Jian Wang, Huanming Yang, Byrappa Venkatesh, Jun Wang, and Qiong Shi. 2014. "Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes." *Nature Communications* 5:5594. doi: 10.1038/ncomms6594

Yu, H., J. Lindsay, Z. P. Feng, S. Frankenberg, Y. Hu, D. Carone, Geoff Shaw, Andrew J. Pask, Rachel O'Neill, Anthony T. Papenfuss, and M. B. Renfree. 2012. "Evolution of coding and non-coding genes in HOX clusters of a marsupial." *BMC Genomics* 13 (1):1.

Zakany, J., and D. Duboule. 2007. "The role of Hox genes during vertebrate limb development." *Current Opinion in Genetics & Development* 17 (4):359-366. doi: 10.1016/j.gde.2007.05.011.

Zerbino, D. R., and E. Birney. 2008. "Velvet: Algorithms for de novo short read assembly using de Bruijn graphs." *Genome Research* 18 (5):821-829. doi: 10.1101/gr.074492.107.

Zuckerkandl, E., and L. Pauling. 1965. "Evolutionary Divergence and Convergence in Proteins." In *Evolving Genes and Proteins*, edited by V. Bryson and H. J. Vogel, 97–166. New York City, New York, USA: Academic Press.