

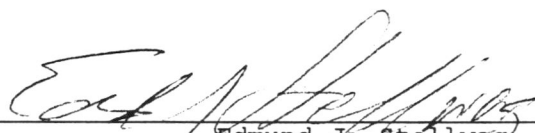
SURVEY OF HOX-LIKE GENES IN A
NEOTELEOST (STRIPED BASS): IMPLICATIONS FOR EVOLUTIONARY
EXPANSION OF THE HOX GENE FAMILY IN VERTEBRATES

by

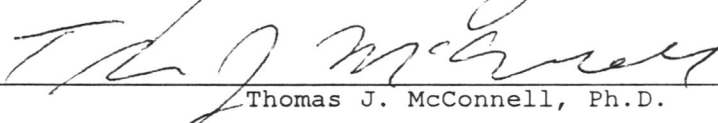
Anthony M. Pavell

APPROVED BY:

DIRECTOR OF THESIS


Edmund J. Stellwag, Ph.D.

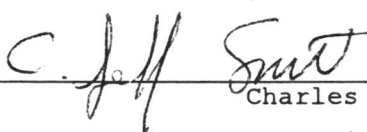
COMMITTEE MEMBER


Thomas J. McConnell, Ph.D.

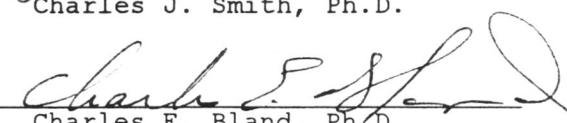
COMMITTEE MEMBER


Charles A. Singhas, Ph.D.

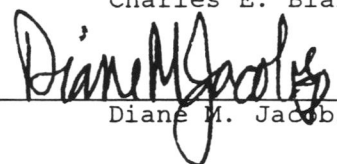
COMMITTEE MEMBER


Charles J. Smith, Ph.D.

CHAIR OF THE DEPARTMENT OF BIOLOGY


Charles E. Bland, Ph.D.

DEAN OF THE GRADUATE SCHOOL


Diane M. Jacobs, Ph.D.

Anthony M. Pavell. SURVEY OF *HOX*-LIKE GENES IN A NEOTELEOST:
IMPLICATIONS FOR EVOLUTIONARY EXPANSION OF THE *HOX* GENE FAMILY IN
VERTEBRATES. (Under the direction of Dr. Edmund J. Stellwag)
Department of Biology, April 1994.

Homeobox (*Hox*) genes encode a class of helix-turn-helix DNA-binding transcription factors characterized by a homeodomain that participates in the regulation of gene expression requisite for axial patterning and specification of segmental identity during embryonic development. Comparison of *Hox* genes among evolutionarily divergent taxa has revealed a hierarchical relationship in which increases in the number of *Hox* genes, their organization into distinct clusters, and cluster number expansion, parallel the elaboration of increasingly complex developmental systems.

The purpose of this thesis was to identify *Hox*-like genes in *Morone saxatilis*, to determine their relationship to similar *Hox* genes in other vertebrates, and to estimate *Hox* cluster number. A PCR-based method was used to survey a defined subset of *Hox*-like genes in the genome of *Morone saxatilis* (striped bass). Based on DNA sequence analysis, eight distinct *Hox*-like genes representing *Hox* clusters A, B, and C were identified from striped bass. Analysis of nucleotide substitution frequencies for selected *Hox*-like genes among four species of fish and the mouse were contrary to established phyletic relationships based on traditional and molecular taxonomic evidence. These results suggest differences in the evolution of *Hox* genes compared to other genes in the evolution of the vertebrates. Comparison of the eight *Hox*-like genes from striped bass to similar

genes from the mouse and zebrafish by computational clustering methods provided evidence consistent with current concepts regarding early evolutionary divergence of genes from the *Hox* 3' *Hox* genes among vertebrate species.

Based on annealing characteristics of the *Hox*-specific primers to corresponding target sequences in mouse and zebrafish, the recovery of *Hox* sequences related to *Hoxa*-3, a-4, a-6, a-7, b-3 to b-5, c-4, c-6 and d-3-like genes was predicted. The predicted representatives of *Hox* clusters A, B, and C were recovered, but there was a notable absence of the expected *Hoxc*-6 and d-3-like genes. These results together with a similar absence of *Hox* D-like genes in other fishes, suggests either these genes may have diverged significantly during evolution or that the *Hox* D cluster may be truncated or absent in fishes relative to mammals. It is proposed that the truncation or absence of the *Hox* D cluster in fishes may account for major evolutionary differences in developmental programs and morphology between the fish and tetrapod lineages. The relevance of these results to the understanding *Hox* D gene function in development and evolution of tetrapods is discussed.

SURVEY OF *HOX*-LIKE GENES IN A
NEOTELEOST (STRIPED BASS): IMPLICATIONS FOR EVOLUTIONARY
EXPANSION OF THE *HOX* GENE FAMILY IN VERTEBRATES

A Thesis
Presented to
the Faculty of the Department of Biology
East Carolina University

in Partial Fulfillment
of the Requirements for the Degree
Master of Science in Molecular Biology and Biotechnology

by
Anthony M. Pavell
April 1994

SURVEY OF *HOX*-LIKE GENES IN A
NEOTELEOST (STRIPED BASS): IMPLICATIONS FOR EVOLUTIONARY
EXPANSION OF THE *HOX* GENE FAMILY IN VERTEBRATES

by

Anthony M. Pavell

APPROVED BY:

DIRECTOR OF THESIS _____
Edmund J. Stellwag, Ph.D.

COMMITTEE MEMBER _____
Thomas J. McConnell, Ph.D.

COMMITTEE MEMBER _____
Charles A. Singhas, Ph.D.

COMMITTEE MEMBER _____
Charles J. Smith, Ph.D.

CHAIR OF THE DEPARTMENT OF BIOLOGY _____
Charles E. Bland, Ph.D.

DEAN OF THE GRADUATE SCHOOL _____
Diane M. Jacobs, Ph.D.

ACKNOWLEDGMENTS

There are many people who helped me succeed in completing my Master's training. I acknowledge the compassion, support, and encouragement given to me by my wife Christy. She gave me the stability I needed to maintain focus throughout the past three years. I owe a debt of gratitude to Dr. Edmund Stellwag. At the beginning of my science career, Dr. Stellwag provided me with much food for thought. The most relevant of his many words were, "if you want to think globally, you must first extract your cranium from your anus." It was this firm yet guiding hand of Dr. Stellwag's that has helped me to maintain my focus and ultimately succeed academically. I acknowledge thoughtful conversations that I had with my thesis committee members, Dr. Thomas McConnell, Dr. C. Jeff Smith, and Dr. Charles Singhas. I am also grateful to the many fellow graduate students with whom I shared lab space and who helped to break up the monotony including; Mark Steffens (Lab Golf Inventor), Jackie Peeler (Shebby Prophet), Bob Walker (Computer Golf Whiz), Keith Coon (.... at noon), and Tonya Mitchell (Anaerobe Odor Supply).

