EVOLUTIONARY DIVERGENCE OF A *HOXA2B* HINDBRAIN ENHANCER IN SYNGNATHIDS MIMICS RESULTS OF FUNCTIONAL ASSAYS

Allison M. Fuiten[1,2] and William A. Cresko[1]

1. Institute of Ecology and Evolution, University of Oregon, Eugene, OR 97403
2. Present address: Department of Dermatology, Oregon Health and Science University, Portland, OR 97239

Correspondence should be addressed to W. A. C. (wcresko@uoregon.edu)

Fuiten ORCID: 0000-0003-3926-5455

Cresko ORCID: 0000-0002-3496-8074

ABSTRACT

Hoxa2 genes provide critical patterning signals during development, and their regulation and function have been extensively studied. We report a previously uncharacterized significant sequence divergence of a highly conserved hindbrain *hoxa2b* enhancer element in the family syngnathidae (pipefishes, seahorses, pipehorses, seadragons). We compared the *hox cis*-regulatory element variation in the Gulf pipefish and two species of seahorse against eight other species of fish, as well as human and mouse. We annotated the *hoxa2b* enhancer element binding sites across three species of seahorse, four species of pipefish, and one species of ghost pipefish. Finally, we performed *in situ* hybridization analysis of *hoxa2b* expression in Gulf pipefish embryos. We found that all syngnathid fish examined share a modified rhombomere 4 *hoxa2b* enhancer element, despite the fact that this element has been found to be highly conserved across all vertebrates examined previously. Binding element sequence motifs and spacing between binding elements have been modified for the *hoxa2b* enhancer in several species of pipefish and seahorse, and that the loss of the Prep/Meis binding site and further space shortening happened after ghost pipefish split from the rest of the syngnathid clade. We showed that 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. Our findings highlight the benefits of studying highly derived, diverse taxa for understanding of gene regulatory evolution and support the hypothesis that natural mutations can occur in deeply conserved pathways in ways potentially related to phenotypic diversity.

KEYWORDS: *hox* genes, hindbrain, enhancers, neural crest, gene regulation, evolution of cis-regulatory elements

DECLARATIONS: none

FUNDING: This work would not have been possible without multiple sources of generous support: grant from the National Science Foundation Doctoral Dissertation Improvement Grant in the Directorate for Biological Sciences (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.

CONFLICTS OF INTEREST/COMPETING INTERESTS: not applicable

ETHICS APPROVAL: The care and use of animals were in accordance under the IACUC protocol (ASSURANCE NUMBER A-3009-01).

CONSENT TO PARTICIPATE: not applicable

CONSENT FOR PUBLICATION: not applicable

AVAILABILITY OF DATA AND MATERIALS: Upon publication, all raw sequencing data described in this study will be available via the NCBI Genbank. *In situ* images will be hosted by the Cresko Lab web server (http://creskolab.uoregon.edu) and distributed by the Cresko Laboratory GitHub account.

CODE AVAILABILITY: not applicable

AUTHORS' CONTRIBUTIONS: Allison Fuiten and Bill Cresko designed this study. Allison Fuiten performed experiments and data analysis. Interpretation of results was by Allison Fuiten and Bill Cresko. Allison Fuiten wrote the original draft, and Bill Cresko reviewed and edited the manuscript. Both authors read and approved the final manuscript.

ACKNOWLEDGEMENTS: We wish to thank S. Bassham, C. Small, M. Currey, A. Amores, Y.L. Yan, J. Poslethwait for providing feedback on the study design and results, as well as other members of the Cresko and Postlethwait labs, H. Mason-Jones and E. Rose for help with obtaining Gulf pipefish embryos, A. Jones and A. Anderson for sending robust ghost pipefish, messmate pipefish, and dwarf seahorse tissue samples from the robust ghost pipefish, messmate pipefish, and dwarf seahorse, and A. Bentley for sending bluestripe pipefish and sculptured pipefish tissue samples from the KU fish tissue collection.

INTRODUCTION

Hox genes are a group of core developmental genes present in all animals. These genes 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). 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) (Supplemental Figure 1a). The vertebrate *Hox* genes are organized into 13 paralogous groups that span these gene clusters (Scott 1992). *Hox* genes exhibit collinearity of expression along the body axis to confer positional identity information, with the vertebrate hindbrain expressing *Hox* genes in paralogous groups 1 through 4 during development (Alexander et al. 2009; Lumsden and Krumlauf 1996; Parker et al. 2016; Tumpel et al. 2009).

Despite the large amount of body plan diversity found in animals, most studies to date have documented that *Hox* genes have maintained a great level of conservation throughout the animal kingdom both in terms of sequence and function (reviewed in (Burglin and Affolter 2016; Gehring et al. 1994; Holland 2013)). This level of conservation in *Hox* genes has been hypothesized to occur because major changes in coding regions of *Hox* genes are likely detrimental to the development of the organism due to extensive antagonistic pleiotropy, and mutations in such core developmental regulators are unlikely to contribute to evolution over short time scales (Carroll 2008; Hoekstra and Coyne 2007; Stern 2000). 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 evolve despite developmental constraints imposed by the antagonistic pleiotropy of the hox genes themselves (Raff 2012; Wilkins 2002). As a result, we might predict that regulation of *Hox* genes may contribute to macroevolution—especially of body plan traits.

