

# An Independent Genome Duplication Inferred from Hox Paralogs in the American Paddlefish—A Representative Basal Ray-Finned Fish and Important Comparative Reference

Karen D. Crow<sup>1,\*</sup>, Christopher D. Smith<sup>1</sup>, Jan-Fang Cheng<sup>2</sup>, Günter P. Wagner<sup>2</sup>, and Chris T. Amemiya<sup>4,5</sup>

<sup>1</sup>Department of Biology, San Francisco State University

<sup>2</sup>Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California

<sup>3</sup>Department of Ecology and Evolutionary Biology, Yale University

<sup>4</sup>Molecular Genetics Program, Benaroya Research Institute at Virginia Mason, Seattle, Washington

<sup>5</sup>Department of Biology, University of Washington

\*Corresponding author: E-mail: crow@sfsu.edu.

**Data deposition:** These sequences can be found at NCBI under the following accession numbers: JX280945–JX280946, JX448769–JX448770

**Accepted:** July 25, 2012

## Abstract

Vertebrates have experienced two rounds of whole-genome duplication (WGD) in the stem lineages of deep nodes within the group and a subsequent duplication event in the stem lineage of the teleosts—a highly diverse group of ray-finned fishes. Here, we present the first full Hox gene sequences for any member of the Acipenseriformes, the American paddlefish, and confirm that an independent WGD occurred in the paddlefish lineage, approximately 42 Ma based on sequences spanning the entire HoxA cluster and eight genes on the HoxD gene cluster. These clusters comprise different HOX loci and maintain conserved synteny relative to bichir, zebrafish, stickleback, and pufferfish, as well as human, mouse, and chick. We also provide a gene genealogy for the duplicated *fzd8* gene in paddlefish and present evidence for the first Hox14 gene in any ray-finned fish. Taken together, these data demonstrate that the American paddlefish has an independently duplicated genome. Substitution patterns of the “alpha” paralogs on both the HoxA and HoxD gene clusters suggest transcriptional inactivation consistent with functional diploidization. Further, there are similarities in the pattern of sequence divergence among duplicated Hox genes in paddlefish and teleost lineages, even though they occurred independently approximately 200 Myr apart. We highlight implications on comparative analyses in the study of the “fin-limb transition” as well as gene and genome duplication in bony fishes, which includes all ray-finned fishes as well as the lobe-finned fishes and tetrapod vertebrates.

**Key words:** *Polyodon spathula*, whole-genome duplication, WGD, rate asymmetry, paralog retention, fin-limb transition.

## Introduction

One of the most challenging problems in evolutionary biology is to understand the types of evolutionary change responsible for generating phenotypic diversity. Gene duplication is widely regarded as the predominant mechanism by which genes with new functions and associated phenotypic novelties arise (Ohno 1970; Holland et al. 1994; Ruddle et al. 1994; Holland and Garcia-Fernandez 1996; Meyer and Schartl 1999;

Lynch and Katju 2004). At the molecular level, duplicate genes provide genetic redundancy that could release one or both gene copies from purifying selection, allowing evolutionary changes to occur while maintaining the ancestral protein function. In this way, gene duplication may be an important genetic mechanism associated with the origin of novel characters (Ohno 1970; Holland et al. 1994; Zhang et al. 2002; Zhang 2003) and diversification of species (Zhou et al. 2001; Scannell et al. 2006, 2007; Semon and Wolfe 2007b). As such, there is

a growing body of evidence implicating genome duplication as a key factor in the evolution of diversity (Werth and Windham 1991; Lynch and Force 2000; Zhou et al. 2001; Postlethwait et al. 2004; Scannell et al. 2006; Roth et al. 2007; Semon and Wolfe 2007a,b), novelty (Holland et al. 1994; Duda and Palumbi 1999; Meyer and Schartl 1999; Zhang et al. 2002), and reduced probability of extinction (Crow and Wagner 2006). However, the types of mutations that contribute to the initial preservation of duplicate genes remain unclear (Lynch and Katju 2004). Several rounds of whole-genome duplication (WGD) have occurred throughout vertebrate evolution (fig. 1), including two genome duplications that preceded the origin of vertebrates and jawed vertebrates (referred to as “2R” for two rounds of duplication) and a third genome duplication (3R) that occurred shortly before the origin of teleosts (Amores et al. 1998; Hawkins et al. 2000; Naruse et al. 2000; Taylor et al. 2003; Christoffels et al. 2004; Hoegg et al. 2004; Jaillon et al. 2004; de Souza et al. 2005; Crow et al. 2006; Schweitzer et al. 2006; Cardoso et al. 2007; Semon and Wolfe 2007a; Salaneck et al. 2008) approximately 285–334 Ma (Vandepoele et al. 2004; Inoue et al. 2005). It has been widely speculated that the extraordinary diversity observed in ray-finned fishes is correlated with the latter and has been referred to as the “teleost-specific genome duplication” (TSGD or 3R).

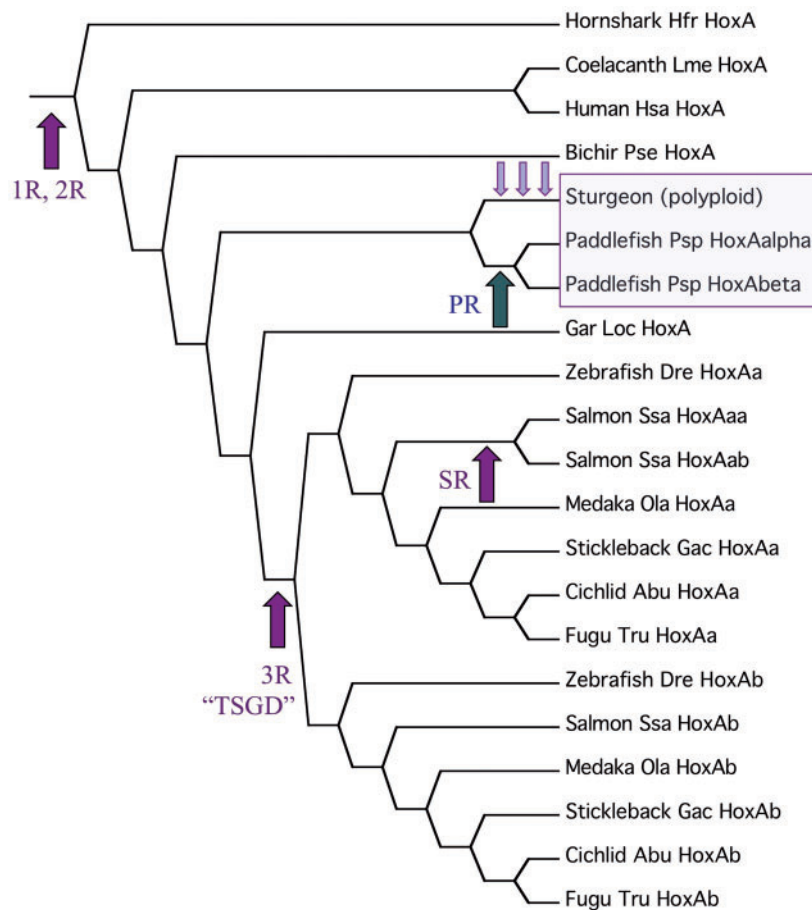
Evidence for these WGDs was, in large part, initially revealed by the discovery of duplicate Hox genes. Hox genes encode transcription factors associated with specification of axial patterning and the development of appendages and organ systems (Ruddle et al. 1994; Burke et al. 1995; Roberts et al. 1995; Warot et al. 1997; Lemons and McGinnis 2006; Mallo et al. 2010). Aspects of Hox gene structure and function are conserved across wide taxonomic distances. However, changes in the protein coding sequences have been linked to the evolution and development of novel characters (Lynch et al. 2008; Crow et al. 2009), and the timing and location of gene expression can cause major phenotypic differences (Gellon and McGinnis 1998). Because they play a key role in determination of body plan morphology, it has been widely assumed that Hox genes play a key role in the evolution of diverse metazoan body plans. Therefore, it is particularly intriguing to understand the role of Hox cluster duplications in the evolution of vertebrate body plans and novelty (Holland et al. 1994; Malaga-Trillo and Meyer 2001; Wagner et al. 2003; Prohaska and Stadler 2004). For example, the posterior (5') Hox genes including paralog groups (PGs) Hox13, Hox12, and Hox11 have been implicated in the evolution of a variety of tetrapod novelties such as the autopod/thumb in humans (Shubin et al. 1997), flippers in cetaceans (Wang et al. 2009), and genital/urogenital organs in various tetrapods (Warot et al. 1997; Lynch et al. 2008; Sifuentes-Romero et al. 2010).

With respect to the TSGD, there are several examples of asymmetric evolution and functional divergence of duplicate

gene paralogs, or “ohnologs” when derived from WGD (Wolfe 2000; Byrne and Wolfe 2005), that are associated with novel features in both non-Hox and Hox genes. For example, divergent paralogs of pigmentation genes specify the unique complexity and diversity of color patterning in teleost fishes, contributing to speciation, and therefore diversity, in this group (Braasch et al. 2006, 2007). Overlapping but divergent expression of *Dlx* paralogs have been implicated in the development of zebrafish pharyngeal dentition, reflecting a redistribution of *Dlx* gene function after the TSGD (Borday-Birraux et al. 2006). Several of the duplicated HoxA cluster genes retained in zebrafish exhibit signatures of positive Darwinian selection (Crow and Wagner 2006) and asymmetric rates of evolution (Crow et al. 2009). Asymmetric rates of evolution and/or positive selection on one or both paralogs often indicate functional divergence and can be important in the development of novel features. For example, hypermutability and functional divergence of the *HoxA13a* paralog in zebrafish and other cypriniform taxa is associated with the evolution and development of a novel feature called the yolk sac extension (Crow et al. 2009), providing a clear link between gene duplication and evolutionary novelty.

The American paddlefish, *Polyodon spathula*, has commanded intense interest in the study of vertebrate evolution. Paddlefish and other members of the Acipenseriformes (the sturgeons) were originally thought to be related to sharks and rays because of their heterocercal tail and cartilaginous skeleton. However, the cartilaginous skeleton is pedomorphic and begins to show ossification in later life history stages (Bemis et al. 1997). It is now well established that paddlefish and sturgeons are bony fishes that occupy an interesting phylogenetic position. They represent one of the basal lineages of ray-finned fishes (fig. 1), and it is currently debated whether they are part of the sister clade of the teleosts (Inoue et al. 2002) or represent a basal lineage to the sister clade of teleosts (Kikugawa et al. 2004). Either way, they have been invoked as a key outgroup taxon for studies investigating the evolution of teleosts because of their phylogenetic position (Metscher and Ahlberg 1999).

The order Acipenseriformes is dynamic and plastic with respect to genome duplication. Although the basal ray-finned fish lineages are generally species poor in terms of extant taxa, the Acipenseriformes is the most diverse group, with 27 extant species (Bemis et al. 1997). The group also has an apparent propensity for genome duplication and polyploidization. Although paddlefish are known to have experienced two ancient genome duplications and are now considered diploidized (Fontana 1994), their close relatives, the sturgeons, have experienced three subsequent genome duplications in various lineages based on chromosome number and inferred ploidy level (Bemis et al. 1997; Ludwig et al. 2001 and fig. 1). As a result, paddlefish have been used as an outgroup taxon with



**Fig. 1.**—Illustration of a HoxA gene genealogy based on a summary of hypotheses from previous studies (e.g., *HoxA11*, Crow et al. 2006) reflecting five independent genome duplication events in the evolutionary history of vertebrates. 1R and 2R refer to two rounds of genome duplication that occurred before the origin of jawed vertebrates; 3R refers to a genome duplication that occurred in the stem lineage of teleosts; SR refers to a subsequent genome duplication that occurred in the salmon lineage (note nested nomenclature, e.g., HoxAaa); and PR refers to an independent genome duplication that occurred in the paddlefish lineage (note the nomenclature HoxA $\alpha$  to reflect nonfirst-order paralogy with teleost HoxAa genes). Light purple arrows indicate multiple subsequent WGDs in various sturgeon lineages.

respect to the multiple independent genome duplications within the sturgeons, and as a basal member of the ray-finned fishes with respect to the TSGD (e.g., Metscher et al. 2005; Wagner et al. 2005; Krieger et al. 2008). Evidence for paddlefish as ancient polyploids is based on the number of chromosomes (Dingerkus and Howell 1976; Ludwig et al. 2001; Leggatt and Iwama 2003). This is supported by paralogous copies of two isozyme loci (Carlson et al. 1982), the *POMC* gene (Danielson et al. 1999), and several microsatellite markers (Heist et al. 2002) that map to a duplication in paddlefish that is independent from the TSGD. However, many authors consider that the paddlefish is now diploidized based on karyotypes and nucleolar organizing regions (Fontana 1994; Peng et al. 2007). As a result, previous studies of Hox expression in paddlefish have been curiously confounded because they have not taken the duplication history of this taxon into account.