Table of Contents

List of Figures and Tables	iv
Introduction	1
Developmental Regulation of <i>Hox</i> Genes	1
Genomic Organization of <i>Hox</i> Genes	2
Evolutionary Conservation of <i>Hox</i> Genes	4
Evolutionary Expansion of <i>Hox</i> Genes	6
Materials and Methods	9
Genomic DNA Isolation	9
Polymerase Chain Reaction Ampification	10
Cloning and Sequencing	11
DNA Sequence Analysis	12
Results	13
Discussion	17
References	31
Figures	41
Tables	53

Figures

1. Multiple sequence alignment of <i>Morone</i> <i>Hoxa</i> -3-like sequences	41
2. Multiple sequence alignment of <i>Morone</i> <i>Hoxa</i> -4-like sequences	42
3. Multiple sequence alignment of <i>Morone</i> <i>Hoxa</i> -6-like sequences	43
4. Multiple sequence alignment of <i>Morone</i> <i>Hoxa</i> -7-like sequences	44
5. Multiple sequence alignment of <i>Morone</i> <i>Hoxb</i> -3-like sequences	45
6. Multiple sequence alignment of <i>Morone</i> <i>Hoxb</i> -4-like sequences	46
7. Multiple sequence alignment of <i>Morone</i> <i>Hoxb</i> -5-like sequences	47
8. Multiple sequence alignment of <i>Morone</i> <i>Hoxc</i> -4-like sequences	49
9. Unrooted evolutionary tree for selected striped bass <i>Hox</i> sequences	50
10. Unrooted evolutionary tree for selected zebrafish <i>Hox</i> sequences	51
11. Unrooted evolutionary tree for selected mouse <i>Hox</i> sequences	52

Tables

1. Comparison and homology assignments of <i>Hox</i> PCR clones from striped bass	53
2. Pair-wise comparison of <i>Hox</i> and <i>Hox</i> -like a-4 and b-5 gene nucleotide sequences	55

INTRODUCTION

In order to develop viable progeny animals need a genetic control system to specify the different cell types necessary for the formation of differentiated tissues such as muscle, bone, and nerves, to respond to signals that establish asymmetry during early embryogenesis (St. Johnston and Nüsslein-Volhard, 1992), and translate those signals into different positional fates. On the anterior-posterior (A-P) axis of most embryos, genetic control systems provide cells with specific positional identities that lead to the development of structures appropriate to the position within the field, i.e., head structures arising from the anterior region of the field and tail structures from the posterior end. A fascinating discovery is a genetic system similar to that identified originally in *Drosophila* specifies A-P positional identities in many other animals, where it apparently serves a similar function (McGinnis and Krumlauf, 1992). The core of this genetic control system consists of a set of structurally similar genes (homeotic selector genes) identified originally in *Drosophila* through mutations resulting in homeotic transformations (duplicated A-P positional values) (Kaufman et al., 1990; Lewis, 1978; Sanchez-Herrero et al., 1985). For example, deletion of the *Drosophila* Antennapedia homeotic selector gene results in the transformation of antennae into legs (Lewis, 1978), whereas mutations within the bithorax complex genes result in transformation of halteres into wings. The proteins encoded by these

homeotic selector genes are related at the structural level in the homeodomain motif.

The term "homeodomain" refers to a class of protein domains that have recognizable similarity to a 60 amino acid motif (encoded by 180 bp homeobox sequences) originally identified in three *Drosophila* homeotic and segmentation proteins (Scott et al., 1989). Those homeodomains that have been tested, through DNA binding assays, contain sequence-specific DNA binding activities and are part of larger proteins that function as transcriptional regulators (Gehring et al., 1990; Laughon, 1991; Levine and Hoey, 1988). Nuclear magnetic resonance and X-ray crystallographic studies on two homeodomains have shown their trihelical structures (α -helices I, II, and III) to be related to the helix-turn-helix motif of prokaryotic DNA-binding proteins (Kissinger et al., 1990; Otting et al., 1990). Examination of amino acid sequence similarities within the homeodomain and in flanking protein sequence revealed that homeodomain proteins and the genes that encode them could be grouped into various classes. One such class is the Antennapedia (*Antp*) class, which includes homeodomains that share 60% or more identity with the Antennapedia-type homeodomain, have identical sequences in et al., 1987; Mavillo et al., 1986; Regulski et al., 1985).

In *Drosophila*, the homeobox genes that encode *Antp* class homeodomain proteins all map to the Antennapedia and Bithorax complexes, which are collectively referred to as the Homeotic

complex, or HOM-C (Akam, 1989). These *Antp* class genes include labial (*lab*), proboscipedia (*pb*), Deformed (*Dfd*), Sex combs reduced (*Scr*), Antennapedia (*Antp*), Ultrabithorax (*Ubx*), and abdominal-A (*abd-A*). Recessive and dominant mutations in these genes result in homeotic transformations of the embryonic and the adult body plan, and are usually lethal. The *Drosophila* HOM-C contains another homeotic selector gene, Abdominal-B (*Abd-B*), whose structure does not correspond to the established criteria of the *Antp* class, but whose sequence is more closely related to *Antp* class proteins than to other structural classes of homeodomain proteins, such as the engrailed, POU, or paired classes (Scott et al., 1989). Other genes encoding non-*Antp* class homeobox proteins (e.g., *bicoid*, *fushi-tarazu*, *zerknüllt*) map within the *Drosophila* HOM-C region, as do genes that are structurally unrelated to the homeobox class (e.g. *amalgam*) (McGinnis and Krumlauf, 1992). The occurrence of additional transcripts from HOM-C genomic DNA sequences suggests that other genes remain to be characterized, although it is likely that all homeobox genes, in *Drosophila*, that can mutate to lethality have been identified (Kaufman et al., 1990; Kuriowa et al., 1985; Sanchez-Herrero et al., 1985).

The *Drosophila* homeotic genes are organized in the HOM-C, wherein the physical order of genes is colinear with their expression along the anterior-posterior axis, and all except *Dfd* are transcribed in the same direction (Lewis, 1978; McGinnis and Krumlauf, 1992).

The most 3' gene of the complex, *lab*, is required for the most anterior HOM-C functional domain (in *lab* mutants structures from the intercalary segment of the head are missing and there are other head defects), and the most 5' gene of the complex, *Abd-B*, is required for the normal development of posterior abdominal segments (Duncan, 1987; Kaufman et al., 1990). The only gene whose domain of function and expression cannot be correlated directly with its position in the HOM-C is proboscipedia (Putz et al., 1988). Homeotic expression domains include and usually extend beyond the boundary of the principal functional domains.

Antp class homeobox genes have been identified in a wide variety of metazoan genomes (Kappen et al., 1993). In the mouse and human genomes, the *Antp* class homeobox (*Hox*) genes show striking similarities in their organization and expression to the selector genes of the *Drosophila* HOM-C (Duboule and Dollé, 1989; Graham et al., 1989). *Antp* class *Hox* genes are clustered in four unlinked gene complexes (*Hoxa* through *Hoxd*; Boncinelli et al., 1988, 1991; Schughart et al., 1988). Individual genes of the mouse *Hox* complexes are arranged and expressed in the same relative order as their structural homologs in the HOM-C. In particular, it is possible to confidently define structural homology relationships between the *Drosophila lab*, *pb*, *Dfd*, and *Abd-B* genes and specific *Hox* genes (Duboule and Dollé, 1989; Graham et al., 1989; Izpisua-Belmonte et al., 1991; Regulski et al., 1987; Rubock et al., 1990). The evidence

for structural homology between specific *Drosophila* HOM-C genes and mammalian *Hox*-C genes is mainly limited to the homeodomain and immediate flanking regions, although for some genes (e.g., *Dfd* and its vertebrate paralog group 4 genes), a moderate amount (60%) of structural similarity exists outside the homeodomain region (Graham et al., 1988; Regulski et al., 1987; Schubert et al., 1993). For *Scr*, *Antp*, *Ubx*, and *Abd-A*, it is more difficult to identify individual homologs in each murine complex on the basis of sequence similarities. The *Scr*, *Antp*, *Ubx*, and *Abd-A* homeodomain sequences are very similar to each other and to unrelated *Hox* genes, and sequence similarities outside the homeodomain are not extensive enough to clearly assign homology relationships, although they are suggestive (McGinnis and Krumlauf, 1992).

Genomic mapping and DNA sequence analysis of *Hox* genes among evolutionarily disparate phyla has revealed a hierarchical relationship between the number of *Hox* genes per cluster, the number of *Hox* clusters per genome, and the developmental repertoire (Holland, 1992; McGinnis and Krumlauf, 1992; Pendleton et al., 1993). Thus, it was proposed that *Hox* genes were organized into a single cluster composed of either 3 or 4 genes in organisms prior to the divergence of the diploblasts and triploblasts (Schubert et al., 1993). Subsequent intra- and inter-cluster duplication events in the progenitors of primitive cephalochordates are postulated to have generated at least 2 clusters, each composed of between 11 and 13 *Hox*

genes (Schubert *et al.*, 1993; Schughart *et al.*, 1989a). Further cluster duplication events, punctuated by loss of individual cognate genes, generated four clusters represented by nearly 40 *Hox* genes in extant mammalian species (Kappen *et al.*, 1989; Schughart *et al.*, 1988).