Of particular interest for this study, the *hoxa2* gene is expressed in the vertebrate hindbrain during development (first described by (Prince and Lumsden 1994)). The hindbrain of all jawed vertebrates is organized into repeated morphological units called rhombomeres, which form through a progression of segmentation during early development (Kiecker and Lumsden 2005; Lumsden 2004). Processes that include the formation of cytoskeletal barriers, cell adhesion and repulsion keep each rhombomere a distinctive unit. As a consequence, each rhombomere contains a separate population of cells that follow different developmental pathways and neurons that are rhombomere specific (reviewed by (Parker et al. 2016; Parker and Krumlauf 2020)). Rhombomeres are a source of cranial neural crest cells and are important regulators for craniofacial and nerve development (reviewed in (Parker et al. 2016; Parker and Krumlauf 2020)). Experimental manipulation of the anterior *Hox* genes have led to cranial phenotypes (Minoux and Rijli 2010; Santagati and Rijli 2003; Trainor and Krumlauf 2000; Trainor and Krumlauf 2001). *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; Parker et al. 2014; Santagati and Rijli 2003). Experimental manipulation of the *hoxa2* gene has been shown to cause various craniofacial phenotypes. Loss-of-function experiments of *Hoxa2* in mice (*Mus musculus*), in *hoxa2b* and *hoxb2a* in zebrafish (*Danio rerio*), and *hoxa2a* and *hoxa2b* in Nile tilapia (*Oreochromis niloticus*) led to duplications of jaw elements (Gendron-Maguire et al. 1993; Hunter and Prince 2002; Le Pabic et al. 2010; Rijli et al. 1993; Santagati et al. 2005).

Gain-of-expression experiments with *hoxa2* led to repression of jaw formation in mice, frog (*Xenopus laevis*), and chicken (*Gallus gallus*) (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; McEllin et al. 2016; Nonchev et al. 1996a; Nonchev et al. 1996b; Parker et al. 2014; Parker et al. 2019a; Parker et al. 2019b; Tumpel et al. 2007; Tumpel et al. 2006). This list described *cis*-regulatory factors 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 et al. 2016; Tumpel et al. 2009). 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. The fugu *hoxa2a* gene has no expression in rhombomere 4 while fugu *hoxa2b* is strongly expressed in rhombomere 4. Sequential conversions of a series of binding sites in the functional fugu *hoxa2b* rhombomere 4 enhancer into paralogous sites derived from the nonfunctional fugu *hoxa2a* enhancer that were fused to a *lacZ* reporter gene and electroporated into chick embryos showed strong to low level of reporter activity in rhombomere 4, depending on the binding sites converted (Tumpel et al. 2006). ~~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).~~ Previous research using chick embryo electroporation and transgenic mouse embryo assays have also reported that site directed mutagenesis on any one of these enhancer binding sites resulted in reduced efficiency of expression of *hoxa2* in rhombomere 4 (Tumpel et al. 2007).

~~Despite the documented deep conservation of the regulatory elements a key question remains; is this finding of a deeply conserved *hoxa2* regulatory element a consequence of the choice of organisms studied?~~ We choose to further examine deeply conserved *hoxa2* regulatory element in the family Syngnathidae, which includes species of pipefish, seahorses, pipehorses, and seadragons. This charismatic teleost family is 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) (Figure 1). 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; Neutens et al. 2014; Wilson and Rouse 2010). The highly modified syngnathid skull results from modified cranial bones in the ethmoid region and Meckel's cartilage (Leysen et al. 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.

We previously reported the *Hox* cluster genes in the Gulf pipefish (*Syngnathus scovelli*), and subsequently the same complement of *Hox* genes have been reported in two species of seahorse (Lin et al. 2016; Lin et al. 2017; Small et al. 2016) (Supplemental Figure 1b). In this study, we compare the *Hox cis*-regulatory elements in syngnathid fish against other vertebrates. We report a previously uncharacterized significant sequence divergence of the hindbrain rhombomere 4 *hoxa2b* enhancer element. We report how the rhombomere 4 enhancer binding sites of

the *hoxa2b* gene are modified in syngnathid fish, and infer possible downstream morphological consequences to the sequence modification of this enhancer element. 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. Our findings highlight the benefits of studying highly derived and diverse taxa for understanding of gene regulatory structure and evolution and support the hypothesis that natural mutations can occur in these deeply conserved pathways in ways potentially related to phenotypic diversity.

METHODS AND MATERIALS

Identification of non-coding elements

CNEs were identified using mVISTA analyses based on levels of sequence conservation within *Hox* clusters across threespine stickleback (*Gasterosteus aculeatus*), fugu (*Takifugu rubripes*), medaka (*Oryzias latipes*), Pacific bluefin tuna (*Thunnus orientalis*), lined seahorse (*Hippocampus erectus*), tiger tail seahorse (*Hippocampus comes*), Gulf pipefish, blue-spotted mudskipper (*Boleophthalmus pectinirostris*), cod (*Gadus morhua*), zebrafish, spotted gar (*Lepisosteus oculatus*), mouse, and human (*Homo sapiens*) (Brudno et al. 2003a; Brudno et al. 2003b; Frazer et al. 2004; Mayor et al. 2000). Sequences for zebrafish, spotted gar, mouse, and human were downloaded from Ensembl. Pacific bluefin tuna sequence was extracted from the Pacific bluefin tuna genome ((Yasuike et al. 2016); http://nrifs.fra.affrc.go.jp/ResearchCenter/5_AG/genomes/Tuna_DNAmicroarray/index.html). Cod sequence was extracted from the cod genome ((Torresen et al. 2017); https://figshare.com/articles/Transcript_and_genome_assemblies_of_Atlantic_cod/3408247). Gulf pipefish sequence was extracted from the Gulf pipefish genome ((Small et al. 2016); https://creskolab.uoregon.edu/pipefish/). The lined seahorse sequence was extracted from the lined seahorse genome ((Lin et al. 2017); NCBI with the project accession PRJNA347499). The tiger tail seahorse sequence was extracted from the tiger tail seahorse genome ((Lin et al. 2016); NCBI with the project accession PRJNA314292). The blue-spotted mudskipper sequence was extracted from the blue-spotted mudskipper genome ((You et al. 2014); NCBI with the project accession PRJNA232434). The fugu sequences were retrieved from Genbank ((Lee et al. 2006); Genbank accessions DQ481663–9). The medaka sequences were retrieved from Genbank ((Kurosawa et al. 2006); AB232918–24). The threespine stickleback sequences were from BAC clones, which were make available by Angel Amores. Sequences were softmasked using RepeatMasker. Threespine stickleback, medaka, and fugu were set as the reference sequences 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. Any significant peak loss was confirmed to be not dependent on the teleost set as a reference nor on whether shuffle-LAGAN or LAGAN algorithm was used.