To understand the comparative and evolutionary significance of the duplicated Hox genes in paddlefish, we have investigated the following questions: What is the complement of Hox genes present on the paddlefish HoxA and HoxD clusters (i.e., have any been lost to mutation?); Does the paddlefish Hox cluster duplication event correspond to a WGD?; Is this duplication independent from the TSGD, and did it occur before paddlefish diverged from sturgeon or after?; When in evolution did the paddlefish duplication occur?; What is the sequence divergence between the HoxA and HoxD paralogs in paddlefish?; and Do any paddlefish paralogs exhibit rate asymmetry or evidence for selection? Finally, we address whether there are similarities or differences in patterns of substitution between the HoxA/D paralogs duplicated in the paddlefish lineage compared with the same genes duplicated approximately 200 Myr earlier in the teleost lineage (i.e., are

there differences in evolutionary processes that occur early after duplication vs. late)?

## Materials and Methods

### Discovery of Hox Duplicates and Characterization of Hox Bacterial Artificial Chromosome Clones

Duplicate paralogs of three HoxA (*HoxA13*, *HoxA11*, and *HoxA1*) and one HoxD (*HoxD4*) genes in paddlefish were discovered by sequencing multiple clones of polymerase chain reaction (PCR) fragments generated with degenerate Hox primers using paddlefish genomic DNA as a template. These sequences exhibited differences that could not be explained by sequencing errors and possessed features indicative of functional paralogs. For example, partial sequences from exon 2 of the paddlefish *HoxA1* genes spanning amino acid 45 through the stop codon and an additional 87 bp in the 3'-UTR (309 bp) revealed two discrete sequences that were differentiated by one triplet indel and 24 bp substitutions (9 nonsynonymous [NS] and 10 synonymous [S]). These differences do not introduce any frame shifts, and the stop codon is intact. In the 3'-UTR region, there were 14 bp differences and a 5 bp indel. Duplicate sequences from *HoxA11* exon 1 (429 bp, spanning amino acids 4–180) exhibited 16 substitutions (5 NS and 11 S). Partial sequences from the *HoxD11* genes spanning 114 bp in exon 2 (from amino acids 223–267) exhibited 13 substitutions (3 NS). These preliminary data provided the rationale to embark on a large-scale sequencing project, and the sequences necessary to construct probes for *HoxA1*, *HoxA13*, and *HoxD11* that subsequently were used to screen an arrayed 10X coverage bacterial artificial chromosome (BAC) genomic library from a single paddlefish specimen that was constructed at the Benaroya Research Institute (Seattle, WA). Specific BAC clones that were positive for HoxA or HoxD probes were subsequently analyzed with *NotI* and *EcoRI* and compared by agarose gel electrophoresis to confirm differences between paralog clones. Scientific names and abbreviation codes for all taxa referred to in this article are given in table 1.

### DNA Sequencing of BAC Clones

DNAs from selected BACs were purified using the Maxiprep kit (Qiagen, Valencia, CA). Shotgun sequencing of HoxA BACs was done using conventional Sanger ABI sequencing. BAC DNA was randomly sheared to 3-kb fragments using a HydroShear (Digilab Genomic Solutions Inc., Holliston, MA), end-repaired, gel-purified, and cloned into the pUC19 vector (Fermentas International Inc., Glen Burnie, MD). Sequencing reactions were performed using standard M13 primers and the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Carlsbad, CA) and sequenced on a 3730x1 DNA Analyzer (Applied Biosystems, Carlsbad, CA) to roughly 10X coverage. Base calling, quality assessment, and assembly were

carried out using the Phred and Phrap (Ewing et al. 1998) and Consed (Gordon et al. 1998). This resulted in full HoxA cluster sequences spanning *HoxA13* to *HoxA1*. Shotgun sequencing of HoxD-containing BACs was done through outsourcing (Macrogen, Korea) by 454 Titanium (Life Technologies, Grand Island, NY) chemistry to ~100X coverage. Sequences were assembled using Newbler (454 Life Technologies) and Phred (Ewing et al. 1998). Because of lack of complete contiguity, assembled fragments were arranged manually using the MAKER gene annotation tool (Cantarel et al. 2008) and multiple sequence alignment relative to published horn shark and coelacanth HoxD genomic sequences. Both paddlefish HoxD clusters spanned *Evx2* to *HoxD8*.

### Gene Annotation and Synteny Analyses of Paddlefish HoxA and HoxD Clusters

We compared the HoxA and HoxD cluster sequences with the single orthologous Hox cluster sequences of the horn shark, coelacanth, and bichir (HoxA only) or gar (HoxD only). Homology of individual genes was established by reciprocal blast, yielding unambiguous assignment of exons. Gene order and summary statistics including sequence divergence, intron length, intergenic length, and overall cluster size were then compared with known Hox clusters from various chordates including horn shark, coelacanth, and bichir to evaluate evolutionary trends. Gene order and synteny were aligned using Multi-LAGAN (Brudno et al. 2003) and visualized with mVISTA (Mayor et al. 2000; Frazer et al. 2004) using the horn shark or coelacanth as the reference sequence.

### Duplicate Paralogs of the *Fzd8* Gene in Paddlefish

The paddlefish *Fzd8* paralogs were amplified and sequenced using degenerate primers based on publicly available sequences of sturgeon, zebrafish, and stickleback. Total genomic DNA was extracted from fin clips or muscle tissue using the Qiagen DNEasy blood and tissue kit (Qiagen Inc., Valencia, CA) following the manufacturer's protocols. PCR was carried out under the following conditions: 35 cycles of 95°C for 30 s, 56°C for 30 s, and 72°C for 1 min. PCR products were purified and subcloned using the pGEM vector system (Promega, Madison, WI) and sequenced using conventional Sanger sequencing.

### Gene Trees and Phylogenetic Analyses

To evaluate the evolutionary history of duplication events, infer first-order paralogy, and determine when in evolution the paddlefish genome duplication occurred, full sequences for five HoxA genes (*HoxA13*, *HoxA11*, *HoxA10*, *HoxA9*, and *HoxA2*), two Hox D genes (*HoxD9* and *HoxD4*), and partial sequences for the *Fzd8* gene were downloaded for representatives from each major clade of jawed vertebrates, including a shark (basal jawed vertebrate), a coelacanth (basal lobe-finned fish), a bichir (basal ray-finned fish), a gar (nonteleost,

**Table 1**

Taxa Referred to in This Study, along with Taxonomic Codes, Common Names, and Source of Sequences

Code	Scientific Name	Referenced as	Common Name	Lineage	Source HoxA a/b	Source HoxD a/b
Cmi	<i>Callorhynchus milli</i>	Ghostshark		Cartilagenous fishes	FJ824598.1	FJ824601.1
Hfr	<i>Heterodontus francisci</i>	Hornshark	Hornshark	Cartilagenous fishes	AF224262.1	AF224263
Lme	<i>Latimeria menadoensis</i>	Coelacanth	Indonesian coelacanth	Lobe-finned fishes	FJ497005	FJ497008.1
Pse	<i>Polypterus senegalus</i>	Bichir	Bichir	Ray-finned fishes	AC126321	
Psp	<i>Polyodon spathula</i>	Paddlefish	Paddlefish	Ray-finned fishes	This study	This study
Sal	<i>Scaphirhynchus albus</i>	Sturgeon	Sturgeon	Ray-finned fishes	DQ119849.1	
Loc	<i>Lepisosteus oculatus</i>	Gar	Spotted gar	Ray-finned fishes	Amores et al. (1998)	Amores et al. (1998)
Aca	<i>Amia calva</i>	Bowfin	Bowfin	Ray-finned fishes	Amores et al. (1998)	Amores et al. (1998)
Dre	<i>Danio rerio</i>	Zebrafish	Zebrafish	Ray-finned fishes, teleosts	NCBI	
Ssa	<i>Salmo salar</i>	Salmon	Atlantic salmon	Ray-finned fishes, teleosts	NCBI	NCBI
Ola	<i>Oryzias latipes</i>	Medaka	Medaka	Ray-finned fishes, teleosts	AB232918.1/AB232919.1	AB232923.1/AB232924.1
Gac	<i>Gasterosteus aculeatus</i>	Stickleback	Threespined stickleback	Ray-finned fishes, teleosts	UCSC genome browser	UCSC genome browser
Abu	<i>Astatotilapia burtoni</i>	Cichlid	Tilapia cichlid	Ray-finned fishes, teleosts	EF594313.1/EF594311.1	EF594315.1/EF594316.1
Tru	<i>Takifugu rubripes</i>	Fugu	Pufferfish	Ray-finned fishes, teleosts	DQ481663.1/DQ481664.1	DQ481668.1/DQ481669.1

member of the sister clade to teleosts), and both paralogs from several teleosts for which duplicate gene data were available (zebrafish, stickleback, tilapia, medaka, and fugu). These sequences were chosen and aligned to both paddlefish paralogs to confirm that the paddlefish duplication was independent from the TSGD and to compare sequence divergences between paddlefish and teleost duplicates. Sequences were aligned using Sequencher 4.1.2 (GeneCodes Corp., Ann Arbor, MI) and SeAl v. 2.0a11 (Rambaut 2002). Gene trees were constructed with PAUP (Swofford 2002), using parsimony, distance (UPGMA), and likelihood algorithms. Bootstrap support for nodes was based on 2,000 replicates. Bayesian analyses were performed in MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001) with model selection determined by the Akaike Information Criteria (AIC), as implemented in MrModeltest 2.3 (Nylander 2008). The Bayesian search ran for 100,000 generations, and log-likelihood scores were plotted to determine when stationarity was achieved. All trees preceding stationarity were discarded, and multiple runs were executed from random trees to ensure that the optimum tree space had been explored, resulting in identical topologies.