Although experimental evidence supporting general aspects of *Hox* gene and cluster expansion during evolution have been provided, the specific nature of the evolutionary events leading to the appearance of four *Hox* clusters in extant mammalian species remain unclear. It has been proposed that genomic duplication within vertebrate progenitors, presumably those expressing only two *Hox* clusters, resulted in the appearance of four *Hox* clusters in all vertebrates (Ohno 1970; Schughart *et al.* 1989a). However, more recent evidence, based on DNA sequence analysis of PCR-amplified *Hox* genes from extant species, including primitive vertebrates and their presumed evolutionary predecessors, suggested it was equally likely that cluster duplication involved an arithmetic progression from one to four clusters in which cluster expansion was marked by appearance of chordate and vertebrate characters (Pendleton *et al.* 1993). If the *Hox* gene system in vertebrates evolved by genomic duplication from a progenitor with two *Hox* clusters, all vertebrates, regardless of developmental complexity would be expected to have four clusters. However, if the *Hox* gene system evolved arithmetically, three clusters might be expected, particularly, but not exclusively, in

those vertebrates expressing depauperate developmental characteristics relative to advanced tetrapods. Support for the arithmetic amplification model is provided by several recent studies involving the analysis of *Hox* gene sequences in the agnathan *Petromyzon marinus* (lamprey), a primitive aquatic vertebrate lacking a well developed skull and paired appendages, and the euteleost *Danio rerio* (zebrafish), a ray-finned fish with a well developed skull and paired appendages (Molven et al., 1990). Although these two aquatic vertebrates differ in developmental complexity, evidence suggests there are only three *Hox* clusters (*Hox* A, B, and C) in each species, a result consistent with an arithmetic amplification model of *Hox* gene-related vertebrate evolution. In contrast, *Hox* genes in both amphibians and mammals have been assigned to each of the four recognized *Hox* gene clusters (*Hox* A, B, C, and D); suggesting, either divergent cluster duplication events for non-mammalian aquatic and terrestrial vertebrate lineages; a common vertebrate ancestor with three clusters; or incorrect estimation of *Hox* cluster number for aquatic vertebrates. Although it is difficult to discriminate among these likely alternatives, a more complete understanding of vertebrate *Hox* cluster composition and number would provide additional information concerning the evolutionary steps leading to *Hox* cluster organization among extant vertebrates and further insight into the function of the four *Hox* clusters found in tetrapod species. An objective of my investigation was to examine the *Hox* gene-related

sequences of a neoteleost fish striped bass, both to discern the most probable number of *Hox* clusters, and to compare the characteristics of *Hox* genes in a higher teleost with those of other vertebrates, particularly fishes and tetrapods.

Based on the high degree of sequence conservation of vertebrate *Hox* genes among distantly related species (Akam, 1989; Kappen et al., 1993), definite homologues of human and murine *Hox* genes have been identified in avians (Wedden et al., 1989), amphibians (Fritz et al., 1989; Harvey et al., 1986), fishes (Njølstad et al., 1990) and invertebrates (Lobe and Gruss, 1989). The highly conserved nature of *Hox* genes allowed Pendleton et al. (1993) to develop a PCR-based screening assay facilitating identification and characterization of *Hox*-like genes in hemichordates, cephalochordates, and primitive vertebrates.

I have used a similar approach to survey a defined subset of *Hox*-like genes, i.e. those genes related to mouse *Hoxa*-3 to a-7, b-3 to b-7, c-4 to c-6, and d-3 cognate groups in striped bass, a higher teleost fish. My results show that representatives of *Hox* genes corresponding to *Hoxa*-3 to a-7, b-3 to b-5, c-4 and c-6 appear to be present in the genome. The notable absence of a predicted *Hoxd*-3-like gene, in conjunction with statistical cluster analysis of the *Morone saxatilis* *Hox* genes, indicate either truncation or absence of the *Hox* D cluster. This view is consistent with my analysis of *Hox* D cluster genes in other fish, and recent analysis of *Hox* D function in

mice and chickens (Dollé et al., 1993; Lufkin et al., 1992; Morgan et al., 1992; Pendleton et al., 1993). I propose that the truncation or absence of *Hox D* cluster genes in fishes may be related to evolutionary differences in the developmental repertoire and morphology between the fish and tetrapod lineages.

MATERIALS AND METHODS

DNA Isolation

Genomic DNA was isolated from whole blood (100 µl) *Morone saxatilis* (striped bass) according to the protocol described by Kan et al. (1977). To 100 µl of whole blood was added 100 µl of lysis buffer composed of 640 mM sucrose, 20 mM Tris (pH 7.8), 10 mM MgCl₂, and 2% Triton X-100. After lysis nuclei were collected by centrifugation for 20 min at 2,000 rpm (4°) using a Sorvall HB-4 rotor. The nuclear pellet was suspended in 2 ml saline-EDTA-SDS (75 mM NaCl, 24 mM EDTA 0.5% SDS) and treated with proteinase K (1 mg/ml) at 54°C overnight. Proteinase K treated samples were extracted once with an equal volume of phenol, phenol:chloroform:isoamyl alcohol (25:24:1 v/v), and chloroform. DNA was precipitated from the aqueous phase by addition of 2 volumes of absolute ethanol, collected by centrifugation, dried and suspended in 2 ml 10 mM Tris (pH 7.5), 0.1 mM EDTA.

Polymerase Chain Reaction (PCR) Amplification

Genomic DNA (100 ng) was used as a substrate for the PCR. The primers used were:

Hox48-71: 5'-GAGAAAGAGTTYCATTTCAACCG-3'

Hox141-163: 5'-CATGCGICGRTTCTGGAACCAG-3'

(Y = C or T, R = A or G, and I = inosine). These primers were directed toward highly conserved regions of the homeobox and were based on conserved sequences within the *Drosophila melanogaster Antennapedia* and zebrafish *Hox* B5, B6, C5 and C6 gene cognate groups. The Hox48-71 primer corresponds to basepairs 48-71 within helix 1 of the *Drosophila Antennapedia* homeobox, whereas the Hox141-163 primer corresponds to a region within recognition helix 3. Together the forward and reverse *Hox* primers had an overall degeneracy of 16-fold, thus providing the potential for highly specific amplification of defined target *Hox* sequences. The PCR reaction buffer consisted of 20 mM Tris (pH 7.5), 10 mM KCl, 10 mM (NH₄)₂SO₄, 1.5 mM MgCl₂, 0.1% Triton X-100, final reaction concentrations. In addition to buffer, reaction mixtures contained 30 pmol of each primer, 2.5 units *Taq* DNA polymerase (Stratagene, La Jolla, CA), 200 μ M of each dNTP, 100 μ g BSA, and template DNA in a total volume of 100 μ l. PCR was carried out for 30 cycles in which the cycling parameters were 94°C for 1 min, 52°C for 1 min, 72°C for 1 min using an M.J. Research thermal cycler (M.J. Research Minicycler, Watertown, MA). A single band corresponding in length to approximately 114 bp was observed

upon agarose and polyacrylamide gel electrophoretic analysis of the PCR products.

Cloning and Sequencing of PCR products

PCR products were extracted twice with phenol/chloroform/isoamyl alcohol (50:48:2 vol/vol) and ethanol precipitated. PCR products at a concentration of 100 ng/ μ l were ligated directly into the pGEM-T vector (pGEM-T, Promega; Madison, Wisconsin) using a insert-to-vector ratio of 5. Ligation products were transformed into *Escherichia coli* XL1 Blue and 120 insert-containing transformants were selected for further analysis. The number of insert containing transformants to be screened was established according to the formula: $N = \ln(1-P) / \ln(1-f)$ where N is the necessary number of recombinants to be sampled, P is the desired probability, and f is the fractional proportion of the genome in a single recombinant (in this case $P=0.999$ and $f=1/10$ where 10 is the number of *Hox* genes predicted to be amplified; Clarke and Carbon, 1976). Plasmids containing inserts were sequenced using the PRISM Ready Reaction Dye-deoxy Terminator Cycle Sequencing Kit (Applied Biosystems Inc.). The PRISM protocol was shortened to include twenty rather than the recommended 25 cycles. The sequencing reaction products were analyzed on an automated DNA sequencer (Model 373A, Applied Biosystems Inc.). The predicted PCR product of 114 base pairs yielded 69 bp of informative DNA sequence.