All conserved noncoding sequences annotated within the Gulf pipefish *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 miRNA coding 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.

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.

Sequence alignments and identification of enhancer binding sites

Hoxa2, *hoxa2b*, and *hoxa2a* sequences from coelacanth (*Latimeria chalumnae*), anole (*Anolis carolinensis*), chicken, zebrafish, spotted gar, mouse, and human were downloaded from Ensembl. The Australian ghostshark (*Callorhinchus milii*) *hoxa2* sequence was retrieved from Genbank. The tamar wallaby (*Notamacropus eugenii*) *hoxa2* sequence was retrieved from Genbank. The fugu *hoxa2a* and *hoxa2b* sequences were retrieved from Genbank ((Lee et al. 2006); Genbank accessions DQ481663–9). The medaka *hoxa2a* and *hoxa2b* sequences were retrieved from Genbank ((Kurosawa et al. 2006); AB232918–24). The threespine stickleback *hoxa2b* sequence was from BAC clones, which were make available by Angel Amores. Pacific bluefin tuna *hoxa2b* sequence was extracted from the Pacific bluefin tuna genome ((Yasuike et al. 2016); http://nrifs.fra.affrc.go.jp/ResearchCenter/5_AG/genomes/Tuna_DNAmicroarray/index.html). The blue-spotted mudskipper *hoxa2b* sequence was extracted from the blue-spotted mudskipper genome ((You et al. 2014); NCBI with the project accession PRJNA232434). The Gulf pipefish *hoxa2b* sequence was extracted from the Gulf pipefish genome ((Small et al. 2016); https://creskolab.uoregon.edu/pipefish/). The lined seahorse *hoxa2b* sequence was extracted from the lined seahorse genome ((Lin et al. 2017); NCBI with the project accession PRJNA347499). The tiger tail seahorse *hoxa2b* sequence was extracted from the tiger tail seahorse genome ((Lin et al. 2016); NCBI with the project accession PRJNA314292).

In addition to the Gulf pipefish, lined seahorse, and tiger tail seahorse genomic sequences, degenerate primers were designed and used to sequence the *hoxa2b* enhancer region for the dwarf seahorse (*Hippocampus zostrae*), the messmate pipefish (*Corythoichthys haematopterus*), bluestripe pipefish (*Doryrhamphus excisus*),

sculptured pipefish (*Choeroichthys sculptus*), and the robust ghost pipefish (*Solenostomus cyanopterus*) (Table 1). 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 dwarf seahorse and messmate pipefish bring additional taxonomic sampling from the Syngnathinae subfamily of Syngnathidae. The sculptured pipefish and bluestripe pipefish are members of the Nerophinae subfamily of Syngnathidae. The robust ghost pipefish is a species from the immediate outgroup to Syngnathidae. This additional taxonomic sampling provided further insight into the change of this enhancer element in this teleost fish family (Figure 1).

The sequences were aligned using MUSCLE through the Geneious software (Edgar 2004). Alignments were corrected manually. 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. Binding site sequences for Pbx/Hox and Prep/Meis were obtained from (Berthelsen et al. 1998; Ferretti et al. 2005; Ferretti et al. 2000; Tumpel et al. 2007).

Whole mount *in situ* hybridization analysis

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. Antisense digoxigenin (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'-gcgcctccttgtacgcacgcgcacctccacccgccctcgtcgtacacgtgcatcagtgacgtgtaccaggaatcctctgatgagggttacctggccgtacccacctgcagcgcggtgacttatcacatggcgccagcctataactcggcgccaaaagccccgctggtggctgactacggcgtgggggggagtctacgccccacaggccaccttcccggaccggaagtcagtggcggcgtacgccttggactccctccgcgtggcccctccgctcacacc-3'. For *hoxa2a*, the probe sequence used was 5'-tggaatccacgcagcaggtccacaatagcagctcggcgagctttgctgctgcaccgctgaacagcaatgagaaaaatctgaaacattttcccaacccgtcacccactgttcccggctgcgtgtcaacaatgggcccaggctcggcatccgtgccggacaatggcgacagtcccccagctttggatgtttctatacacgacttccaagctttctcgtcggattcctgcttgcaactmtccgacgctgcctcgccgagcttgtctgaatcgctggacagtcccgtgg-3'. For *hoxa2b*, the probe sequence used was 5'-gcgaaggacctttggaagagcagccagccaaggggcagaggtatttccaggaaaattgtttcaattcacaacattgtcctaatagccacaatggsgacaatgattcgacttgtgcataagtgagaaaaatgccaaacatcttccggactgcgctcccaccacggctcccttctgtgcgcccgaaataggcccggagaataatatytttcccacgtctcgcacagtgaatactccccggatttggacgcctctttgcgggagcttcctcgagcatcctcgttctcgcaagactggtccgattcaactccgct-3'.