### Estimating the Age of the Paddlefish Genome Duplication Event

To estimate the age of the paddlefish duplication event, we used full coding sequences for five HoxA genes, two HoxD genes, and the *Fzd8* locus for which homologous sequences were publicly available for vertebrate taxa spanning our calibration nodes. Branch lengths for all loci were estimated with maximum likelihood (ML) implemented in PAUP using the model selected by hierarchical likelihood ratio test or the AIC, as implemented in MrModeltest 2.3 (Nylander 2008) for individual loci (table 2). Divergence times between paddlefish paralogs were estimated using the software r8s

(Sanderson 2002, 2003), which does not assume a molecular clock and requires a calibration based on the fossil record for at least one node. We used the penalized likelihood method that combines likelihood and a nonparametric rate smoothing penalty function (Sanderson 2002). This permits specification of the relative contribution of the rate smoothing and the data-fitting parts of the estimation procedure. A cross validation procedure was performed to provide a data-driven method for finding the optimal level of smoothing for each locus individually using a single fixed node (Sarcopterygian/Actinopterygian=450 Ma) before running the penalized likelihood algorithm according to Sanderson (2003). We checked the stability of the solution using the “checkgradient” command in r8s. Because of the uncertainty in the placement of the root node, the branch length leading to the outgroup, horn shark in this case, is incorrect. Therefore, the horn shark was omitted, shifting the root to the next node for which branch lengths are estimated accurately in the r8s program. In this case, the root node then becomes the divergence between the lobe-finned fishes (Sarcopterygii) and the ray-finned fishes (Actinopterygii), which was also our fixed calibration point. We used a fixed age of 450 Myr for the divergence time between sarcopterygians and actinopterygians based on both fossils and molecular data (Gardiner 1993; Hedges and Kumar 2003). We used an additional constraint of 210–330 Myr as the minimum and maximum for the origin of teleosts based on fossils 216–203 Ma (Arratia 2004) and molecular data 285–334 Ma (Vandepoele et al. 2004), respectively. Because we did not have access to genomic sequences from a basal teleost, we used the TSGD as the origin of teleosts, which has previously been estimated to have occurred within 3–5 Myr before the origin of teleosts (Crow et al. 2006). The estimated age of teleosts is further supported by mitochondrial genomic data (Inoue et al. 2005) as 284.7–333.8 Ma. Finally, we used a constraint of 141 Myr for the origin of neopterygians

**Table 2**

Age Estimates of WGDs Inferred from Full Coding Region Sequences Using the Program r8s

Locus	Model	bp	Age Estimate (r8s)		Age Estimate (r8s)		
			Psp Dup	CV: Smoothing Parameter	SalmonDup	TeleostDup	Actinops
<i>HoxA13</i>	<i>SYM + I + G</i>	954	<b>36.99</b>	250	26.95	202.46	325.87
<i>HoxA11</i>	<i>TrN + I + G</i>	936	<b>24.96</b>	200	33.8	232.82	305.4
<i>HoxA10</i>	<i>TrN + I + G</i>	1,167	<b>57.81</b>	400	28.8	190.15	281.42
<i>HoxA2</i>	<i>HKY + I + G</i>	1,194	<b>43.75</b>	250	44.88	210.3	354.49
<i>HoxD11</i>	<i>HKY + G</i>	654	<b>57.36</b>	200	110.17		
<i>HoxD10</i>	<i>GTR + I + G</i>	1,065	<b>28.4</b>	500	65.64		
<i>HoxD9</i>	<i>TrN + I + G</i>	864	<b>73.3</b>	500	59.27		
<i>Fzd8</i>	<i>TrN + I + G</i>	1,326	<b>11.02</b>	32		280.08	
<b>All loci Xbar</b>			<b>41.69875</b>		52.787	223.162	316.795
Range: 11.02–73.3 Myr							
HoxA_Concat		4,251	42.7	250	23.63	220	301.29
HoxA Xbar			40.8775		33.61	208.93	75.3225
HoxD Xbar			53.02		60.76		
<i>Fzd8</i>			11.02			280.08	

NOTE.—Model and smoothing parameter were estimated for each locus independently. Concatenated data set for three HoxA genes included *HoxA13*, *HoxA11*, and *HoxA2*. Bold indicates estimates of the paddlefish WGD (Psp Dup).

(gars, bowfins, and teleosts) based on the oldest lepisosteid fossil (from the Cretaceous, Gardiner 1993) for loci for which sequences from the spotted gar (*Lepisosteus oculatus*) or the Florida gar (*Lepisosteus platyrhynchus*) were available. The divergence time between the paddlefish paralogs was estimated for each locus individually and on a concatenated data set for three HoxA genes (*HoxA13*, *HoxA11*, and *HoxA2*; table 2).

#### Estimation Rate Asymmetry and Evidence for Selection between Paddlefish Paralogs

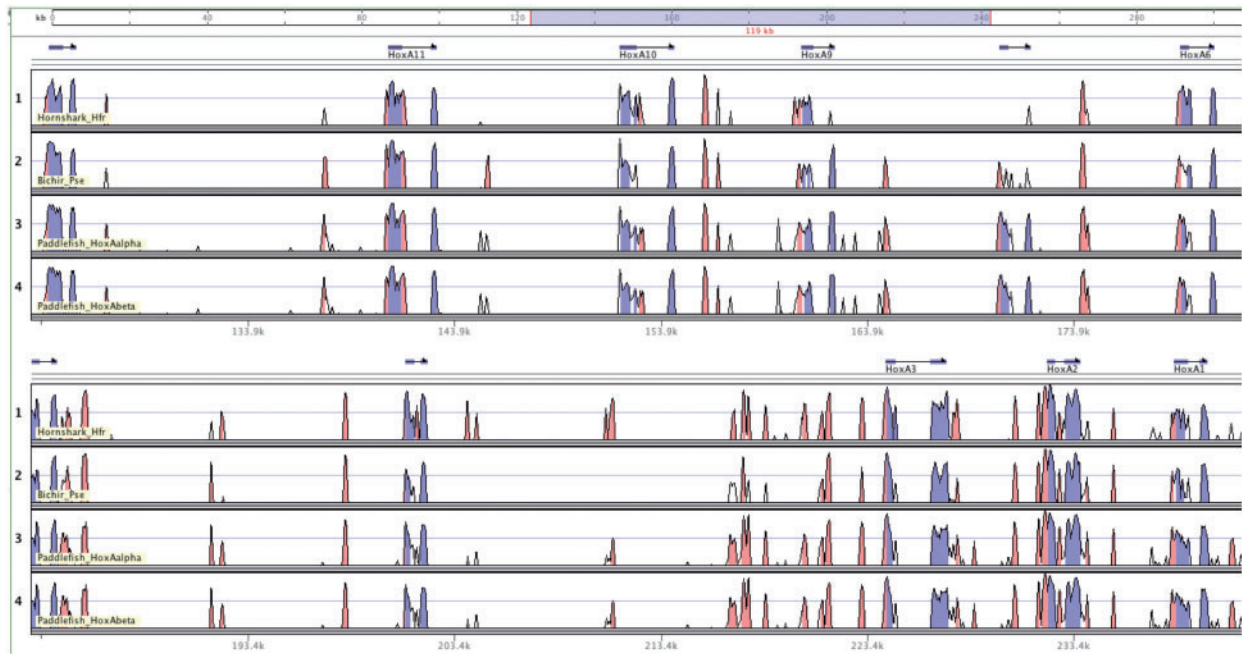
To test for asymmetric rates of evolution between the paddlefish paralogs, we performed pairwise relative rate comparisons using the software package HyPhy (Pond et al. 2005) using the codon model of Goldman and Yang (1994) and the bichir (*Polypterus senegalus*) as the outgroup for the HoxA genes, and the coelacanth (*Latimeria menadoensis*) or horn shark (*Heterodontus francisci*) for the HoxD genes. Variation in lineage-specific dN/dS rate ratios (selection) was estimated using HyPhy (Pond et al. 2005; Kosakovsky Pond et al. 2011).

## Results

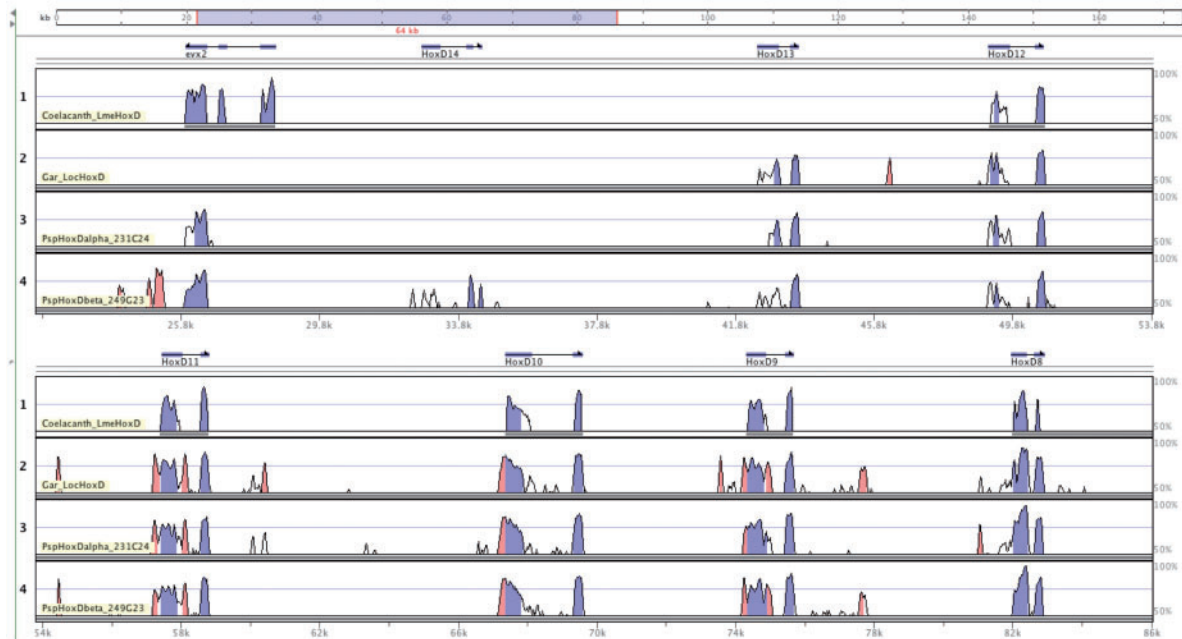
### Inventory of Hox Genes Present on the Paddlefish HoxA and HoxD Clusters

We present evidence for two HoxA and HoxD clusters in paddlefish. We obtained full HoxA cluster sequences spanning *HoxA13* to *HoxA1*, encompassing 115 kb (fig. 2) and partial

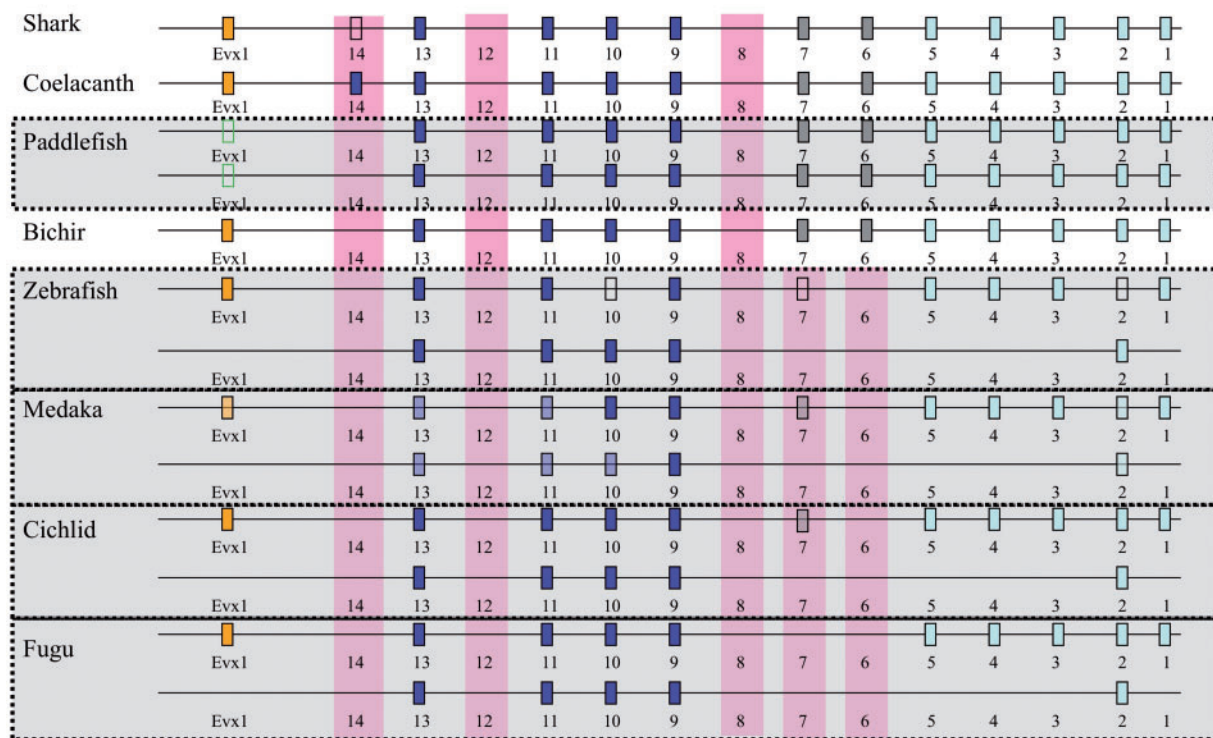
sequences of the HoxD clusters spanning from *Evx2* to *HoxD8* (fig. 3). Two HoxD partial BAC clones were assembled from nine 454 contigs each that were assembled manually resulting in contigs spanning 21,414 bp (BAC\_231C24) and 32,875 bp (BAC\_249G23). The full HoxA and partial HoxD clusters from paddlefish were annotated and compared with full cluster sequences from horn shark, coelacanth, and bichir or gar (figs. 2 and 3). The shark, coelacanth, and bichir HoxA clusters have 11 genes, and all 11 genes were present on each of the paddlefish HoxA gene clusters with complete conservation of synteny (figs. 2 and 4). No HoxA genes have been lost (or gained) in the paddlefish HoxA clusters relative to these ancestral reference genomes. The HoxD clusters of the horn shark, human, and coelacanth are more variable with respect to gene loss and have 12, 9, and 8 HoxD genes, respectively (figs. 3 and 5). Of these, the horn shark is the most ancestral species with a HoxD gene cluster that has lost fewer genes than the human or coelacanth, with the following complement of genes (from 5' to 3'): *Evx2*, *HoxD14*, *HoxD13*, *HoxD12*, *HoxD11*, *HoxD10*, *HoxD9*, and *HoxD8*, *HoxD5*, *HoxD4*, *HoxD3*, *HoxD2*, *HoxD1*. Our paddlefish BAC clones contained the HoxD cluster portion spanning from *Evx2* to *HoxD8* (figs. 3 and 5 and as underlined above). Both paddlefish HoxD BAC clones contained the 3' partial sequences of the *Evx2* gene, but only one contained a *HoxD14* homolog. The remaining six HoxD genes including *HoxD13*, *HoxD12*, *HoxD11*, *HoxD10*, *HoxD9*, and *HoxD8* are present on both paddlefish HoxD clusters with conserved synteny (fig. 5).