DNA Sequence analysis

DNA sequences from 100 independent clones were compiled and grouped according to nucleotide sequence data using the Genetics Computer Group programs (GCG; Devereux *et al.*, 1984). A single representative DNA sequence from each group was compared to homeobox sequences from other vertebrates using the BLASTx search program (NCBI; Altschul *et al.*, 1990; Gish and States, 1993). Based on BLASTx search results the most homologous mouse homeobox sequence was identified and used to provide a tentative *Hox*-like gene assignment for the corresponding striped bass sequence. Within each nucleotide sequence group, certain individual DNA sequences differed by 1 or 2 random nucleotide changes. Most of these differences were attributed to sequencing or amplification errors.

Nucleotide substitution frequencies for the *Hox* and *Hox*-like a-4 and b-5 genes from the mouse, striped bass, zebrafish, goldfish, and lamprey were determined by pairwise reciprocal comparison of DNA sequences bracketed by my forward and reverse PCR primers. The total number of nucleotide substitutions were then divided by the total number of nucleotides to determine the frequency of substitution per base pair.

Unrooted evolutionary trees for mouse, striped bass and zebrafish *Hox* and *Hox*-like sequences were constructed from nucleotide substitution matrices using the two-parameter model of Kimura (1980) from the REAP program (version 4.0; McElroy *et al.*, 1991). *Hox* and

Hox-like gene sequences were clustered by a neighbor-joining algorithm using the NJOIN program of NTSYS-pc (version 1.70; Rohlf, 1992).

RESULTS

Based on the annealing characteristics of the *Hox*-specific primers, as determined by comparison to corresponding target sequences in the mouse *Hox* cognate groups, and assuming conservation of cognate groups and a maximum of two base substitutions within the region corresponding to the PCR primer annealing sites, the amplification of sequences similar to ten *Hox* genes representing four *Hox* clusters was predicted (Geada et al., 1992; Graham et al., 1988; Kessel et al., 1987; Krumlauf et al., 1987; Lonai et al., 1987; Mavillo et al., 1986; Schughart et al., 1989b; Wolgemuth et al., 1986); including *Hoxa*-3, a-4, a-6, a-7, b-3 to b-5, c-4, c-6, and d-3. However, based on published target sequences from zebrafish *Hox* cognate groups (Molven et al., 1992; Njølstad et al., 1988a, 1988b, 1988c, 1990; Runstadler and Kocher, 1991), amplification of only eight *Hox* genes representing three *Hox* clusters; including *Hoxa*-3, a-4, a-6, a-7, b-3 to b-5, and c-4 was predicted. Alignment of nucleotide sequence data obtained from 103 independent clones provided evidence for eight distinct sequences corresponding to *Hox* genes a-3, a-4, a-6, a-7, b-3 to b-5 and c-4, but not c-6 or d-3 (Table 1). The number of individual clones corresponding to each

Hox-like gene was as follows: a-3, 15 clones (Figure 1); a-4, 1 clone (Figure 2); a-6, 27 clones (Figure 3); a-7, 2 clones (Figure 4); b-3, 5 clones (Figure 5); b-4, 20 clones (Figure 6); b-5, 29 clones (Figure 7); and c-4, 1 clone (Figure 8; Table 1). In addition to the 100 *Hox*-like clones, 3 clones derived from PCR amplified DNA had inserts greater than 114 bp in length (250, 330, and 1050 bp). However, none of these three sequences were similar to *Hox* sequences (A.M.P., personal observations). Although the number of striped bass clones corresponding to *Hox* genes from mice or zebrafish varied significantly, ranging from a low value of one for *Hoxa*-4 and c-4 to an upper value of 29 for *Hoxb*-5, 100% of the expected sequences related to zebrafish *Hox* genes and 80% of the sequences related to mouse *Hox* genes were recovered (Table 1). The high efficiency recovery of predicted *Hox*-like genes from striped bass provided direct support for the effectiveness of my PCR-based assay.

Based on informative DNA sequence bracketed by the forward and reverse PCR primer sequences it was determined *Hox*-like genes a-4, a-6, a-7, b-3, b-4, and c-4 were represented by single or multiple invariant DNA sequences (Table 1). However, sequences corresponding to *Hox*-like genes a-3 and b-5 were divided into two subgroups each, a-3.1 and a-3.2 as well as b-5.1 and b-5.2. The *Hoxa*-3.1 subgroup (11 clones) was differentiated from the *Hoxa*-3.2 subgroup (4 clones) on the basis of a C to G transversion at bp 92 relative to the *Drosophila melanogaster* Antennapedia homeobox

sequence, whereas the *Hoxb*-5.1 (17 clones) and *Hoxb*-5.2 (12 clones) subgroups were differentiated by a C to T transition at bp 87. These results suggested that the *Hoxa*-3.1 and -3.2 as well as the *Hoxb*-5.1 and -5.2 subgroups represented alleles at the *Hoxa*-3 and *Hoxb*-5 loci.

Reciprocal pairwise comparisons of the similarities and differences between the nucleotide sequences in the region between the forward and reverse primer annealing sites for the *Hox* and *Hox*-like genes a-4 and b-5 genes of the mouse, striped bass, zebrafish, goldfish and lamprey were used to estimate the extent of *Hox* gene divergence among these species. These two genes were chosen for analysis because they are the only two characterized *Hox* genes common to each of these vertebrate species. Comparisons to the mouse revealed far less nucleotide substitution between the *Hoxa*-4 and b-5 like genes from the lamprey, a petromyzontiform, and striped bass, a perciform fish, than between these two genes in the zebrafish and goldfish, two cypriniform species (Table 2). However, nucleotide substitution frequencies between the *Hoxa*-4 and b-5 sequences were far less frequent between the lamprey and the mouse than between the lamprey and any of the teleosts (Table 2). A restricted comparison including only the teleosts demonstrated there was less nucleotide substitution between the *Hoxa*-4 like sequences of the striped bass and goldfish, members of two different taxonomic orders, than between the zebrafish and goldfish, members of the same taxonomic family (Table 2). The virtually identical nucleotide substitution

frequencies of *Hoxb-5* like sequences among the teleosts contrasted markedly with the *Hoxa-4* results and demonstrated that the *Hox* nucleotide substitution frequencies between these two genes were dissimilar in identical pairwise comparisons between species (Table 2).

Comparative analysis of the eight *Hox*-like genes from *Morone saxatilis* to related genes from mice and zebrafish by computational clustering yielded trees composed of two major branches. Similar analyses of lamprey, goldfish, and salmon were not conducted for various reasons including the absence of genes from an entire *Hox* cognate group in the lamprey (Pendleton et al., 1993), restricted representation of *Hox* central class genes in the goldfish (Levine and Schecter, 1993), and a very limited number of characterized *Hox* genes (3 genes) in Atlantic salmon (Fjose et al., 1988, 1990). Of or among the species analyzed, *Hox* genes linked to the two major branches corresponded either to members of the 3' (*Hox* cognate group 3 genes) or central class (*Hox* cognate groups 4-7) of homeobox genes (Figs. 9-11). *Hox* cognate group 3 and 3-like sequences from all three species were located on a distinct branch separate from the other major branch represented by *Hox* central class genes. Comparison of related mouse and striped bass *Hox* and *Hox*-like central class genes revealed considerable variation in the topology of the branches linked to these genes. Despite the variability in topology of the branches corresponding to central class genes, *Hoxb-5* and *c-4* genes from mouse

and striped bass shared a single branch. Direct comparison of the branching patterns between striped bass and zebrafish central class *Hox* and *Hox*-like genes was precluded because only 6 of the 13 characterized *Hox*-like central class genes were shared between these two species.

I predicted the isolation of sequences similar to mouse *Hoxc*-6 and d-3 based on intracognate sequence conservation of the mouse *Hox* genes. Although a significant number of sequences corresponding to *Hox* genes in both cognate groups were identified, I was unable to recover sequences corresponding to the mouse *Hoxc*-6 and d-3 genes.