Whole-mount *in situ* hybridization analyses were performed as described in Thisse and Thisse (2008). One to five Gulf pipefish 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 Gulf pipefish were collected in Tampa Bay, Florida on May 5, 2017. Breeding tanks were set up at the University of Tampa. Embryos from pregnant male pipefish were collected at 1, 2, 3, 4, 5, 6 dpf. Additionally, embryos from wild caught pregnant male Gulf pipefish were collected. Pregnant male pipefish were euthanized in 0.017% Tricaine-S, and then fixed in 4% paraformaldehyde either overnight at 4°C or for 5 hours at room temperature, embryos were extracted from the male brood pouch 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 divergence of a *hoxa2b* enhancer is shared across syngnathid fish

We used lined seahorse, tiger tail seahorse and Gulf pipefish as the syngnathid representatives and compared their conservation noncoding element (CNE) content to percomorph teleost fish (threespine stickleback, fugu, medaka, Pacific bluefin tuna), non-percomorph teleost fish (blue-spotted mudskipper, cod, zebrafish), non-teleost fish (spotted gar), and two non-fish vertebrates (mouse and human) in the *Hox* clusters. From examining the VISTA plots, there were five instances of syngnathid-specific peak losses of *Hox* cluster CNEs among the species examined.

Two of the five peak losses were independent losses of *Hox* cluster microRNAs—*mir19b* and *mir10a*—that are reported to be lost convergently in other teleosts. We previously documented *mir196b* as an independent loss in the Gulf pipefish (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 (Supplemental Figure 2). *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 (Supplemental Figure 3).

We also identified two syngnathid specific CNE peak losses in *HoxCa*—one between *hoxc8a* and *hoxc6a* and another between *hoxc4a* and *hoxc3a* (Supplemental Figures 4 and 5). Both of these CNEs were only found among the acanthomorph fish examined (cod, blue-spotted mudskipper, fugu, medaka, Pacific bluefin tuna, and threespine stickleback). It is unknown whether these CNEs serve a functional role or are merely the result of neutral sequence conservation ~~(Supplemental Figure 6a)~~.

Our most surprising finding was of a significant sequence divergence of a highly conserved noncoding element that is shared across the included syngnathid species (Figure 2 and Supplemental Figure 6). 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 (~~Supplemental~~ Figure 6d). 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 fugu *hoxa2a* paralog when fused to a lacZ reporter gene and electroporated into chick embryos ~~in fish~~, despite exhibiting the binding motifs of this enhancer, apparently does not drive expression of the *hoxa2a* gene in rhombomere 4 the chick hindbrain (Tumpel et al. 2006).

Significant sequence changes to the Pbx/Hox and Prep/Meis syngnathid binding sites

The *hoxa2b* enhancer element comprises four Pbx/Hox binding sites and one Prep/Meis binding site. One of the four Pbx/Hox binding sites is 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 3) (Parker et al. 2014; Tumpel et al. 2007; Tumpel et al. 2006). In order to further examine the degree of conservation of this binding site among vertebrates and determine the pattern of divergence in syngnathids, we examined the enhancer element binding motifs across Vertebrata using the Australian ghostshark, coelacanth, anole, chicken, tamar wallaby, human, mouse, spotted gar, zebrafish, fugu, medaka, threespine stickleback, Pacific bluefin tuna, and blue-spotted mudskipper, along with syngnathid species (Figure 4, Table 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 et al. 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 et al. 1996; Manzanares et al. 2001).

We found that teleost fish have the 5'-TGAT-3' motif in the Pbx/Hox 1 site, with the exception of the blue-spotted mudskipper, the bluestripe pipefish, the sculptured pipefish, the messmate pipefish, the Gulf pipefish, and the dwarf seahorse. Teleost fish, except for the blue-spotted 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 robust 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 robust 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.~~ For the first four bases in the Pbx/Hox 3 binding site, all teleost fish examined have sequences 5'-TGGC-3' or 5'-TGGA-3', instead of following the more common sequence pattern of 5'-TGAT-3' or the 5'-NNAT-3' described by (Ferretti et al. 2005). Based on alignments, the second half of Pbx/Hox 3 binding sequence appeared to have been lost in the syngnathid species. The robust ghost pipefish and Syngnathidae fish have sequences 5'-CGAT-3' in the first four bases in the Pbx/Hox 4 binding sites instead of the 5'-TGAT-3' found in the other examined teleost fish. For the last four bases in the Pbx/Hox 4 binding site, examined teleost fish had either 5'-

NNGT-3', 5'-NNTT-3', or 5'-NNCT-3' (with the exception of the Pacific bluefin tuna) instead of the more common the 5'-NNAT-3' sequence motif (Table 2).

We also found that the Prep/Meis binding site was conserved across taxa examined, including in the robust ghost pipefish, with the exception of the syngnathid species. Members of the Syngnathinae subfamily (lined seahorse, tiger tail seahorse, dwarf seahorse, messmate pipefish, and Gulf pipefish) were missing the Prep/Meis binding site. Based on alignments, it appeared that the two species from the Nerophinae subfamily, bluestripe pipefish and sculptured pipefish had only the "ACA" nucleotides remaining from this binding site (Figure 4).

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 3). The intron lengths spanned from 924 bases in anole to 417 bases in fugu. The intron length in syngnathid species were all less than 275 bases.

The spacing between each of the binding sites was also shorter in the syngnathid species relative to the other species (Table 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 3).

Robust 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 robust 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 robust ghost pipefish which had a longer spacing of 356 bases (Table 3).

Loss of Prep/Meis and further space shortening evolved after robust 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 1).