**FIG. 2.**—Sequence identity plots of the HoxA cluster genes for multiple taxa using mVISTA (Mayor et al. 2000; Frazer et al. 2004) and Multi-LAGAN (Brudno et al. 2003) with coelacanth as the reference sequence to visualize and compare the HoxA gene complements from horn shark, bichir, and paddlefish paralog clusters. Exons are shown in blue and are based on homology with annotated coelacanth genes (National Center for Biotechnology Information [NCBI] FJ497005).



**FIG. 3.**—Sequence identity plots of the HoxD cluster genes for multiple taxa using mVISTA (Mayor et al. 2000; Frazer et al. 2004) and Multi-LAGAN (Brudno et al. 2003) with horn shark as the reference sequence to visualize and compare the HoxD gene complement in coelacanth, gar, and paddlefish paralog clusters. Exons are shown in blue and are based on homology with annotated horn shark genes (National Center for Biotechnology Information [NCBI] AF224263).



**FIG. 4.**—Gene complements of the HoxA clusters in horn shark, coelacanth, paralogous paddlefish clusters, bichir, and paralogous clusters in four teleosts (zebrafish, medaka, cichlid, and fugu). Boxes with dotted lines highlight duplicated Hox clusters. Numbers indicate paralog group. Posterior Hox genes that share an ancestral sequence with *abd-A* and *abd-B* in fruit fly are shown in purple. Central Hox genes are shown in gray. Anterior Hox genes that share an ancestral sequence with the *lab*, *pb*, *Dfd*, and *Scr* in fruit fly are shown in light blue. The *Evx* gene that lies upstream of Hox14 on the opposite strand is shown in orange. Clear boxes indicate no data and/or presence of pseudogene. Pink bars indicate paralog groups that have been lost in one or more taxa. Note that all ancestral genes are present on both paddlefish HoxA clusters with no gene loss.

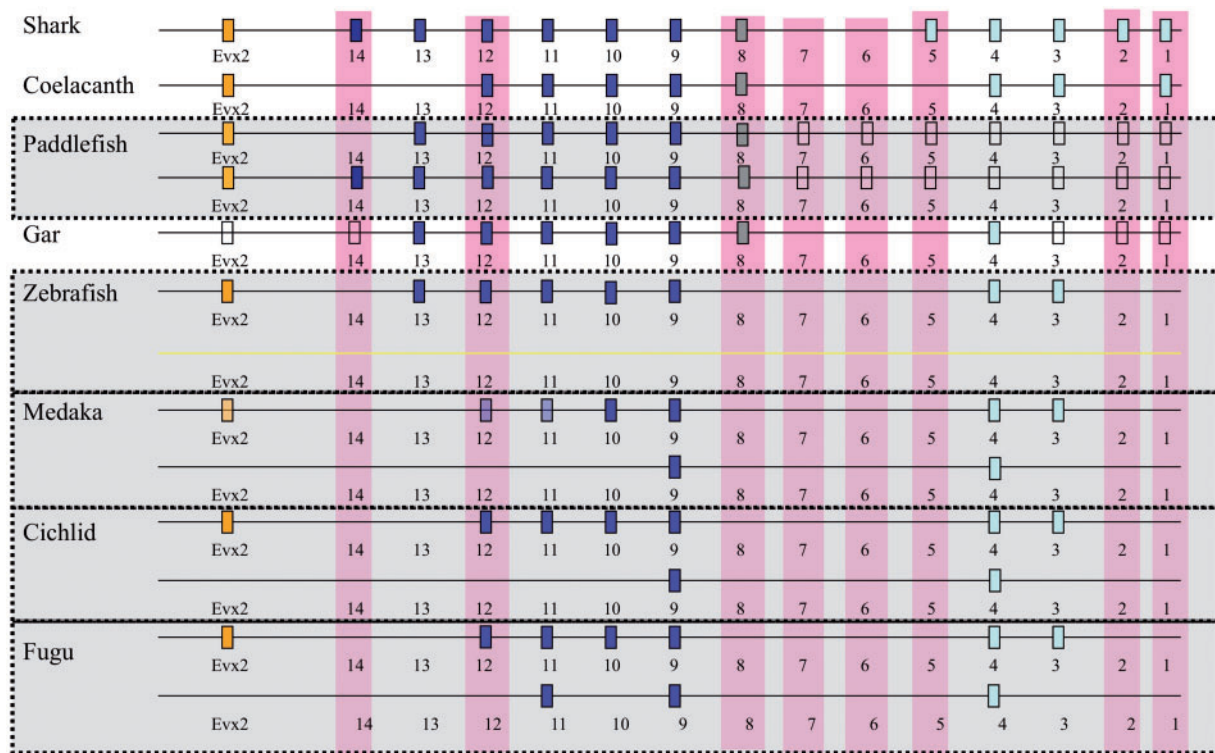
The Hox14 PG genes were first described by Powers and Amemiya (2004) for the coelacanth (*HoxA14*) and horn shark (*HoxD14*), and have since been described for two species of lamprey (*Hox14a*), several cartilaginous fishes (*HoxD14* and pseudogenes of other group 14 paralogs), and two species of lungfish (*HoxA14*) (Feiner et al. 2011; Liang et al. 2011). In all cases, the respective gene is encoded by three exons and exhibits the diagnostic homeodomain third alpha helix motif WFQNQR (as opposed to the usual WFQNR) (Powers and Amemiya 2004). Previously, no Hox14 gene had been identified in any ray-finned fish lineage (Amemiya et al. 2010). Here, we identify an intact HoxD14 gene from paddlefish (Polyodon BAC\_249G23, figs. 3 and 5) that exhibits all the hallmarks of a PG 14 gene. A paralogous HoxD14 gene was not identified from BAC\_231C24, despite exhaustive Blast searches on both the assembly and the raw 454 sequence reads. Similarly, no HoxA14 gene was seen for the two HoxA clusters. It is highly unlikely that the observed paddlefish HoxD14 could be a duplicate of another Hox gene, given the unique structure and sequence of the PG14 genes and because it is confidently placed within the PG14 clade in phylogenetic analyses (supplementary figs. S1 and S2, Supplementary Material online).

Paddlefish paralog clusters were arbitrarily assigned the terminology of “alpha” and “beta” to avoid inference of first-order paralogy with teleost Hox gene paralogs. The paddlefish HoxA $\alpha$  paralogs were isolated from BAC\_352P4 and the HoxA $\beta$  paralogs were isolated from BAC\_370N10. Similarly, the HoxD $\alpha$  cluster genes were isolated from BAC\_231C24 and the HoxD $\beta$  paralogs were isolated from BAC\_249G23.

#### Evidence for a WGD in Paddlefish

Our initial inference of a WGD in paddlefish was based on paralogous HoxA sequences, and we subsequently added paralogous sequences of HoxD cluster genes and the *Fzd8* gene. The HoxA and HoxD clusters are located on different chromosomes in other vertebrates for which the genome has been sequenced, including several ray-finned fishes (Ruddle et al. 1994; Jaillon et al. 2004). Therefore, the WGD in paddlefish is indicated by large, duplicate gene clusters that are likely located at different chromosomal loci. In addition, we found duplicate paralogs of the *Fzd8* gene based on 1,575 bp, extending the single known paddlefish (GB DQ307742.1) sequence by 150 bp and adding a second paralog sequence.





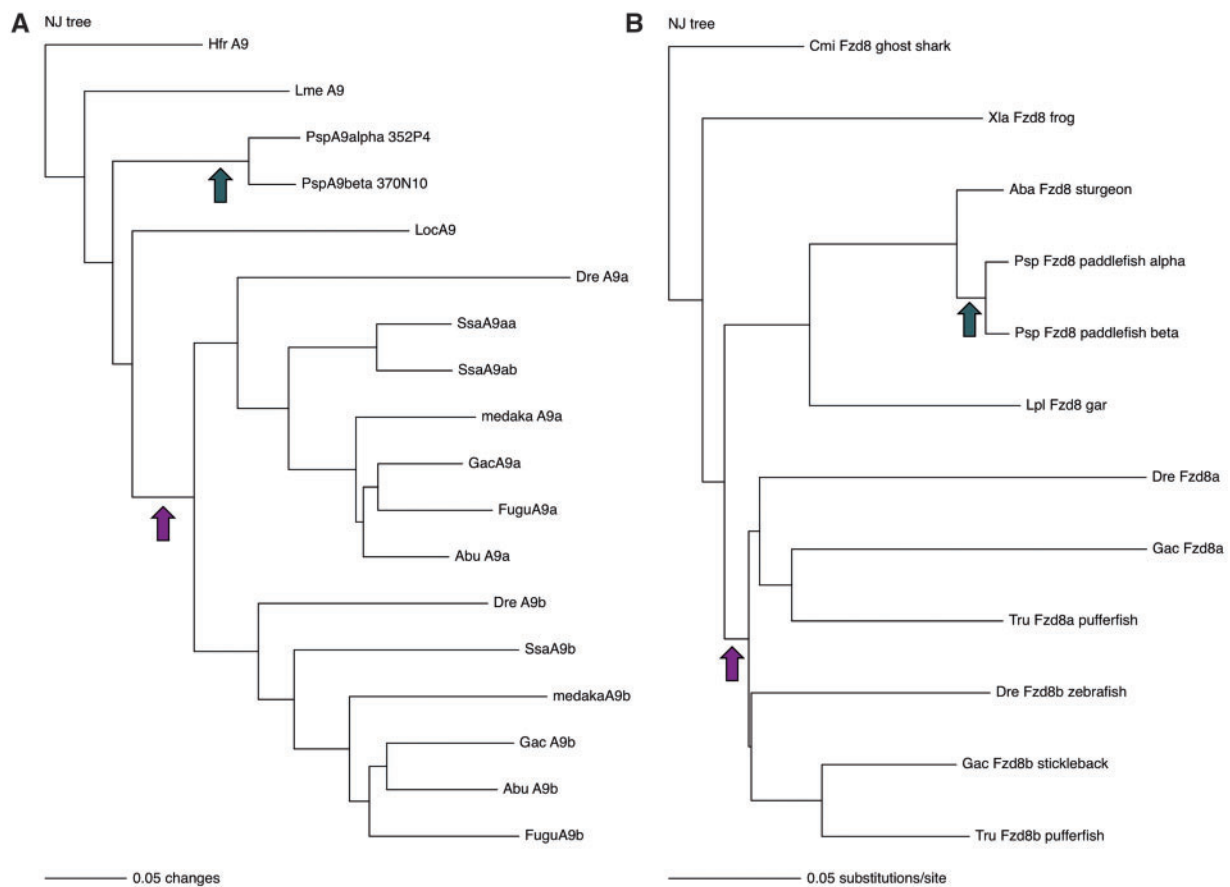
**Fig. 5.**—Gene complements of the HoxD clusters in horn shark, coelacanth, partial sequences for the paralogous paddlefish clusters, gar, and paralogous clusters in four teleosts (zebrafish, medaka, cichlid, and fugu). Numbers indicate paralog group. Posterior Hox genes that share an ancestral sequence with *abd-A* and *abd-B* in fruit fly are shown in purple. Central Hox genes are shown in gray. Anterior Hox genes that share an ancestral sequence with the *lab*, *pb*, *Dfd*, and *Scr* in fruit fly are shown in light blue. The *Evx* gene that lies upstream of Hox14 on the opposite strand is shown in orange. Clear boxes indicate no data and/or presence of pseudogene. Pink bars indicate paralog groups that have been lost in one or more taxa. Note that both paddlefish paralogs are present for seven of the eight HoxD genes detected relative to horn shark, with loss of only the *HoxD14b* paralog in paddlefish.