DISCUSSION

The nearly 40 *Hox* genes constituting the 4 *Hox* clusters of extant vertebrate species are postulated to have been derived from a single ancestral *Hox* gene by repeated duplication (Schubert et al. 1993). As a consequence, *Hox* genes in each of the four clusters are evolutionarily related and form paralogous groups. The extent to which the identity of *Hox* gene paralogs is preserved reflects the evolutionary relationships among individual genes and the degree to which different genes in the *Hox* network differ.

PCR-based surveys of *Hox* genes offer a relatively rapid and widely applicable method for identification of *Hox* type genes among diverse taxa (Levine and Schecter, 1993; Murtha et al., 1991;

Pendleton et al., 1993). As such, these surveys provide valuable information for comparative studies of *Hox* gene evolution. However, application of this methodology is restricted to so few species that details of *Hox* gene number and evolution, particularly among vertebrates, remain uncertain (Levine and Schecter, 1993; Murtha et al., 1991; Pendleton et al., 1993). In this study a modification of the PCR-based *Hox* gene identification method developed by Murtha et al. (1991) was employed to screen for a defined subset of *Hox*-like genes from striped bass. Based on the primer annealing sequences from the *Hox* genes of zebrafish and the sequence of the *Hox*-specific PCR primers, 100% of the predicted *Hox*-like genes were recovered from among my striped bass clones. The eight *Hox*-like genes recovered from striped bass represented three *Hox* clusters A, B, and C, and they numerically equal, or exceed, the number of *Hox*-like genes characterized for any aquatic vertebrate with the exception of the lamprey. Isolation and characterization of the eight *Hox*-like genes from striped bass provides the opportunity to compare *Hox* genes among vertebrates. Further, recovery of all the genes targeted by the assay provides direct support for the accuracy of the PCR-based method and suggests this method may provide a generally applicable and effective approach for identification of defined subsets of *Hox* genes among related vertebrate species.

Based on the theoretical annealing characteristics of the *Hox*-specific PCR primers, amplification of only those templates with

fewer than two mismatches in the primer annealing sequences of target genes was predicted. Because all *Hox*-like genes predicted based on zebrafish were recovered, the results suggest the primer annealing regions of the *Hox*-like genes a-3, a-4, a-6, a-7, b-3 to b-5, and c-4 from zebrafish and striped bass diverged less than 10% since the evolutionary separation of the cypriniform and perciform fishes. In this regard, it is noteworthy that only 80% of the predicted mouse *Hox* genes targeted by the assay were recovered from striped bass. The two genes related to mouse *Hoxc*-6 and d-3 were not identified from the clone collection. Provided the number of PCR-derived clones surveyed (103 clones) and the predicted number of *Hox* genes available as targets for amplification based on mouse (10 *Hox* gene targets), there was a 99.997% probability that each of the 10 target genes should have been recovered if present in the clone collection (Clark and Carbon, 1977). Although striped bass homologs of the mouse *Hoxc*-6 and d-3 may have been absent from the clone collection as a result of evolutionary gene loss or intron insertion precluding their amplification, the results suggest evolutionary divergence within the primer annealing target sequences as a likely explanation for their absence. Moreover, my results suggest that the primer annealing sequences of the mouse diverged from striped bass more extensively than the primer annealing sequences of the zebrafish.

Reciprocal pairwise comparison of nucleotide substitutions in the region between the forward and reverse primer annealing sites

from the *Hox* and *Hox*-like a-4 and b-5 genes of the mouse, striped bass, zebrafish, goldfish and lamprey revealed these regions differed in the extent of nucleotide substitution in a species and gene dependent manner. Unexpectedly, nucleotide substitution frequencies among the four species of fish and the mouse reflected differences that were contrary to established phyletic relationships (Jamieson, 1991). Based on current views of vertebrate taxonomy, the lamprey, an agnathan, is thought to have diverged from the gnathostomes, represented by the mouse, striped bass, zebrafish and goldfish, prior to the divergence of the teleosts (striped bass, zebrafish, and goldfish) and the tetrapod, mouse (Løvtrup, 1977). However, comparison of the *Hoxa*-4 and b-5 like sequences from the five vertebrate species, demonstrated a lower nucleotide substitution frequency between the *Hoxa*-4 and b-5 like sequences of the lamprey and the mouse than between the lamprey and the teleosts (Tables 2 and 3). This result suggests a closer relationship between the *Hoxa*-4 and b-5 genes of the lamprey and the mouse than between the lamprey and the teleosts. Further, comparisons of the nucleotide substitution frequencies of the *Hoxa*-4 and b-5 sequences within the teleosts revealed reduced levels of nucleotide substitution between members of two different taxonomic orders, the perciforms (striped bass) and cypriniforms (zebrafish and goldfish) than between the individual cypriniforms. These results are contrary to accepted views of vertebrate phylogeny based on traditional and molecular

taxonomic evidence (Hillis and Dixon, 1989; Løvtrup, 1977). It is interesting to speculate that *Hox* gene evolution among vertebrates transcends the boundaries of traditional phylogenetics.

Evolutionary trees were constructed for those species from which a sufficient number of related 3' and central class *Hox* genes were available. Consequently, evolutionary trees were prepared for the mouse, zebrafish, and striped bass, but not the lamprey (no cognate group 3 genes), goldfish (paucity of central class genes), or Atlantic salmon (only three *Hox* genes characterized). Comparison of evolutionary trees revealed a similar deep division separating branches corresponding to the 3' and central class *Hox* genes were available. Consequently, evolutionary trees were prepared for the mouse, zebrafish, and striped bass, but not the lamprey (no cognate group 3 genes), goldfish (paucity of central class genes), or Atlantic salmon (only three *Hox* genes characterized). Comparison of evolutionary trees revealed a similar deep division separating branches corresponding to the 3' (cognate group 3) and central (cognate groups 4-7) class genes in all three species. This arrangement is consistent with the view that members of the 3' class were related and diverged from the central and 5' classes of the ancestral homeobox complex early in metazoan evolution (Schubert et al., 1993). Further, deep subdivisions within individual cognate group 3 genes of striped bass and the mouse suggest that, although related, these genes diverged from one another early in the evolution

of *Hox* cognate genes. Shallower subdivisions among members of the cognate group 4 relative to the cognate group 3 genes from striped bass and the mouse provides evidence that indicates cognate group 4 genes diverged from one another later than corresponding cognate group 3 genes.

Further comparisons of evolutionary trees were restricted to the mouse and striped bass because they were the only two species for which a sufficient number of related *Hox* 3' and central class genes were available for analysis. Considerable variation in the topology of the branches linking *Hox* and *Hox*-like central class genes suggests marked differences in the evolutionary divergence among these genes between striped bass and the mouse. However, preservation in the topology of the branching patterns corresponding to *Hoxb*-5 and *c*-4 in evolutionary trees from both striped bass and the mouse indicates a possible restricted evolutionary divergence between these two genes compared to other *Hox* central class genes.

My analysis and comparison of *Hox* gene sequences from various fish and tetrapod species revealed a distinct difference in the genomic complement of *Hox* cluster genes between these two groups of vertebrates. In addition to this study, one agnathan and three other teleost species have been examined for genomic composition of *Hox* genes including; lamprey, goldfish, zebrafish, and the Atlantic salmon. The four fish species all differ in the number of *Hox* genes isolated and cluster representation. In the lamprey, Pendleton et

al.(1993) provide evidence for three *Hox* clusters but state that the presence of a fourth cluster could not be rigorously excluded. Conclusions in the lamprey study must be considered tentative because a statistically significant number of clones was not sampled. The PCR survey attempted to identify all (40) *Hox*-related sequences, however, only 97 clones were sampled (80% identification probability) instead of the 400 clones necessary to be screened to assure a 99% probability of identifying all *Hox*-like genes. Interestingly though, none of the clones that were isolated contained sequences similar to any *Hox D* gene.