We found that the robust ghost pipefish had all five binding sites for this enhancer element and an intermediately sized intron of 350 bases (Tables 2 and 3). This observation can be interpreted as that the loss of the

Prep/Meis binding site happened after the robust ghost pipefish diverged from Syngnathidae clade (Figure 1). The spacing of the motifs were already shortening before robust ghost pipefish split from Syngnathidae, but more extreme shortening of the binding site spacing after robust 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 were knocked out in chicken and mouse using site directed mutagenesis. When these binding sites were knocked out, there was reduced *Hoxa2* expression in rhombomere 4, demonstrating that these binding sites were necessary for the function of the enhancer. 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 (Figure 5a). 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 (Figures 5b–d). At five days post fertilization, *hoxa2b* is expressed in rhombomeres 3, 4, and 5 and in the tailbud (Figures 5e and 5f). We found that *hoxa2a* was expressed in the hindbrain during development. At three, four, and five days post fertilization, *hoxa2a* was expressed in rhombomeres 2, 3, and 4 and pharyngeal arch 2 (Figures 5m–p). Expression of *hoxa2b* appears to be reduced in rhombomere 4 relative to neighboring rhombomeres 3 and 5 in Gulf pipefish. This relative reduction in expression of *hoxa2b* in rhombomere 4 aligns with previous functional tests (Tumpel et al. 2007).

DISCUSSION

Significant sequence divergence of the *hoxa2b* R4 enhancer is a synapomorphy of syngnathid fish

Slight natural variants of this *hoxa2b* enhancer element has been previously reported in Tumpel et al. 2006, Tumpel et al. 2007 and Parker et al. 2014, but variation in 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 as we show specifically for syngnathid fish.

Syngnathid fish all share a much more 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. The robust 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 robust ghost pipefish split from Syngnathidae. Interestingly, the length of the spacing of the binding sites in the robust 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 6).

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

Despite the significant sequence divergence, it is possible that functionality of the regulatory elements is maintained. To address this question, we examined expression of the *hoxa2b* gene in rhombomere 4 and found that it is lower relative to the surrounding rhombomeres. ~~This change in expression is consistent with the loss of the element having functional effects through changes to cranial neural crest cells.~~ Other studies have reported changes to regulatory elements that have resulted in interesting phenotypic modifications to body plans (reviewed in (Carroll 2008; Gehrke and Shubin 2016; Rebeiz and Tsiantis 2017; Wray 2007)). 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 of deeply conserved 'master regulatory genes' 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 (Gendron-Maguire et al. 1993; Hunter and Prince 2002; Le Pabic et al. 2010; 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. In addition, 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 (Brown 2010; Kimmel et al. 2017; Leysen et al. 2010). ~~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 sequence divergence of the *hoxa2b* enhancer element is tied to the highly modified skull in syngnathid fish, but more functional work will need to be done with this enhancer for rhombomere 4. Previous functional studies have demonstrated different roles, with various levels of redundancy, between *Hox* paralog group 2 genes in teleost (Hunter and Prince 2002; Le Pabic et al. 2010). Future studies exploring the functional roles within the three *Hox* paralog group 2 genes (*hoxa2a*, *hoxa2b*, and *hoxb2a*) will need to be done to further understand the functional consequences of the unique change in this enhancer in syngnathid fish.

In addition to bones, rhombomere 4 is important for nerve cell and Mauthner cell development.

Intriguingly, syngnathids have reportedly lost their Mauthner cells (Benedetti et al. 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 argue against this connection~~. Although we cannot infer direct causation of the extreme sequence changes in this regulatory element of *hoxa2b* and the highly derived craniofacial morphologies of syngnathids, the possible connections are intriguing and beg further study.

CONCLUSIONS

Making use of the increasing number of 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 sequence divergence 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. The possible connections between divergence of regulation of highly conserved developmental genes and evolutionary are fascinating, but of course subsequent studies will be necessary to fully challenge these causative hypotheses. 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.

FIGURE AND TABLE LEGENDS

Figure 1: The Syngnathidae family contain morphologically diverse fish encompassing pipefish, seahorses, seadragons and pipehorses. Illustrations depict representative species: a) dwarf seahorse (*Hippocampus zostrae*), b) tiger tail seahorse (*Hippocampus comes*), c) lined seahorse (*Hippocampus erectus*), d) Gulf pipefish (*Syngnathus scovelli*), e) leafy seadragon (*Phycodurus eques*), f) weedy seadragon (*Phyllopteryx taeniolatus*), g) messmate pipefish (*Corythoichthys haematopterus*), h) sculptured pipefish (*Choeroichthys sculptus*), i) bluestripe pipefish (*Doryrhamphus excisus*), j) robust ghost pipefish (*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. Clades sampled in this study are highlighted with blue boxes. Cladogram based on molecular phylogeny published by Hamilton et al. 2017.

Figure 2: A conserved non-coding element is not detectable in the pipefish *HoxAb* cluster. a) One CNE present in other teleost fish and mammals is missing from the intron of *hoxa2b* in the Gulf pipefish, tiger tail seahorse and lined seahorse assemblies (red arrows). b) Syngnathid fish are not missing CNEs from the intron of *hoxa2a* in the Gulf pipefish, tiger tail seahorse and lined seahorse assemblies. Exons are highlighted in blue, CNEs in pink. The reference, Gac, is threespine stickleback; Tru, fugu; Ola, medaka; Tor, Pacific bluefin tuna; Hco, tiger tail seahorse; Her, lined seahorse; Ssc, Gulf pipefish; Bpe, blue-spotted mudskipper; Gmo, cod; Dre, zebrafish; Loc, spotted gar; Mmu, mouse; Hsa, human. Red arrows indicate missing CNE in syngnathid fish.

Figure 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.

Figure 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.