The paddlefish WGD is also supported by duplicate paralogs of the *Pomc* gene (Danielson et al. 1999). Taken together, the presence of duplicated HoxA clusters and HoxD clusters, along with duplicate paralogs of the *Fzd8* and *Pomc* genes, provides strong evidence for a WGD event in the paddlefish lineage.

### Is the Paddlefish Duplication Independent from Other Genome Duplications?

Ray-finned fishes have experienced multiple WGDs in various lineages throughout their evolution. The jawed vertebrates experienced two ancestral rounds of WGD, and another occurred in the stem lineage of teleosts, with a subsequent WGD in the salmon lineage (fig. 1). Sturgeon and paddlefish belong to the same order, the Acipenseriformes, and clearly genome stability and ploidy level are plastic in this taxon, with a WGD in paddlefish, and three or more subsequent WGDs in the sturgeon lineage based on cytogenetic and genome size data (Blackledge and Bidwell 1993; Birstein et al. 1997; Ludwig et al. 2001). However, the precise timing and independence of these multiple WGDs in the Acipenseriformes remain to be demonstrated.

To first verify that the paddlefish duplication occurred independently from the TSGD, we generated gene trees for five HoxA genes with complete sequence representation for the horn shark, coelacanth, bichir, both paddlefish paralogs, and duplicate paralogs from six additional teleosts. The paddlefish WGD occurred independently from the TSGD (fig. 6 and supplementary figs. S3–S7, Supplementary Material online), as indicated by high levels of statistical support (bootstrap support values 100% in all analyses including neighbor joining [NJ], MP, and BI) at the node uniting paddlefish paralogs, indicating that they are more closely related to one another than to any other vertebrate or teleost sequence. In addition, the teleost Hox duplicates form reciprocally monophyletic paralog clades with statistical support in all analyses. The same pattern was consistently supported in five *HoxD* gene trees (supplementary figs. S8–S13, Supplementary Material online, *HoxD13*, *HoxD12*, *HoxD11*, *HoxD10*, and *HoxD9*) with the same level of statistical support but reduced taxon sampling due to limited availability of full Hox gene sequences (supplementary figs. S8–S13, Supplementary Material online). Zebrafish have lost one of their HoxD clusters, and the derived



**FIG. 6.**—Gene trees supporting topology illustrated in figure 1 for a representative Hox and non-Hox gene. (A) *HoxA9* gene genealogy. (B) *Fzd8* gene genealogy. Purple arrows indicate TSGD, and green arrows indicate the paddlefish WGD. Support joining paddlefish paralogs is 100% in all analyses including NJ/MP/BI. Taxonomic codes defined in table 1. [Supplementary figures S3–S13, Supplementary Material](#) online, exhibit similar topologies from 10 additional genes indicating that the paddlefish WGD occurred independent from the TSGD.

percomorph teleosts have a reduced *HoxD $\beta$*  cluster with only two genes for which both paralogs are maintained (*HoxD9* and *HoxD4*). Interestingly, the *HoxD9b* sequences in percomorphs are highly divergent, rendering the alignment ambiguous for variable regions in exon 1. Whether we excluded the *HoxD9b* sequences creating an unambiguous alignment for all taxa using only the *HoxD9a* teleost paralogs or excluded exon 1 and produced a gene tree based on exon 2 for all taxa and both paralogs, the paddlefish duplication is nonetheless supported as independent ([supplementary figs. S12 and S13, Supplementary Material](#) online). The paddlefish BAC clones that we sequenced did not include the *HoxD4* genes. The *Fzd8* gene tree also supports the paddlefish WGD as independent from the TSGD with the same high level of statistical support (fig. 6B). These data sets were not combinable into a concatenated data set due to different taxonomic representation or missing data due to gene losses in various lineages (e.g., zebrafish lack *HoxA10a*, *HoxA2a*, and the *HoxDb* cluster). However, all individual gene

trees were generally consistent with the topology represented in figure 1, with clear evidence that the paddlefish WGD occurred independently from the TSGD ([supplementary figs. S3–S13, Supplementary Material](#) online).

Although complete mitochondrial genomic sequences are available for most members of the Acipenseriformes, the nuclear genome is not available for any sturgeon or paddlefish. Further, there are no full Hox gene sequences available for any member of the Acipenseriformes, but here, we present the first full Hox gene sequences for the American paddlefish. As such, there is a limited amount of data available to evaluate the evolutionary history of WGDs within the order. We previously sequenced portions of three HoxA genes and one HoxB gene for the basal pallid sturgeon *Scaphirhynchus albus* (Crow et al. 2006), and here, we compare these data with the newly acquired paddlefish Hox sequences. We also added sequences of the paddlefish *Fzd8* paralogs to evaluate an independent origin from sturgeon using publicly available sequences for several vertebrate taxa. The data from individual gene trees

are equivocal with respect to the timing of the paddlefish WGD, relative to their sturgeon relatives. Gene trees from *Fzd8* ( $n = 1,575$  bp), exon1 of the *HoxA13* ( $n = 574$  bp), and *HoxA11* ( $n = 478$  bp) indicate that the paddlefish paralogs were duplicated after their divergence from the sturgeon with high levels of statistical support (fig. 6B and [supplementary fig. S14, Supplementary Material](#) online). Alternatively, sequences from exon 2 of the *HoxA1* gene ( $n = 246$  bp) and exon 1 of the *HoxB5* ( $n = 572$  bp, Crow et al. 2006) indicate that the paddlefish paralogs originated before the divergence from sturgeon with high bootstrap support ([supplementary fig. S14, Supplementary Material](#) online). When we combine our data from partial sequences from four Hox genes into a concatenated data set, again, the data are equivocal with support for different outcomes in different analyses. The NJ tree indicates paralogs originated before the divergence of paddlefish and sturgeon, the maximum parsimony tree supports duplication after divergence of paddlefish and sturgeon, and the ML analysis infers a polytomy ([supplementary fig. S15, Supplementary Material](#) online). When we consider additional data supporting duplication before the divergence from sturgeon with high bootstrap support such as the duplicated *Pomc* gene (Danielson et al. 1999) and the ancestral number of chromosomes inferred for the basal members of the Acipenseriformes (Ludwig et al. 2001), it may appear more parsimonious to suggest the “duplication before divergence” scenario. However, the observed pattern could equally be explained by independent duplication events that occurred shortly after the divergence of these two lineages, leaving little time to build up phylogenetic signal. A more complete data set that includes complete representation of sturgeon paralog sequences will be necessary to fully address this question.

#### Sequence Divergence between HoxA and HoxD Ohnologs in Paddlefish

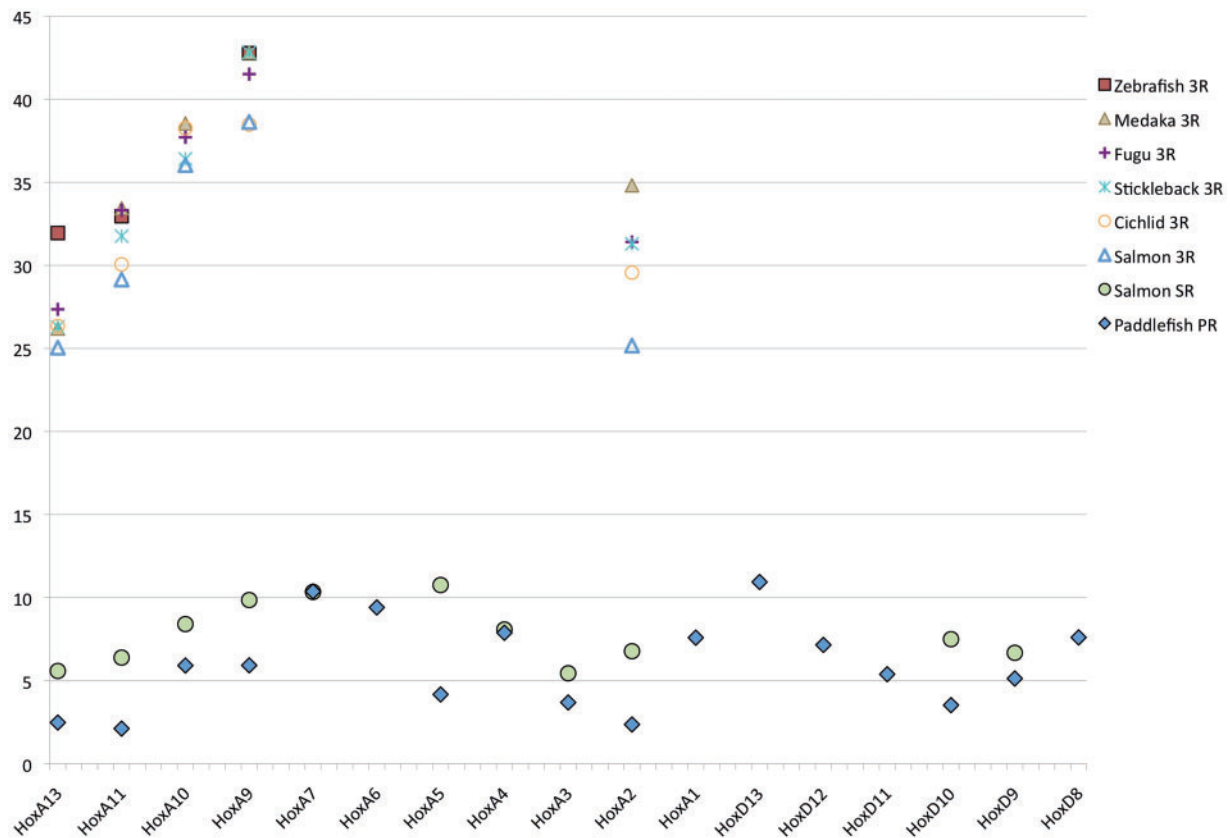
Paddlefish have duplicate paralogs for 11 HoxA genes and 6 HoxD genes, whereas the horn shark, bichir, and gar have only single copies (figs. 2, 4, and 5). Teleosts have experienced multiple gene losses and have only maintained duplicate paralogs in five HoxA genes and two to three HoxD genes. We plotted the nucleotide percent sequence divergence for all HoxA and six HoxD genes in paddlefish and compared these values with divergences between paralogs from other independent WGDs including the TSGD and an independent WGD in the salmon lineage (SR, figs. 1 and 7).

Percent sequence divergence between the full coding regions of the paddlefish HoxA and HoxD paralogs varied among genes, ranging from 2.12% (*HoxA11*) to 10.94% (*HoxD13*, fig. 7). Percent sequence divergences for teleost HoxA paralogs that originated in the TSGD were far greater, indicating that this WGD occurred much earlier in evolution. Teleost paralog divergences range from 25.05% (*HoxA13*) to

42.8% (*HoxA9*) and vary among taxa and gene loci. For example, *HoxA2* exhibited the greatest range among teleosts with 25.18% in salmon to 34.82% in medaka. Within a single taxon, the stickleback (*Gasterosteus aculeatus*) HoxA genes exhibited divergences ranging from 26.29% to 42.8% for *HoxA13* and *HoxA9*, respectively (fig. 7). Duplicate HoxD paralogs have been maintained in teleosts for only two genes: *HoxD9* and *HoxD4*. We found that the teleost *HoxD9a* paralogs were so divergent from *HoxD9b*, that the alignment was ambiguous, and therefore did not include values in this figure; however, the percent sequence divergence for *HoxD9* exon 2 only for four teleosts (medaka, stickleback, cichlid, and fugu) ranged from 21.55 to 27.44 (data not shown). Although it is clear that the TSGD is ancient, and the paddlefish WGD (PR) is relatively young, the independent WGD that occurred in the salmon lineage (SR) is only slightly older than the PR. Salmon are not closely related to paddlefish, and the independence of these distinct WGD events is clearly supported in all gene trees (fig. 6A, [supplementary figs. S3–S13, Supplementary Material](#) online, and illustrated in fig. 1).