Levine and Schecter (1993) used a PCR approach similar to that described here, but just as Pendleton et al (1993) they only sampled a small number of clones (52). Levine and Schecter (1993) believe that they had isolated sequences in the goldfish that were similar to human *Hox D* genes. In my interpretation of their data, I discovered that the clones they identified as similar to *Hoxd-10* were actually most similar to *Hoxc-10*. The authors also identified two sets of clones that they classified as similar to *Hoxb-5*. Because the authors provide evidence for alleles to each of the b-5-like groups and it is only possible to have one allele per gene (goldfish are diploid), it is unlikely that the goldfish genome contains four copies of a *Hoxb-5*-like gene. Based on amino acid similarity between the two goldfish b-5-like groups and murine *Hox* genes, I have reclassified one of the groups as a *Hoxa-4*-like group. Therefore,

this apparent misclassification of *Hox*-like clones (*Hoxd*-10 and two *Hoxb*-5 like clones) does not provide evidence for the existence of four *Hox* clusters in fish, but it does agree with other data suggesting only three *Hox* clusters in fish.

Drosophila Antp and *Scr* homeoboxes have been used as probes to identify and clone *Hox* genes from zebrafish and Atlantic salmon recombinant DNA libraries (Fjose et al., 1988, 1990; Njølstad et al., 1988a, 1988b, 1988c, 1990). *Hox* genes representative of clusters A, B, and C have been identified in both species using the *Drosophila* probes. Again, notably absent is any evidence for the existence of *Hoxd*-like genes or a fourth *Hox* cluster in fish. Therefore, fish species, including members of the most primitive and advanced lineages, appear to be notably devoid of *Hox* D-like sequences.

Based on statistical clustering, chromosomal arrangement, and nucleotide sequence evidence from *Drosophila*, *Tribolium*, *C. elegans*, mice and humans, the progenitor to the vertebrates and invertebrates seems to have possessed a single homeobox gene complex (Schubert et al., 1993). Throughout invertebrate evolution a single HOM cluster has been maintained, and alternative RNA splicing mechanisms are used to generate the numerous gene products necessary to specify position and cytodifferentiation within the invertebrate embryo (Schubert et al., 1993). Conversely, the vertebrate lineage has expanded the number of individual *Hox* genes through amplification of individual clusters in order to generate the requisite number of developmentally

important genes (Schubert et al., 1993). However, the mechanism of cluster expansion from one *Hox* cluster to the four clusters present in extant mammals is not known. The favored hypothesis is that *Hox* cluster number increased from one to four in all vertebrates by geometric amplification accomplished through total genome duplication (Ohno, 1970; Schughart et al., 1989a). Alternatively, arithmetic amplification of the *Hox* clusters is equally feasible and tentatively supported by the apparent lack of *Hox* D-like genes in fish. A third possibility is that all vertebrates possess four *Hox* clusters, obtained through geometric amplification, and that there has been a secondary loss or truncation of the fourth *Hox* complex in fishes. The reasoning supporting geometric *Hox* cluster amplification is based on concordant duplication of all gene loci creating no problem with regard to the dosage relationship of functionally related genes and their regulatory cistrons. Gene dosage imbalance does have profound effects on the vitality of an organism and total genome duplication is indeed a mechanism of avoiding such problems (Ohno, 1970). However, if related regulatory and structural genes are juxtaposed on the chromosome or if a gene has autoregulatory capabilities, then it or they are not limited to functional amplification only through total genome duplication because duplication of that region results in proportional duplication of related regulatory and structural genes. Autoregulatory genes along with physically linked structural and regulatory genes would be able to be amplified through tandem

duplication along a chromosome or duplication of an individual chromosome. Such a system of linked auto- and cross-regulatory genes exists in the *Hox* and HOM gene networks (McGinnis and Kuziora, 1994). For example in *Drosophila*, the Deformed gene is autoregulated; the Ultrabithorax protein represses the expression of the Antennapedia gene; and Abdominal-B regulates itself and other genes in the HOM complex, including Antennapedia, Ultrabithorax, and Abdominal-A. Furthermore, transgenic expression of human and mouse *Hox* genes in *Drosophila* exhibit similar patterns of expression or mis-expression suggesting that vertebrate *Hox* genes regulate themselves and others similarly to the methods used in invertebrates (McGinnis and Kuziora, 1994). Since many of the *Hox* genes possess auto- and cross-regulatory functions, duplication of a single cluster would not result in an imbalance between the expression of regulatory and structural genes. Therefore, assuming that total genome duplication is the only mechanism to alleviate problems associated with gene dosage instability is fallacious. It is possible that *Hox* cluster amplification and genome expansion proceeded via arithmetic amplification.

Differences in the number of *Hox* clusters among extant vertebrate species representative of fish and tetrapod lineages would suggest that *Hox* cluster evolution in vertebrates may have proceeded along either congruent or divergent paths. The presence of *Hox* clusters A, B, and C, but not D in fishes would be consistent with a

congruent evolutionary path marked by arithmetic *Hox* cluster amplification in which the progenitor of the fishes possessed two *Hox* clusters and the fishes, with three *Hox* clusters, served as progenitor to the tetrapods. Alternatively, it is equally possible that *Hox* cluster evolution in vertebrates proceeded along divergent paths by way of a common progenitor possessing two *Hox* clusters. Proceeding from a common two *Hox* cluster-containing progenitor, the fishes could have arisen by arithmetic amplification of the *Hox* clusters, whereas the tetrapods could have arisen by geometric amplification. Although it is impossible to discriminate among these alternative *Hox* gene expansion mechanisms, further elucidation of *Hox* gene organization and cluster number, particularly *Hox D* genes, in evolutionary pivotal species will provide additional insight into the possible evolutionary dynamics of *Hox* cluster amplification.

Despite a high probability of recovery, based on nucleotide sequence conservation in other vertebrates, I was unable to identify sequences in *Morone* that were similar to *Hoxd-3*. My recovery of clones from *Hox*-like genes that are evolutionarily more divergent than the *Hox 3* cognate group infers that the more conserved *Hox* cognate group 3-like genes, including the presumptive *Hoxd-3*-like gene, should have been recovered, if present. The likely absence of a *Hoxd-3*-like paralog in *Morone* is compatible with my sequence analysis of *Hox*-like genes in other fishes. This analysis revealed that *Hoxd-3*-like sequences, and perhaps all *Hox D*-like sequences, are

absent in those fish species examined. Based on the results of the present and previous studies of *Hox* genes in fish, it would be consistent with empirical observations to propose that either the *Antennapedia*-related *Hox D* cluster central class region or, more generally, the entire *Hox D* cluster is absent in fish. Although rigorous proof of the absence of the *Hox D* cluster in fish will require additional genomic mapping and expression studies, the available evidence suggests the *Hox D* cluster may be truncated or absent in fishes.

Truncation or absence of the *Hox D* cluster in fishes would also provide an explanation for the fundamental differences in development and morphogenesis between fishes and tetrapods. It has been established that mammalian *Hox D* cluster genes are expressed during critical phases of head morphogenesis (Hunt *et al.*, 1991) and limb ontogeny (McGinnis and Krumlauf, 1992). Disruption of the *Hoxd-13* gene in the mouse has been demonstrated to result in defective limb development, manifested by reduction in morphological complexity of limb structures characteristic of neoteny. Moreover, expression of the *Hoxd-11* gene under control of the type-A replication-competent avian retroviral vector (RCAS) promoter after transfection into chicken limb buds resulted in posterior homeotic transformation in which an anterior digit was converted to a more posterior-like digit along with reduction of the distal leg segment and aberrations in foot morphology (Morgan *et al.*, 1992). In addition to the limb

skeletal abnormalities induced by disruption or mis-expression of *Hox* D genes, ectopic expression of the *Hoxd-4* gene under the control of the *Hoxa-1* promoter in mice resulted in a homeotic posteriorization of the occipital bones in which affected mice were observed to have developed a skull resembling that of the adult lamprey, a primitive jawless fish (Lufkin *et al.*, 1992). These observations suggest that *Hox* D cluster genes are involved in specification of limb and skull morphology in advanced tetrapods, and that defects in *Hox* D gene function are evident as palingenetic reversions, perhaps reflecting ancestral phylogenetic relationships among vertebrates. Based on these results, truncation or absence of *Hox* D cluster genes would be expected to severely restrict the developmental potential of tissue primordia responsible for specification of those limb and cranial structures characteristic of tetrapods. This view of *Hox* D gene function suggests fish may lack the complement of *Hox* genes necessary for elaboration of tetrapod-related structural features including limbs and a neocranium. According to this view, it would be appropriate to conclude that the tetrapod limb and fish fin are evolutionarily analogous. In support of this interpretation, is the observation that the mesenchyme of the fin bud is derived from mesodermal and neural crest tissue, whereas the mesenchyme of the tetrapod limb is derived solely from mesodermal tissue (Tabin, 1991; Thorogood, 1991). Furthermore, the transition from fin to limb involves loss of any trace of the dermal (neural crest or mesodermal)