Figure 5: *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–p show expression of *hoxa2a* in Gulf pipefish embryos. Images q–t 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; o) *hoxa2a* ~5dpf left lateral; p) *hoxa2a* ~5dpf dorsal; q) *hoxa2a* with *krox20a* 3dpf lateral; r) *hoxa2a* with *krox20a* 4dpf dorsal; s) *hoxa2a* with *krox20a* ~5dpf left lateral; t) *hoxa2a* with *krox20a* ~5dpf dorsal. *Krox20a* marks rhombomeres 3 and 5. R3, Rhombomere 3; R5, Rhombomere 5; PA2, Pharyngeal Arch 2.

Figure 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 robust ghost pipefish. d) Cladogram shows evolutionary relationships between vertebrates included in the CNE analysis. 1 vertebrate CNE was uniquely, significantly diverged in the syngnathid fish.

Table 1: Degenerate primer pairs used on syngnathid species for *hoxa2b*.

Table 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.

Table 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.

Supplemental Figure 1: Evolution of *Hox* complex. a) Evolutionary timing of *Hox* complex duplications are denoted on the animal phylogeny based on (Carroll et al. 2013), with updates from (Pascual-Anaya et al. 2018; Ravi et al. 2009). Dashed arrow indicates current uncertainty where the second vertebrate *Hox* cluster duplication occurred relative to agnathans. b) A cartoon of the previously reported *Hox* clusters in Gulf pipefish, lined seahorse and tiger tail seahorse with boxes representing genes and circles representing microRNAs arranged along chromosome segments oriented left to right 5' to 3'. The hollow box represents the *hoxa7a* pseudogene as described in Small et al. 2016.

Supplemental Figure 2: VISTA plots for the *HoxB* clusters with threespine stickleback *HoxBa* set as reference sequence. Exons are highlighted in blue, CNEs in pink, 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, fugu; 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 specific peak losses are in red boxes.

Supplemental Figure 3: VISTA plots for the *HoxB* clusters with threespine stickleback *HoxBb* set as reference sequence. Exons are highlighted in blue, CNEs in pink, 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, fugu; 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 specific peak losses are in red boxes.

Supplemental Figure 4: VISTA plots for the *HoxC* clusters. ~~**with threespine stickleback *HoxCa* set as reference sequence.**~~ Exons are highlighted in blue, CNEs in pink. ~~LAGAN alignment was used. The reference,~~ Gac, is the threespine stickleback; Tru, fugu; 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 specific peak losses are in red boxes. a) Gac is set as the reference with Shuffle-LAGAN alignment used. b) Ola is set as the reference with Shuffle-LAGAN alignment used. c) Tru is set as the reference with Shuffle-

LAGAN alignment used. d) Gac is set as the reference with LAGAN alignment used. e) Ola is set as the reference with LAGAN alignment used. f) Tru is set as the reference with LAGAN alignment used.

Supplemental Figure 5: VISTA plots for the *HoxC* clusters. ~~**with threespine stickleback *HoxCa* set as reference sequence.**~~ Exons are highlighted in blue, CNEs in pink. ~~LAGAN alignment was used. The reference,~~ Gac, is the threespine stickleback; Tru, fugu; 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 specific peak losses are in red boxes. a) Gac is set as the reference with Shuffle-LAGAN alignment used. b) Ola is set as the reference with Shuffle-LAGAN alignment used. c) Tru is set as the reference with Shuffle-LAGAN alignment used. d) Gac is set as the reference with LAGAN alignment used. e) Ola is set as the reference with LAGAN alignment used. f) Tru is set as the reference with LAGAN alignment used.

~~**Supplemental Figure 6: Distribution of syngnathid-specific losses CNEs cataloged within the syngnathid *Hox* clusters.** Cladograms show evolutionary relationships between vertebrates included in the CNE analysis. 2 acanthomorph CNEs were uniquely lost in the syngnathid clade.~~

Supplemental Figure 6: A conserved non-coding element is not detectable in the pipefish *HoxAb* cluster with different species set as the reference and mVISTA alignment algorithms. Exons are highlighted in blue, CNEs in pink. Gac, is the threespine stickleback *HoxAb* sequence; Tru, fugu *HoxAb* sequence; Ola, medaka *HoxAb* sequence; Tor, tuna *HoxAb* sequence; Ssc, pipefish *HoxAb* sequence; Hco, tiger tail seahorse *HoxAb* sequence; Her, lined seahorse *HoxAb* sequence; Bpe, mudskipper *HoxAb* sequence; Gmo, cod *HoxAb* sequence; Dre, zebrafish *HoxAb* sequence; Loc, spotted gar *HoxA* sequence; Mmu, mouse *HoxA* sequence; Hsa, human *HoxA* sequence. Red arrows indicate missing CNE in syngnathid fish. a) Threespine stickleback (gac) is set as the reference with LAGAN alignment used. b) Fugu (tru) is set as the reference with LAGAN alignment used. c) Medaka (ola) is set as the reference with LAGAN alignment used. d) Fugu (tru) is set as the reference with Shuffle-LAGAN alignment used. e) Medaka (ola) is set as the reference with Shuffle-LAGAN alignment used.