#### When in Evolution Did the Paddlefish Duplication Occur?

We estimate that the paddlefish WGD occurred approximately 41.7 Ma (table 2) based on the mean from eight individual loci using the program r8s (Sanderson 2003) with a fixed calibration point of 450 Ma for the ray-finned/lobe-finned fish split (Kumar and Hedges 1998). Estimates for the timing of the paddlefish WGD from individual loci were variable, ranging from 11.02 to 73.3 Ma (table 2). However, the mean from eight loci (41.698 Ma) was in good agreement with the estimate from a concatenated data set of three HoxA genes (42.7 Ma, the only loci with replicate taxon sampling) and the mean of four HoxA loci (40.88 Ma). There were only 10 loci with available sequences spanning our calibration nodes to perform these analyses. We excluded two loci because the gene tree topology did not agree with the generally accepted bony fish phylogeny, and as such, the estimates of divergence times were outliers. For example, the gene tree topology for *HoxA9* did not place the bichir as the basal actinopterygian. Rather the paddlefish paralogs were inferred as the ancestral lineage yielding divergence estimates 3–10 times greater than other loci (156.45 Ma). Further, when these data were trimmed and/or topological constraints were enforced, divergence estimates increased further, indicating instability in this locus as an estimator of divergence time. For the *HoxD12* locus, the gar was inferred as ancestral to paddlefish and as such yielded an estimate of 137.08 Ma for the paddlefish duplication. The remaining eight loci were consistent with the generally accepted vertebrate phylogeny (sensu Inoue et al. 2005 or Kikugawa et al. 2004 and illustrated in fig. 1) at the major nodes supporting monophyly of lobe-finned fishes, ray-finned fishes with bichir as the



**Fig. 7.**—Sequence divergence between Hox paralogs. Each datum is the nucleotide percent sequence divergence between paralogs for genes in which both copies have been retained. Only one paralog (ohnolog) has been lost in teleosts for Hox paralog groups 6–3 and Hox1; therefore the comparison is not possible. Note that the paddlefish duplication is independent from the “teleost-specific genome duplication” and occurred much more recently. Another independent WGD occurred in the salmon lineage, subsequent to the TSGD. Percent sequence divergence was calculated using the uncorrected “*p*” distance matrix of pairwise comparisons for nucleotide alignments of full coding sequences (i.e., exons 1 and 2).

basal taxon, monophyletic teleost, and TSGD paralog clades. We did not make any attempt to infer the sister taxon of teleosts nor the branching order of the derived and polyphyletic percomorphs (sensu Miya et al. 2003, e.g., medaka, stickleback, cichlid, and fugu). For these eight loci, we ran the divergence time analyses multiple times with various parameters, including enforcing a constraint of 210–330 Ma for the origin of teleost paralogs, and when data were available for the gar, we constrained the origin of neopterygians (gars, bowfins, and teleosts) to a minimum of 141 Myr based on the oldest lepisosteid fossil (from the Cretaceous, Gardiner 1993). All these analyses yielded results similar to the unconstrained estimate, using only a single fixed calibration node, and therefore, we report the latter for consistency (i.e., some loci did not have gar sequences, and teleosts did not retain both paralogs for some loci). We note that our estimation of the origin of teleosts (223.16 Myr) is in good agreement with Arratia (2000), and the age estimate of the salmon WGD (SR) was older than the paddlefish duplication, consistent with sequence divergence illustrated in figure 7.

#### Do Paddlefish Paralogs Exhibit Rate Asymmetry or Evidence for Selection?

We looked for evidence of asymmetric rates of evolution between the paddlefish paralogs and found one gene with a significant increase in one paralog relative to the other (table 3). Upon closer examination, we observed a pattern of faster divergence for linked genes along an entire paralog cluster for both the HoxA and HoxD gene clusters in paddlefish. In these analyses, the ML estimate is calculated for a 3-taxon tree with independent rates of evolution and then again with the two paralog branches constrained to be equal. The likelihood ratio test is performed to determine whether the null model (i.e., no difference in evolutionary rates between paralogs) is a better fit. When evaluating NS substitutions only, we found evidence for significant rate asymmetry between paralogs of only a single gene, *HoxA6* ( $P = 0.045646$ ), which is no longer significant when a multiple comparisons correction is applied (table 3). Because of the relatively young age of the paddlefish genome duplication,

**Table 3**

Pairwise Relative Rates for Paddlefish Hox Ohnologs and dN/dS Rate Ratios for Individual Loci

Genes	Outgroup	dN				dS				dN/dS		dN/dS	
		Alpha	Beta	LR	P	Alpha	Beta	LR	P	Alpha	Trend	Beta	Trend
<i>HoxA1</i>	Bichir	0.136958	<b>0.88323</b>	0.1385	0.709798	0.13702	<b>2.41192</b>	0.0338	0.854049	0.9996	Neutral	0.3662	Purifying
<i>HoxA2</i>	Bichir	0.0810301	<b>0.15343</b>	0.4787	0.489	0.08099	<b>0.736921</b>	0.0147	0.903506	1.0005	Neutral	0.2082	Purifying
<i>HoxA3</i>	Bichir	0.136586	<b>0.48306</b>	0.4721	0.492024	0.14557	<b>1.5668</b>	0.6562	0.417916	0.9383	Neutral	0.3083	Purifying
<i>HoxA4</i>	Bichir	0.140877	<b>0.805414</b>	0.2471	0.619112	0.14118	<b>0.256454</b>	2.8984	0.088668	0.9978	Neutral	3.1406	Pos selection
<i>HoxA5</i>	Bichir	0.0821808	<b>0.334836</b>	0.3305	0.565351	0.09112	<b>1.41641</b>	0.0383	0.844889	0.9019	Neutral	0.2364	Purifying
<i>HoxA6</i>	Bichir	0.174912	<b>1.37713</b>	3.9946	0.045646*	0.19717	<b>5.50726</b>	0.7821	0.376492	0.8871	Neutral	0.2501	Purifying
<i>HoxA9</i>	Bichir	0.146106	<b>0.883757</b>	0.515	0.472992	0.14621	<b>0.68322</b>	1.0213	0.312209	0.9993	Neutral	1.2935	Pos selection
<i>HoxA10</i>	Bichir	0.109361	<b>0.814478</b>	0.263	0.608069	0.10126	<b>1.34568</b>	0.0026	0.959316	1.0800	Neutral	0.6053	Purifying
<i>HoxA11</i>	Bichir	0.073913	<b>0.207006</b>	0.9957	0.318349	0.07351	<b>0.785544</b>	0.2572	0.61206	1.0055	Neutral	0.2635	Purifying
<i>HoxA13</i>	Bichir	0.0784	<b>0.302403</b>	0.4357	0.509224	0.08701	<b>0.182551</b>	1.2287	0.267652	0.9010	Neutral	1.6565	Pos selection
<i>HoxD13</i>	Hfr	0.218708	<b>1.76621</b>	0.0002	0.989621	0.21819	<b>4.23132</b>	2.3545	0.12492	1.0024	Neutral	0.4174	Purifying
<i>HoxD12</i>	Lme	0.16994	<b>1.90892</b>	3.0667	0.079909	0.16991	<b>1.3417</b>	0.1503	0.698257	1.0002	Neutral	1.4228	Pos selection
<i>HoxD11</i>	Lme	0.169774	<b>0.479902</b>	0	1	0.17014	<b>2.25434</b>	0.0824	0.774026	0.9978	Neutral	0.2129	Purifying
<i>HoxD10</i>	Lme	0.179818	<b>0.418399</b>	0.1297	0.718693	0.17881	<b>1.71416</b>	0.4886	0.484534	1.0056	Neutral	0.2441	Purifying
<i>HoxD9</i>	Lme	0.148756	<b>0.22226</b>	0.4326	0.510699	0.1417	<b>1.34173</b>	0.8316	0.361818	1.0498	Neutral	0.1657	Purifying
<i>HoxD8</i>	Lme	0.135323	<b>0.59449</b>	0.0468	0.828764	0.13525	<b>3.66521</b>	0.2238	0.636169	1.0006	Neutral	0.1622	Purifying

NOTE.—The relative rates were estimated using ML and the codon model of Goldman and Yang (1994) as implemented in HyPhy. dN, NS substitution rate; dS, S substitution rate; LR, likelihood ratio. Outgroup for each comparison is indicated. Bold indicates the paralog with the longer branch length. \*Indicates a significant difference.

it is not surprising that rate asymmetries have not accumulated to the level of statistical significance. However, there was a surprising pattern when we compared which Hox paralogs exhibit the longer branch length across all duplicate Hox genes. We found that the  $\beta$  paralogs diverge faster for all 16 genes analyzed. The a priori probability that the 10 HoxA $\beta$  genes occurring on the same cluster would exhibit increased substitution rates by chance is highly unlikely ( $P=0.00097$ ). Even the probability of the six HoxD $\beta$  genes exhibiting consistently higher rates is significantly unlikely ( $P=0.01562$ ). In fact, the inferred NS substitution rates on the  $\beta$  paralog clusters were 1.49–11.23 times greater than substitution rates on the  $\alpha$  paralog clusters (dNb/dNa, from table 3). Because the uncorrected sequence divergence between paralogs varied among loci, we also evaluated rate asymmetries based on S substitutions as well. Again, S substitution rates were higher in  $\beta$  paralogs versus  $\alpha$  paralogs in all comparisons for 16 genes. S substitution rates for the  $\beta$  paralogs were greater than  $\alpha$  paralogs by a factor of 1.8–27.93. These values were generally greater for dS  $\alpha/\beta$  comparisons than dN  $\alpha/\beta$ , indicating that S substitutions are the predominant substitution class. Overall, both NS and S substitutions are accruing faster in all genes represented in one paralog cluster relative to the other, even though this rate asymmetry was not significant in individual gene comparisons. This suggests cluster wide, or regional, differences in the pattern and process of molecular evolution between first order Hox paralogs. Finally, dN/dS rate ratios for individual gene loci yielded interesting results. We found that dN/dS rate ratios for only the  $\alpha$  paralogs were close to one, consistent with neutral

evolution. In contrast, dN/dS rate ratios for genes in the  $\beta$  paralog were either less than or greater than 1, indicating either purifying or positive selection. We note that dN/dS values  $> 1$  could also be explained by relaxed purifying selection.

## Discussion

The American paddlefish (*Polyodon spathula*) lineage experienced a WGD (PR, fig. 1) approximately 42 Ma. This event was clearly independent from the salmonid WGD and the TSGD (3R) that occurred approximately 285–334 Ma based on gene genealogies from 10 loci (supplementary figs. S3–S13, Supplementary Material online). However, it is unclear whether this event occurred in the stem lineage of the Acipenseriformes or if independent WGDs occurred in both the paddlefish and sturgeon lineages. Peng et al. (2007) estimate that the split between paddlefish and sturgeon occurred approximately 184.4 Ma, which is much older than our estimate of the paddlefish WGD and that the divergence between the Chinese and American paddlefish occurred approximately 68 Ma. These data support the idea that the paddlefish duplication is exclusive, and occurred after their divergence from sturgeon. It is unclear whether paddlefish should be considered polyploid or rediploidized, and the terminology for both has been invoked in the literature. The genome duplication history in paddlefish has been further confounded based on analyses of chromosome number and c value data because the bichir (*Polypterus senegalus*) has experienced chromosome reduction (Morescalchi et al.