component and considerable variability would be expected in the fin-limb transition in a situation where the basal cartilaginous elements were expanding to replace apical elements (Hall, 1991). Within the fishes the differences between the two fin types (ray and lobed) are so great, however, that it is very difficult to imagine what structural intermediates would be like or how they could function. It is easier to envisage the replacement as a single-step event caused by a simple "switch" in gene expression (Ahlberg, 1992). A major aspect of coelacanth morphology, lobed fins instead of ray fins, may thus have arisen quite suddenly, rather than through gradual change. The mixed germ layer derivation of the mesenchyme found in the appendages of non-tetrapod vertebrates, in contrast to the single germ layer derivation of mesenchyme associated with the tetrapod limb, and lack of paleontological evidence for ray fin to tetrapod limb transition supports an interpretation favoring independent evolutionary origins for fish fins and tetrapod limbs. It is interesting to speculate that *Hox D* cluster genesis may have been the "single evolutionary step" allowing for the morphological elaboration that has resulted in tetrapod limbs and the neocranium.

REFERENCES

- Ahlberg, P. (1992) Coelacanth fins and evolution. *Nature* 358:459.
- Altschul, S., W. Gish, W. Miller, E. Myers, and D. Lipman. (1990)
Basic local alignment search tool. *J. Mol. Biol.* 215:
403-410.
- Akam, M. (1989) Hox and Hom: homologous gene clusters in insects and
vertebrates. *Cell* 57:347-349.
- Boncinelli, E., R. Somma, D. Acampora, M. Pannese, M. D' Esposito,
A. Faiella, and A. Simeone. (1988) Organization of human
homeobox genes. *Human Reproduction* 3:880-886.
- Boncinelli, E. A. Simeone, D. Acampora, and F. Mavillo. (1991). HOX
gene activation by retinoic acid. *Trends Genet.* 7:329-334.
- Clarke, L., and J. Carbon. (1976). A colony bank containing
synthetic Col El hybrid plasmids representative of the
entire *E. coli* genome. *Cell* 9:91.
- Devereux, J., P. Haeberli, and O. Smithies. (1984) A comprehensive
set of sequence analysis programs for the VAX. *Nucleic Acids
Research* 12:387-395.
- Dollé, P., A. Dierich, M. LeMeur, T. Schimmang, B. Schuhbaur, P.
Chambon, and D. Duboule. (1993) Disruption of the Hoxd-13
gene induces localized heterochrony leading to mice with
neotenic limbs. *Cell* 75:431-441.

- Duboule, D., and P. Dollé. (1989) The structural and functional organization of the murine *Hox* gene family resembles that of *Drosophila* homeotic genes. *EMBO J.* 8:1497-1505.
- Duncan, I. (1987) The bithorax complex. *Annu. Rev. Genet.* 21:285-319.
- Fjose, A., A. Molven, and H. Eiken. (1988). Molecular cloning and characterization of homeo-box-containing genes from Atlantic salmon. *Gene* 62:141-152.
- Fjose, A., H. Eiken, P. Njølstad, A. Molven, and I. Hordvik. (1990) Homeobox sequences of Atlantic salmon and zebrafish. *Aquaculture* 85:51-60.
- Fritz, A., K. Cho, C. Wright, B. Jegalian, and E. De Roberts. (1989) Duplicated homeobox genes in *Xenopus*. *Developmental Biology* 131:584-588.
- Gada, A., S. Gaunt, M. Azzawi, S. Shimeld, J. Pearce, and P. Sharpe. (1992). Sequence and embryonic expression of the murine *Hox-3.5* gene. *Development* 116:497-506.
- Gehring, W., M. Muller, M. Affolter, A. Percival-Smith, M. Billeter, Y. Qian, G. Otting, and K. Wuthrich. (1990) The structure of the homeodomain and its functional implications. *Trends In Genetics* 6:323-329.
- Gish, W. and D. States. (1993) Identification of protein coding regions by database similarity search. *Nature Genetics* 3:266-272.

- Graham, A., N. Papalopulu, J. Lorimer, J. McVey, E. Tuddenham, and R. Krumlauf. (1988) Characterization of a murine homeobox gene, *Hox-2.6*, related to the *Drosophila deformed* gene. *Genes Dev.* 2:1424-1438.
- Graham, A., N. Papalopulu, and R. Krumlauf. (1989) The murine and *Drosophila* homeobox gene complexes have common features of organization and expression. *Cell* 57:367-378.
- Hall, B. (1991) The evolution of connective and skeletal tissues. in *Developmental Patterning of the Vertebrate Limb*, J. Hinchliff, J. Hurle and D. Summerbell, eds. Vol. 205:303-311.
- Harvey, R., C. Tabin, and D. Melton. (1986) Embryonic expression and nuclear localization of *Xenopus* homeobox (*Xhox*) gene products. *The EMBO Journal* 5:1237-1244.
- Hillis, D. and M. Dixon. (1989). Vertebrate phylogeny: evidence from 28S ribosomal DNA sequences In *The Hierarchy of Life*. B. Fernholm, K. Bremer and H. Jörnvall eds. pp. 355-367. Elsevier Science Publishers, Amsterdam.
- Holland, P. (1992) Homeobox genes in vertebrate evolution. *Bioessays* 4(14):276-273.
- Hunt, P., M. Gulisano, M. Cook, M. Sham, A. Faiella, D. Wilkinson, E. Boncinelli, and R. Krumlauf. (1991) A distinct Hox code for the branchial region of the vertebrate head. *Nature* 353:861-864.

- Izpisúa-Belmonte, J., H. Falkenstein, P. Dollé, A. Renucci, and D. Duboule. (1991). Murine genes related to the *Drosophila Abd-B* homeotic genes are sequentially expressed during development of the posterior part of the body. *The EMBO Journal* 10:2279-2289.
- Jamieson, B. (1991). *Fish Evolution and Systematics: Evidence from Spermatozoa*. Cambridge: Cambridge University Press.
- Kan, Y., A. Dozy, R. Trecartin, and D. Todd (1977) Identification of a nondeletion defect in alpha-thalassemia. *N. Eng. J. Med.* 297:1081-1084.
- Kappen, C., K. Schughart, and F. Ruddle. (1989) Two steps in the evolution of Antennapedia-class vertebrate homeobox genes. *Proc. Natl. Acad. Sci. USA* 86:5459-5463.
- Kappen, C., K. Schughart, and F. Ruddle. (1993) Early evolutionary origin of major homeodomain sequence classes. *Genomics* 18:54-70.
- Kaufman, T., M. Seeger, and G. Olsen. (1990) Molecular and genetic organization of the Antennapedia gene complex of *Drosophila melanogaster*. In *Advances in Genetics: Genetic Regulatory Hierarchies in Development*, P. Wright, ed. Vol 27:309-362.
- Kessel, M., F. Schulze, M. Fibi, and P. Gruss. (1987). Primary structure and nuclear localization of a murine homeodomain protein. *Proc. Natl. Acad. Sci. USA* 84:5306-5310.

- Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences, *J. Mol. Evol.* 16:111-120.
- Kissenger, C., B. Liu, E. Martin-Blanco, T. Kornberg, and C. Pabo. (1990) Crystal structure of an engrailed homeodomain-DNA complex and 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell* 63:579-590.
- Krumlauf, R., P. Holland, J. McVey, and B. Hogan. (1986) Developmental and spatial patterns of expression of the mouse homeobox gene, 2.1. *Development* 99:603-617.
- Kuriowa, A., U. Kloter, P. Baumgartner, and W. Gehring. (1985) Cloning of the homeotic *Sex combs reduced* in *Drosophila* and *in situ* localization of its transcripts. *The EMBO Journal* 4:3757-3764.
- Laughon, A. (1991) DNA binding specificity of homeodomains. *Biochemistry* 30(48):11357-11367.
- Levine, E. and T. Hoey. (1988) Homeobox proteins as sequence-specific transcription factors. *Cell* 55:537-540.
- Levine, E. and N. Schechter. (1993) Homeobox genes are expressed in the retina and brain of adult goldfish. *Proc. Natl. Acad. Sci. USA* 90:2729-2733.
- Lewis, E. (1978) A gene complex controlling segmentation in *Drosophila*. *Nature* 276:565-570.