References

Alexander T, Nolte C, Krumlauf R (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

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

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

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

Brown R (2010) Craniofacial development in pipefish: A morphological and molecular analysis. University of Oregon

Brudno M, Do CB, Cooper GM, Kim MF, Davydov E (2003a) LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA Genome Research 13:721-731 doi:10.1101/gr.926603

Brudno M, Malde S, Poliakov A, Do CB, Couronne O, Dubchak I (2003b) Glocal alignment: finding rearrangements during alignment Bioinformatics 19:i54-i62 doi:10.1093/bioinformatics/btg1005

Burglin TR, Affolter M (2016) Homeodomain proteins: an update Chromosoma 125:497-521 doi:10.1007/s00412-015-0543-8

Carroll SB (1995) Homeotic genes and the evolution of arthropods and chordates Nature 376:479-485 doi:10.1038/376479a0

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

Carroll SB, Grenier JK, Weatherbee SD (2013) From DNA to diversity: molecular genetics and the evolution of animal design. Blackwell Pub., Malden, Massachusetts, USA

Chan SK, Ryoo HD, Gould A, Krumlauf R, Mann RS (1997) Switching the in vivo specificity of a minimal Hox-responsive element Development 124:2007-2014

Chan YF et al. (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer Science 327:302-305 doi:10.1126/science.1182213

Chang CP, Shen WF, Rozenfeld S, Lawrence HJ, Largman C, Cleary ML (1995) Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins Genes and Development 9:663-674

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput Nucleic Acids Research 32:1792-1797 doi:10.1093/nar/gkh340

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

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

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics Nucleic Acids Research 32:W273-W279 doi:10.1093/nar/gkh458

556 Gavalas A, Davenne M, Lumsden A, Chambon P, Rijli FM (1997) Role of Hoxa-2 in axon
557 pathfinding and rostral hindbrain patterning Development 124:3693-3702
558 Gehring WJ, Affolter M, Bürglin T (1994) Homeodomain proteins Annual review of
559 biochemistry 63:487-526
560 Gehrke AR, Shubin NH (2016) Cis-regulatory programs in the development and evolution of
561 vertebrate paired appendages Seminars in Cell and Developmental Biology 57:31-39
562 doi:10.1016/j.semcdb.2016.01.015
563 Gendron-Maguire M, Mallo M, Zhang M, Gridley T (1993) Hoxa-2 mutant mice exhibit
564 homeotic transformation of skeletal elements derived from cranial neural crest Cell
565 75:1317-1331
566 Grammatopoulos GA, Bell E, Toole L, Lumsden A, Tucker AS (2000) Homeotic transformation
567 of branchial arch identity after Hoxa2 overexpression Development 127:5355-5365
568 Herald ES (1959) From pipefish to seahorse - a study of phylogenetic relationships Proceedings
569 of the California Academy of Sciences 29
570 Hoegg S, Boore JL, Kuehl JV, Meyer A (2007) Comparative phylogenomic analyses of teleost
571 fish Hox gene clusters: lessons from the cichlid fish Astatotilapia burtoni BMC Genomics
572 8:317 doi:10.1186/1471-2164-8-317
573 Hoekstra HE, Coyne JA (2007) The locus of evolution: evo devo and the genetics of adaptation.
574 Evolution: International Journal of Organic Evolution 61:995-1016
575 Holland PW (2013) Evolution of homeobox genes Wiley Interdisciplinary Reviews:
576 Developmental Biology 2:31-45 doi:10.1002/wdev.78
577 Hunter MP, Prince VE (2002) Zebrafish Hox Paralogue Group 2 Genes Function Redundantly as
578 Selector Genes to Pattern the Second Pharyngeal Arch Developmental Biology 247:367-
579 389 doi:10.1006/dbio.2002.0701
580 Infante CR, Mihala AG, Park S, Wang JS, Johnson KK, Lauderdale JD, Menke DB (2015)
581 Shared Enhancer Activity in the Limbs and Phallus and Functional Divergence of a
582 Limb-Genital cis-Regulatory Element in Snakes Developmental Cell 35:107-119
583 doi:10.1016/j.devcel.2015.09.003
584 Kiecker C, Lumsden A (2005) Compartments and their boundaries in vertebrate brain
585 development Nature Reviews Neuroscience 6:553-564 doi:10.1038/nrn1702
586 Kimmel CB, Small CM, Knope ML (2017) A rich diversity of opercle bone shape among teleost
587 fishes PLoS ONE 12:e0188888 doi:10.1371/journal.pone.0188888
588 Kitazawa T et al. (2015) Distinct effects of Hoxa2 overexpression in cranial neural crest
589 populations reveal that the mammalian hyomandibular-ceratohyal boundary maps within
590 the styloid process Developmental Biology 402:162-174
591 doi:10.1016/j.ydbio.2015.04.007
592 Knoepfler PS, Lu Q, Kamps MP (1996) Pbx-1 Hox heterodimers bind DNA on inseparable half-
593 sites that permit intrinsic DNA binding specificity of the Hox partner at nucleotides 3' to
594 a TAAT motif Nucleic Acids Research 24:2288-2294
595 Krumlauf R (1994) Hox genes in vertebrate development Cell 78:191-201
596 Kurosawa G et al. (2006) Organization and structure of hox gene loci in medaka genome and
597 comparison with those of pufferfish and zebrafish genomes Gene 370:75-82
598 doi:10.1016/j.gene.2005.11.015
599 Le Pabic P, Scemama JL, Stellwag EJ (2010) Role of Hox PG2 genes in Nile tilapia pharyngeal
600 arch specification: implications for gnathostome pharyngeal arch evolution Evolution and
601 Development 12:45-60 doi:10.1111/j.1525-142X.2009.00390.x

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

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

Lin Q et al. (2016) The seahorse genome and the evolution of its specialized morphology Nature 540:395-399 doi:10.1038/nature20595

Lin Q et al. (2017) Draft genome of the lined seahorse, Hippocampus erectus Gigascience 6:1-6 doi:10.1093/gigascience/gix030

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

Lumsden A, Krumlauf R (1996) Patterning the vertebrate neuraxis Science 274:1109-1115

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

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

Manzanares M et al. (2001) Independent regulation of initiation and maintenance phases of Hoxa3 expression in the vertebrate hindbrain involve auto- and cross-regulatory mechanisms Development 128:3595-3607

Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA (2000) VISTA: visualizing global DNA sequence alignments of arbitrary length Bioinformatics 16:1046-1047 doi:10.1093/bioinformatics/16.11.1046

McEllin JA, Alexander TB, Tumpel S, Wiedemann LM, Krumlauf R (2016) Analyses of fugu hoxa2 genes provide evidence for subfunctionalization of neural crest cell and rhombomere cis-regulatory modules during vertebrate evolution Developmental Biology 409:530-542 doi:10.1016/j.ydbio.2015.11.006

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

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

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

Nonchev S et al. (1996a) 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 93:9339-9345

Nonchev S et al. (1996b) Segmental expression of Hoxa-2 in the hindbrain is directly regulated by Krox-20 Development 122:543-554

Parker HJ, Bronner ME, Krumlauf R (2014) A Hox regulatory network of hindbrain segmentation is conserved to the base of vertebrates Nature 514:490-493 doi:10.1038/nature13723

Parker HJ, Bronner ME, Krumlauf R (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 38:526-538 doi:10.1002/bies.201600010

Parker HJ, Bronner ME, Krumlauf R (2019a) An atlas of anterior hox gene expression in the embryonic sea lamprey head: Hox-code evolution in vertebrates Dev Biol 453:19-33 doi:10.1016/j.ydbio.2019.05.001

Parker HJ et al. (2019b) A Hox-TALE regulatory circuit for neural crest patterning is conserved across vertebrates Nat Commun 10:1189 doi:10.1038/s41467-019-09197-8

Parker HJ, Krumlauf R (2020) A Hox gene regulatory network for hindbrain segmentation Curr Top Dev Biol 139:169-203 doi:10.1016/bs.ctdb.2020.03.001

Pascual-Anaya J et al. (2018) Hagfish and lamprey Hox genes reveal conservation of temporal colinearity in vertebrates Nature Ecology and Evolution 2:859-866 doi:10.1038/s41559-018-0526-2

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

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

Raff RA (2012) The shape of life: genes, development, and the evolution of animal form. University of Chicago Press, Chicago, Illinois, USA

Ravi V, Lam K, Tay BH, Tay A, Brenner S, Venkatesh B (2009) Elephant shark (Callorhinchus milii) provides insights into the evolution of Hox gene clusters in gnathostomes Proceedings of the National Academy of Sciences of the United States of America 106:16327-16332 doi:10.1073/pnas.0907914106

Rebeiz M, Tsiantis M (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

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

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

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

Scott MP (1992) Vertebrate homeobox gene nomenclature Cell 71:551-553

Small CM et al. (2016) The genome of the Gulf pipefish enables understanding of evolutionary innovations Genome Biology 17:258 doi:10.1186/s13059-016-1126-6

Stern DL (2000) Perspective: evolutionary developmental biology and the problem of variation. Evolution 54:1079-1091

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

Teske PR, Beheregaray LB (2009) Evolution of seahorses' upright posture was linked to Oligocene expansion of seagrass habitats Biology Letters 5:521-523 doi:10.1098/rsbl.2009.0152

692 Torresen OK et al. (2017) An improved genome assembly uncovers prolific tandem repeats in
693 Atlantic cod BMC Genomics 18:95 doi:10.1186/s12864-016-3448-x
694 Trainor PA, Krumlauf R (2000) Patterning the cranial neural crest: hindbrain segmentation and
695 Hox gene plasticity Nature Reviews Neuroscience 1:116-124 doi:10.1038/35039056
696 Trainor PA, Krumlauf R (2001) Hox genes, neural crest cells and branchial arch patterning
697 Current Opinion in Cell Biology 13:698-705
698 Tumpel S, Cambronero F, Ferretti E, Blasi F, Wiedemann LM, Krumlauf R (2007) Expression of
699 Hoxa2 in rhombomere 4 is regulated by a conserved cross-regulatory mechanism
700 dependent upon Hoxb1 Developmental Biology 302:646-660
701 doi:10.1016/j.ydbio.2006.10.029
702 Tumpel S, Cambronero F, Wiedemann LM, Krumlauf R (2006) Evolution of cis elements in the
703 differential expression of two Hoxa2 coparalogous genes in pufferfish (Takifugu
704 rubripes) Proceedings of the National Academy of Sciences of the United States of
705 America 103:5419-5424 doi:10.1073/pnas.0600993103
706 Tumpel S, Wiedemann LM, Krumlauf R (2009) Hox genes and segmentation of the vertebrate
707 hindbrain Current Topics in Developmental Biology, 88:103-137 doi:10.1016/S0070-
708 2153(09)88004-6
709 Wilkins AS (2002) The evolution of developmental pathways. Sinauer Associates Inc,
710 Sunderland, Massachusetts, USA
711 Wilson NG, Rouse GW (2010) Convergent camouflage and the non-monophyly of ‘seadragons’
712 (Syngnathidae: Teleostei): suggestions for a revised taxonomy of syngnathids Zoologica
713 Scripta 39:551-558 doi:10.1111/j.1463-6409.2010.00449.x
714 Wray GA (2007) The evolutionary significance of cis-regulatory mutations Nature Reviews
715 Genetics 8:206-216 doi:10.1038/nrg2063
716 Yasuike M et al. (2016) A functional genomics tool for the Pacific bluefin tuna: Development of
717 a 44K oligonucleotide microarray from whole-genome sequencing data for global
718 transcriptome analysis Gene 576:603-609 doi:10.1016/j.gene.2015.10.023
719 You X et al. (2014) Mudskipper genomes provide insights into the terrestrial adaptation of
720 amphibious fishes Nature Communications 5:5594 doi:10.1038/ncomms6594
721 https://www.nature.com/articles/ncomms6594#supplementary-information
722