2008) and members of the Acipenseriformes exhibit microchromosomes that may be the result of chromosome splitting (van Eenennaam et al. 1998; Kim et al. 2005). Curiously, these considerations have not been previously accounted for with respect to Hox cluster duplication or gene expression studies, even though paddlefish are often included in such studies because of their importance as a basal ray-finned fish representative. Here, we show that first-order paralogs (ohnologs) exhibit various levels of sequence divergence and clear patterns of substitution processes that indicate diploidization processes are ongoing and quantifiable. However, several authors refer to the paddlefish genome as 4N (Birstein and DeSalle 1998; Ludwig et al. 2001), and we do not disagree with this nomenclature because it highlights the fact that the paddlefish genome is duplicated and that this is an important consideration in studies involving gene expression or molecular evolution.

#### *HoxD14* Is Present in a Ray-Finned Fish

The discovery of a *HoxD14* gene in *Polyodon* was surprising given the apparent absence of PG14 genes in any ray-finned fish to date (Amemiya et al. 2010). Its retention in this lineage as an intact gene suggests that it may be functional, though expression data have not yet been obtained. Whether its expression will resemble the noncanonical PG14 expression patterns found in lamprey and shark (Kuraku et al. 2008; Oulion et al. 2011) is yet to be determined. The absence of an ohnolog of *HoxD14* in the duplicated D $\alpha$  cluster suggests that it has been lost in the relatively short time since the WGD. Notably, although *HoxD14* is present in *Polyodon*, the lobed-finned fishes, coelacanth, and lungfish possess *HoxA14* genes, suggesting divergent resolution of PG14 genes in these lineages. The retention and loss of PG14 genes, however, is probably not a simple matter because cartilaginous fishes have likely undergone mutations in different PG14 genes independently as inferred by the presence of pseudogenes (Powers and Amemiya 2004; Ravi et al. 2009).

#### Comparing Processes of Molecular Evolution in the HoxA/D Cluster Genes Duplicated Recently in the Paddlefish Lineage with the Same Genes Duplicated Much Earlier in the Teleost Lineage

The analysis of the duplicated Hox genes in the paddlefish provides a remarkable opportunity to investigate the proximate changes in molecular evolution that occur relatively shortly after a duplication event and to compare those processes with the same genes that were duplicated independently before the origin of teleosts approximately 250 Myr earlier.

There are recurring patterns of molecular evolution that are apparent in paralogs that originated from independent duplication events. First, there is a pattern of increasing divergence, or relaxed constraint, in the posterior HoxA genes in six teleost taxa, with divergence increasing from *HoxA13* to

*HoxA9* (5' to 3', fig. 7) that is mirrored in the salmon and paddlefish posterior HoxA genes (albeit, not as precisely for the paddlefish paralogs that were duplicated most recently). Therefore, processes structuring molecular evolution appear to be consistent for paralog clusters originating from three different, independent duplication events (3R, SR, and PR). This trend is not correlated with the size of coding regions (i.e., it is unlikely that the trend is a function of increasing number of unconstrained sites, with 888, 849, 1,014, and 777 bp in the coding regions of *HoxA13*, *HoxA11*, *HoxA10*, and *HoxA9* paralogs, respectively). Hox genes exhibit spatial and temporal colinearity with nested and overlapping expression domains suggesting coregulation by upstream segment/trait specification genes or by the same processes controlling expression of specification genes (Tabin and Wolpert 2007). This would explain the observed conservation of Hox cluster integrity in vertebrates. However, this would not explain why the build up of divergence in the protein coding sequences may be increasingly constrained with increasing paralog number with a repeating pattern of increasing sequence divergence in the posterior HoxA genes from multiple independent duplication events that occurred at various times in evolution.

Second, we observed a consistent pattern of faster divergence in linked genes along entire paralog clusters in both the HoxA and HoxD gene clusters in paddlefish suggesting independent processes of molecular evolution between first-order Hox ohnologs. This rate asymmetry between ohnologs was not significant in single gene comparisons, but the probability that a consistent pattern in all 10 HoxA $\beta$  genes and all six HoxD $\beta$  genes occurring on the same cluster would exhibit increased substitution rates by chance is significantly unlikely. Further, the pattern was consistent when NS substitutions were considered and when S substitutions were considered, suggesting regional differences in mutation rates between paralog clusters. In addition, dN/dS rate ratios for paralogs for individual gene loci yielded interesting results. We found that dN/dS rate ratios for the  $\alpha$  paralogs were close to 1, consistent with neutral evolution. In contrast, dN/dS rate ratios for genes on the  $\beta$  paralog cluster were either less than or greater than 1, indicating purifying or positive (relaxed purifying) selection. A similar pattern of consistently higher substitution rates, for both NS and S substitutions, in genes from one paralog cluster relative to the other was found in five HoxA genes in pufferfish *Takifugu rubripes* (Wagner et al. 2005). One possible explanation for this pattern is that one paralogous Hox cluster is transcriptionally inhibited due to chromatin structure alterations such as heterochromatinization. This could potentially explain dN/dS ratios close to 1 along an entire cluster, because limited expression could keep them veiled from selection. This model could also explain the difference in the magnitude of both, dN and dS. We find that the alpha paralogs have consistently lower substitution rates than the beta paralogs, which is consistent with lower

mutation rates. Heterochromatinization could explain lower mutation rates because it protects DNA from lesions resulting in replication errors (Boulikas 1992). Transcriptionally active genes experience higher mutation rates, but they also exhibit higher rates of DNA repair (Boulikas 1992). As a result, patterns of positive and purifying selection are associated with transcriptionally active regions of euchromatin (Babbitt and Kim 2008; Fudenberg et al. 2011). This model would suggest that although paddlefish have a duplicated genome, they may be functionally diploid due to transcriptional inhibition (for one copy of the HoxA and HoxD clusters) by chromatin structure analogous to the X-chromosomal dosage compensation mechanism in mammals. However, to date, there are no data comparing differential expression or chromatin structure between paralogs that would support this hypothesis. In summary, the  $\beta$  paralog HoxA and HoxD gene clusters are more dynamic with respect to gene retention, NS and S substitutions, and may be differentially maintained by natural selection.

### Significance of Genome Duplication in Paddlefish

Clarifying the status of the duplicated paddlefish genome bears on the current paradigm of paired limb evolution in jawed vertebrates—a historical comparative developmental genetics model (reviewed in Mabee 2000). Tetrapods express a third wave of Hox gene expression that is associated with digit formation in the autopod. This was thought to be a synapomorphy shared by tetrapods based on comparisons of Hox gene expression patterns between zebrafish and mice. Because zebrafish are representatives of a derived lineage, it has been suggested repeatedly in the literature that a basal ray-finned fish, such as the paddlefish, must be examined to confirm this hypothesis. Recently, it was shown that the paddlefish also exhibits a late phase of Hox gene expression indicating that this expression pattern is not a synapomorphy specific to tetrapods but may in fact be the ancestral regulatory pathway that was in place before the divergence of ray-finned and lobe-finned fish (Davis et al. 2007). Davis et al. showed that the third phase Hox gene expression pattern in paddlefish and tetrapods involves several *HoxD* cluster genes but not the HoxA cluster genes, *HoxA11* and *HoxA13*, that are normally expressed in phase 2 of limb formation in both zebrafish and mouse. However, it is possible that the expression pattern described in paddlefish is obscured by the Hox gene cluster duplications reported here and could be secondarily derived. In other words, the Hox gene expression patterns shown in previous studies could represent expression of one or both paralogs, whereas expression of the other has gone undetected or undifferentiated. These studies have not taken gene duplication into account. This is an important consideration because it could largely change the current interpretation of the fin–limb transition in vertebrate evolution. If so,

this would implicate a novel and independently derived pathway in fin development that is associated with divergence of duplicate genes.

### Supplementary Material

Supplementary figures S1–13 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

### Acknowledgments

The authors thank John Postlethwait for procuring the *Polyodon* sample that was used for the BAC library, Alicia Hill, Tsutomu Miyake, Andy Stuart, and Deb Tinnemore for help in construction and screening of the BAC library, Yi Peng for help with DNA sequencing, and Kent Susick for help obtaining partial sequences of the paddlefish *Fzd8* paralogs. This work was supported by National Science Foundation grants IOS-1022509 (to K.D.C.), IOS-0321470 (to G.P.W.), and IOS-0321461 and MCB-0719558 (to C.T.A.), and National Institutes of Health grants HL66728 (to E. Rubin and J.-F.C.) and RR14085 (to C.T.A.). HoxA cluster genes for the gar (*Lo*) were obtained from Angel Amores.

### Literature Cited

- Amemiya CT, et al. 2010. Complete HOX cluster characterization of the coelacanth provides further evidence for slow evolution of its genome. *Proc Natl Acad Sci U S A*. 107:3622–3627.
- Amores A, et al. 1998. Zebrafish Hox clusters and vertebrate genome evolution. *Science* 282:1711–1714.
- Arratia G. 2000. New teleostean fishes from the Jurassic of southern Germany and the systematic problems concerning the ‘pholidophoriforms’. *Paläontologische Zeitschrift* 74:113–143.
- Arratia G. 2004. Systematics, paleoenvironments and biodiversity. In: Arratia G, Tintori A, editors. *Mesozoic fishes 3*. München, Germany: Verlag Dr. Friedrich Pfeil. p. 279–315.
- Babbitt GA, Kim Y. 2008. Inferring natural selection on fine-scale chromatin organization in yeast. *Mol Biol Evol*. 25:1714–1727.
- Bemis WE, Findeis EK, Grande L. 1997. An overview of Acipenseriformes. *Environ Biol Fishes*. 48:25–71.
- Birstein VJ, DeSalle R. 1998. Molecular phylogeny of Acipenserinae. *Mol Phylogenet Evol*. 9:141–155.
- Birstein VJ, Hanner R, DeSalle R. 1997. Phylogeny of the Acipenseriformes: cytogenetic and molecular approaches. *Environ Biol Fishes*. 48:127–155.
- Blackledge KH, Bidwell CA. 1993. Three ploidy levels indicated by genome quantification in Acipenseriformes of North America. *J Hered*. 84:427–430.
- Borday-Birraux V, et al. 2006. Expression of *Dlx* genes during the development of the zebrafish pharyngeal dentition: evolutionary implications. *Evol Dev*. 8:130–141.
- Boulikas T. 1992. Evolutionary consequences of nonrandom damage and repair of chromatin domains. *J Mol Evol*. 35:156–180.
- Braasch I, Salzburger W, Meyer A. 2006. Asymmetric evolution in two fish-specifically duplicated receptor tyrosine kinase paralogs involved in teleost coloration. *Mol Biol Evol*. 23:1192–1202.
- Braasch I, Schartl M, Volff J-N. 2007. Evolution of pigment synthesis pathways by gene and genome duplication in fish. *BMC Evol Biol*. 7:74.

- Brudno M, et al. 2003. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* 13: 721–731.
- Burke AC, Nelson CE, Morgan BA, Tabin C. 1995. Hox genes and the evolution of vertebrate axial morphology. *Development* 121:333–346.
- Byrne KP, Wolfe KH. 2005. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res.* 15:1456–1461.
- Cantarel BL, et al. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 18: 188–196.
- Cardoso J, et al. 2007. Persistence of duplicated PAC1 receptors in the teleost, *Sparus auratus*. *BMC Evol Biol.* 7:221.
- Carlson DM, Kettler MK, Fisher SE, Whitt GS. 1982. Low genetic variability in paddlefish populations. *Copeia* 1982:721–725.
- Christoffels A, et al. 2004. Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol Biol Evol.* 21:1146–1151.
- Crow KD, Amemiya CT, Roth J, Wagner GP. 2009. Hypermutability of Hoxa13A and functional divergence from its paralog are associated with the origin of a novel developmental feature in zebrafish and related taxa (Cypriniformes). *Evolution* 63:1574–1592.
- Crow KD, Wagner GP. 2006. What is the role of genome duplication in the evolution of complexity and diversity in vertebrates? *Mol Biol Evol.* 23:887–892.
- Crow KD, et al. 2006. The “fish specific” Hox cluster duplication is coincident with the origin of teleosts. *Mol Biol Evol.* 23:121–136.
- Danielson PB, et al. 1999. Duplication of the POMC gene in the paddlefish (*Polyodon spathula*): analysis of gamma-MSH, ACTH, and beta-endorphin regions of ray-finned fish POMC. *Gen Comp Endocrinol.* 116:164–177.
- Davis MC, Dahn RD, Shubin NH. 2007. An autopodial-like pattern of Hox expression in the fins of a basal actinopterygian fish. *Nature* 447: 473–476.
- de Souza FSJ, Bumashny VF, Low MJ, Rubinstein M. 2005. Subfunctionalization of expression and peptide domains following the ancient duplication of the proopiomelanocortin gene in teleost fishes. *Mol Biol Evol.* 22:2417–2427.
- Dingerkus G, Howell W. 1976. Karyotypic analysis and evidence of tetraploidy in the North American paddlefish, *Polyodon spathula*. *Science* 194:842–844.
- Duda TF, Palumbi SR. 1999. Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc Natl Acad Sci U S A.* 96:6820–6823.
- Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8: 175–185.
- Feiner N, Ericsson R, Meyer A, Kuraku S. 2011. Revisiting the origin of the vertebrate Hox14 by including its relict sarcopterygian members. *J Exp Zool B Mol Dev Evol.* 316:515–525.
- Fontana F. 1994. Chromosomal nucleolar organizer regions in four sturgeon species as markers of karyotype evolution in Acipenseriformes (Pisces). *Genome Res.* 37:888–892.
- Frazer KA, et al. 2004. VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32:W273–W279.
- Fudenberg G, Getz G, Meyerson M, Mirny LA. 2011. High order chromatin architecture shapes the landscape of chromosomal alterations in cancer. *Nat Biotechnol.* 29:1109–1113.
- Gardiner BG. 1993. Osteichthyes: basal actinopterygians. In: Benton M, editor. *The fossil record 2*. London: Chapman & Hall, p. 611–619.
- Gellon G, McGinnis W. 1998. Shaping animal body plans in development and evolution by modulation of Hox expression patterns. *BioEssays* 20:116–125.
- Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol.* 11: 725–736.
- Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* 8:195–202.
- Hawkins MB, et al. 2000. Identification of a third distinct estrogen receptor and reclassification of estrogen receptors in teleosts. *Proc Natl Acad Sci U S A.* 2000:10 751–710 756.
- Hedges SB, Kumar S. 2003. Genomic clocks and evolutionary timescales. *Trends Genet.* 19:200–206.
- Heist EJ, Nicholson EH, Sipiorski JT, Keeney DB. 2002. Microsatellite markers for the paddlefish (*Polyodon spathula*). *Conserv Genet.* 3:205–207.
- Hoegg S, Brinkmann H, Taylor JS, Meyer A. 2004. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J Mol Evol.* 59:190–203.
- Holland PW, Garcia-Fernandez J, Williams NA, Sidow A. 1994. Gene duplications and the origins of vertebrate development. *Development Suppl.* 125–133.
- Holland PWH, Garcia-Fernandez J. 1996. Hoxgenes and chordate evolution. *Dev Biol.* 173:382–395.
- Huelsenbeck JP, Fredrik Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
- Inoue JG, Miya M, Tsukamoto K, Nishida M. 2002. Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the “ancient fish.” *Integr Comp Biol.* 42:1249–1249.
- Inoue JG, Miya M, Venkatesh B, Nishida M. 2005. The mitochondrial genome of Indonesian coelacanth *Latimeria menadoensis* (Sarcopterygii: Coelacanthiformes) and divergence time estimation between the two coelacanths. *Gene* 349:227–235.
- Jaillon O, et al. 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431: 946–957.
- Kikugawa K, et al. 2004. Basal jawed vertebrate phylogeny inferred from multiple nuclear DNA-coded genes. *BMC Biol.* 2:3.
- Kim DS, et al. 2005. Karyotype of North American shortnose sturgeon *Acipenser brevirostrum* with the highest chromosome number in the Acipenseriformes. *Ichthyol Res.* 52:94–97.
- Kosakovsky Pond SL, et al. 2011. A random effects branch-site model for detecting episodic diversifying selection. *Mol Biol Evol.* 28:3033–3043.
- Krieger J, et al. 2008. The molecular phylogeny of the order Acipenseriformes revisited. *J Appl Ichthyol.* 24:36–45.
- Kumar S, Hedges SB. 1998. A molecular timescale for vertebrate evolution. *Nature* 392:917–920.
- Kuraku S, et al. 2008. Noncanonical role of Hox14 revealed by its expression patterns in lamprey and shark. *Proc Natl Acad Sci U S A.* 105: 6679–6683.
- Leggatt RA, Iwama GK. 2003. Occurrence of polyploidy in the fishes. *Rev Fish Biol Fisheries.* 13:237–246.
- Lemons D, McGinnis W. 2006. Genomic evolution of Hox gene clusters. *Science* 313:1918–1922.
- Liang D, et al. 2011. A general scenario of Hox gene inventory variation among major sarcopterygian lineages. *BMC Evol Biol.* 11:25.
- Ludwig A, et al. 2001. Genome duplication events and functional reduction of ploidy levels in sturgeon (Acipenser, Huso and Scaphirhynchus). *Genetics* 158:1203–1215.
- Lynch M, Force A. 2000. The origin of interspecific genomic incompatibility via gene duplication. *Am Nat.* 156:590–605.
- Lynch M, Katju V. 2004. The altered evolutionary trajectories of gene duplicates. *Trends Genet.* 20:544–549.
- Lynch VJ, et al. 2008. Adaptive changes in the transcription factor HoxA-11 are essential for the evolution of pregnancy in mammals. *Proc Natl Acad Sci U S A.* 105:14928–14933.
- Mabee PM. 2000. Developmental data and phylogenetic systematics: evolution of the vertebrate limb. *Am Zool.* 40:789–800.



- Malaga-Trillo E, Meyer A. 2001. Genome duplication and accelerated evolution of Hox genes and cluster architecture in teleost fishes. *Am Zool.* 41:676–686.
- Mallo M, Wellik DM, Deschamps J. 2010. Hox genes and regional patterning of the vertebrate body plan. *Dev Biol.* 344:7–15.
- Mayor C, et al. 2000. VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* 16:1046–1047.
- Metscher BD, Ahlberg PE. 1999. Zebrafish in context: uses of a laboratory model in comparative studies. *Dev Biol.* 210:1–14.
- Metscher BD, et al. 2005. Expression of *Hoxa-11* and *Hoxa-13* in the pectoral fin of a basal ray finned fish, *Polyodon spathula*: implications for the origin of tetrapod limbs. *Evol Dev.* 7:186–195.
- Meyer A, Scharl M. 1999. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr Opin Cell Biol.* 11:699–704.
- Miya M, et al. 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. *Mol Phylogenet Evol.* 26:121–138.
- Morescalchi M, et al. 2008. Karyotypic characterization and genomic organization of the 5S rDNA in *Polypterus senegalus* (Osteichthyes, Polypteridae). *Genetica* 132:179–186.
- Naruse K, et al. 2000. A detailed linkage map of Medaka, *Oryzias latipes*: comparative genomics and genome evolution. *Genetics* 154:1773–1784.
- Nylander JAA. 2008. MrModeltest2.3. Distributed by author. Uppsala, Sweden: Department of Systematic Zoology, Uppsala University.
- Ohno S. 1970. Evolution by gene duplication. New York: Springer-Verlag.
- Oulion S, et al. 2011. Evolution of repeated structures along the body axis of jawed vertebrates, insights from the *Scyliorhinus canicula* Hox code. *Evol Dev.* 13:247–259.
- Peng Z, et al. 2007. Age and biogeography of major clades in sturgeons and paddlefishes (Pisces: Acipenseriformes). *Mol Phylogenet Evol.* 42:854–862.
- Pond SL, Frost SD, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21:676–679.
- Postlethwait J, et al. 2004. Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet.* 20:481–490.
- Powers TP, Amemiya CT. 2004. Evidence for a Hox14 paralog group in vertebrates. *Curr Biol.* 14:R183–R184.
- Prohaska SJ, Stadler PF. 2004. The duplication of the Hox gene clusters in teleost fishes. *Theory Biosci.* 123:89–110.
- Rambaut A. 2002. SE-AL v. 2.0a11: sequence alignment program, Available from: <http://tree.bio.ed.ac.uk/software/seal/>.
- Ravi V, et al. 2009. Elephant shark (*Callorhynchus milii*) provides insights into the evolution of Hox gene clusters in gnathostomes. *Proc Natl Acad Sci U S A.* 106:16327–16332.
- Roberts DJ, et al. 1995. Sonic hedgehog is an endodermal signal inducing Bmp-4 and Hox genes during induction and regionalization of the chick hindgut. *Development* 121:3163–3174.
- Roth C, et al. 2007. Evolution after gene duplication: models, mechanisms, sequences, systems, and organisms. *J Exp Zool B Mol Dev Evol.* 308B:58–73.
- Ruddle FH, et al. 1994. Evolution of Hox genes. *Annu Rev Genet.* 28:423–442.
- Salaneck E, Larsson T, Larson E, Larhammar D. 2008. Birth and death of neuropeptide Y receptor genes in relation to the teleost fish tetraploidization. *Gene* 409:61–71.
- Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol Biol Evol.* 19:101–109.
- Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19:301–302.
- Scannell DR, Butler G, Wolfe KH. 2007. Yeast genome evolution—the origin of the species. *Yeast* 24:929–942.
- Scannell DR, et al. 2006. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* 440:341–345.
- Schweitzer J, et al. 2006. Evolution of myelin proteolipid proteins: gene duplication in teleosts and expression pattern divergence. *Mol Cell Neurosci.* 31:161–177.
- Semon M, Wolfe KH. 2007a. Consequences of genome duplication. *Curr Opin Genet Dev.* 17:505–512.
- Semon M, Wolfe KH. 2007b. Reciprocal gene loss between Tetraodon and zebrafish after whole genome duplication in their ancestor. *Trends Genet.* 23:108–112.
- Shubin N, Tabin C, Carroll S. 1997. Fossils, genes and the evolution of animal limbs. *Nature* 388:639–648.
- Sifuentes-Romero I, Merchant-Larios H, Garcia-Gasca A. 2010. Hox gene expression in the embryonic genital system of the sea turtle *Lepidochelys olivacea* (Eschscholtz, 1829), a species with temperature-dependent sex determination. *Gene Expr Patterns.* 10:290–298.
- Swofford DL. 2002. Phylogenetic analysis using parsimony (\* and other methods). Version 4. Sunderland (MA): Sinauer Associates.
- Taylor JS, et al. 2003. Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Res.* 13:382–390.
- van Eenennaam AL, Murray JD, Medrano JF. 1998. Synaptonemal complex analysis in spermatocytes of white sturgeon *Acipenser transmontanus* Richardson (Pisces, Acipenseridae), a fish with a very high chromosome number. *Genome* 41:51–61.
- Vandepoele K, et al. 2004. Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. *Proc Natl Acad Sci U S A.* 101:1638–1643.
- Wagner GP, Amemiya C, Ruddle F. 2003. Hox cluster duplications and the opportunity for evolutionary novelties. *Proc Natl Acad Sci U S A.* 100:14603–14606.
- Wagner GP, et al. 2005. Molecular Evolution of duplicated ray finned fish HoxA clusters: increased synonymous substitution rate and asymmetrical co-divergence of coding and non-coding sequences. *J Mol Evol.* 60:665–676.
- Wang Z, et al. 2009. Adaptive evolution of 5′HoxD genes in the origin and diversification of the cetacean flipper. *Mol Biol Evol.* 26:613–622.
- Warot X, et al. 1997. Gene dosage-dependent effects of the Hoxa-13 and Hoxd-13 mutations on morphogenesis of the terminal parts of the digestive and urogenital tracts. *Development* 124:4781–4791.
- Werth CR, Windham MD. 1991. A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate gene expression. *Am Nat.* 137:515–526.
- Wolfe K. 2000. Robustness—it's not where you think it is. *Nat Genet.* 25:3–4.
- Zhang JZ. 2003. Evolution by gene duplication: an update. *Trends Ecol Evol.* 18:292–298.
- Zhang JZ, Zhang YP, Rosenberg HF. 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat Genet.* 30:411–415.
- Zhou RJ, Cheng HH, Tiersch TR. 2001. Differential genome duplication and fish diversity. *Rev Fish Biol Fisheries.* 11:331–337.

Associate editor: Bill Martin