- Lobe, C., and P. Gruss. (1989) Mouse version of fly developmental control genes: legitimate or illegitimate relatives. *The New Biologist* 1(1):9-18.
- Lonai, P., E. Arman, H. Czosnek, F. Ruddle and C. Blare. (1987). New murine homeoboxes: Structure, chromosomal assignment and differential expression in adult erythropoiesis. *DNA* 6:409-418.
- Løvtrup, S. (1977). *The Phylogeny of Vertebrata*. London: Wiley & Sons.
- Lufkin, T., M. Mark, C. Hart, P. Dollé, M. LeMeur, and P. Chambon. (1992) Homeotic transformation of the occipital bones of the skull by ectopic expression of a homeobox gene. *Nature* 359:835-841.
- Mavillo, F., A. Simeone, A. Giampaolo, A. Faiella, V. Zappavigna, D. Acampora, G. Poiana, G. Russo, C. Peschle, and E. Boncinelli. (1986) Differential and stage-related expression in embryonic tissues of a new human homeobox gene. *Nature* 324:664-668.
- McElroy, D., P. Moran, E. Bermingham, and I. Kornfeld. (1991) REAP version 4.0
- McGinnis, W. and M. Kuziora. (1994) The molecular architects of body design. *Scientific American* 220(2):58-66.
- McGinnis, W. and R. Krumlauf. (1992) Homeobox genes and axial patterning. *Cell* 68:283-302.

- Molven, A., I. Hördvik, P. Njølstad, M. Van Ghelue, and A. Fjose. (1992). The zebrafish homeobox gene *hox(zf-114)*: Primary structure, expression pattern and evolutionary aspects. *Int. J. Dev. Biol.* 36:229-237.
- Molven, A., C. Wright, R. Bremiller, E. De Robertis, and C. Kimmel. (1990) Expression of a homeobox gene product in normal and mutant zebrafish embryos: evolution of the tetrapod body plan. *Development* 109:279-288.
- Morgan, B., J. Izpisua-Belmonte, D. Duboule, and C. Tabin. (1992) Targeted misexpression of Hox-4.6 in the avian limb bud causes apparent homeotic transformations. *Nature* 358:236-239.
- Murtha M., J. Leckman, and F. Ruddle. (1991) Detection of homeobox genes in development and evolution. *Proc. Natl. Acad. Sci. USA* 88:10711-10715.
- Njølstad, P., A. Molven, H. Eiken, and A. Fjose. (1988a). Structure and neural expression of a zebrafish homeobox sequence. *Gene* 73:33-46.
- Njølstad, P., A. Molven, and A. Fjose. (1988b). A zebrafish homologue of the murine Hox-2.1 gene. *FEBS Lett.* 230:25-30.
- Njølstad, P., A. Molven, I. Hördvik, J. Apold, and A. Fjose. (1988c). Primary structure, developmentally regulated expression and potential duplication of the zebrafish homeobox gene ZF-21. *Nucleic Acids Res.* 16:9097-9111.

- Njølstad, P., A. Molven, J. Apold, and A. Fjose. (1990) The Zebrafish homeobox gene *hox-2.2*: transcription unit, potential regulatory regions and in situ localization of transcripts. *The EMBO Journal* 9:515-524.
- Ohno, S. (1970) *Evolution by gene duplication*. Springer, Heidelberg. p. 89-109.
- Otting, G., Y. Qian, M. Billeter, M. Müller, M. Affolter, W. Gehring, and K. Wüthrich. (1990) Protein-DNA contacts in the structure of homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *The EMBO Journal* 9:3085-3092.
- Pendleton, J., B. Nagai, M. Murtha, and F. Ruddle. (1993) Expansion of the Hox gene family and the evolution of chordates. *Proc. Natl. Acad. Sci. USA* 90:6300-6304.
- Putz, M., R. Diederich, D. Cribbs, and T. Kaufman. (1988) The proboscipedia locus of the Antennapedia complex: a molecular and genetic analysis. *Genes Dev.* 2:901-920.
- Regulski, M., K. Harding, R. Kostriken, F. Karch, M. Levine, and W. McGinnis. (1985) Homeobox genes of the Antennapedia and Bithorax complexes of *Drosophila*. *Cell* 43:71-80.
- Regulski, M., N. McGinnis, R. Chadwick, and W. McGinnis. (1987) Developmental and molecular analysis of Deformed; a homeotic gene controlling *Drosophila* head formation. *The EMBO Journal* 6:767-777.

- Rohlf, F. (1992). NTSYS-pc version 1.70.
- Rubock, M., Z. Larin, M. Cook, N. Papalopulu, R. Krumlauf, and H. Lehrach. (1990) A yeast artificial chromosome containing the mouse homeobox cluster *Hox-2*. *Proc. Natl. Acad. Sci. USA* 87:4751-4755.
- Runstadler, J. and T. Kocher. (1991). A new Antennapedia-class gene from the zebrafish. *Nucleic Acids Res.* 19:5434.
- St. Johnston, D., and C. Nüsslein-Volhard. (1992) The origin of pattern and polarity in the *Drosophila* embryo. *Cell* 68:201-219.
- Sanchez-Herrero, E., I. Vernos, R. Marco, and G. Morata. (1985) Genetic organization of *Drosophila* Bithorax complex. *Nature* 313:108-113.
- Schubert, F., K. Nieselt-Struwe, and P. Gruss. (1993) The Antennapedia-type homeobox genes have evolved from three precursors separated early in metazoan evolution. *Proc. Natl. Acad. Sci. USA* 90:143-147.
- Schughart, K., C. Kappen, and F. Ruddle. (1988) Mammalian homeobox-containing genes: genome organization, structure, expression and evolution. *Br. J. Cancer Suppl.* 9(58):9-13.
- Schughart, K., C. Kappen, and F. Ruddle. (1989a) Duplication of large genomic regions during the evolution of vertebrate homeobox genes. *Proc. Natl. Acad. Sci. USA* 86:7067-7071.

- Schughart, K., D. Pravtcheva, M. Newman, L. Hunihan, Z. Jiang, and F. Ruddle. (1989b). Isolation and regional localization of the murine homeobox-containing gene Hox-3.3 to mouse chromosome region 15E. *Genomics* 5:76-83.
- Scott, M., J. Tamkun, and G. Hartzell. (1989) The structure and function of the homeodomain. *Biochemica et Biophysica Acta*. 989:25-48.
- Tabin C. (1991) Retinoids, homeoboxes, and growth factors: toward molecular models for limb development. *Cell* 66:199-217.
- Thorogood, P. (1991) The Development of the teleost fin and implications for our understanding of tetrapod limb evolution. In *Developmental Patterning of the Vertebrate Limb*. J. Hinchliffe, J. Hurle, and D. Summerbell, eds. Vol. 205:347-354
- Wedden, S., K. Pang, and G. Eichele. (1989) Expression pattern of homeobox-containing genes during chick embryogenesis. *Development*. 105:639-650.
- Wolgemuth, D., E. Engelmeyer, R. Duggal, E. Gizang-Ginsberg, G. Mutter, C. Ponzetto, C. Viviano, and A. Zakari. (1986). Isolation of a mouse cDNA coding for a developmentally regulated testis-specific transcript containing homeobox homology. *The EMBO Journal* 5:1229-1235.

Figure 1: Multiple sequence alignment of *Morone Hoxa-3*-like sequences. All *Morone* sequences are shown in reference to the *Drosophila Antennapedia homeobox* sequence. Dashes signify identity with the *Drosophila Antennapedia* sequence.

	1					60
<i>Droso.</i>	GAGAAGGAGT	TTCACTTCAA	TCGCTACTTG	ACCCGTCGGC	GAAGGATCGA	GATCGCCCAC
Psc212	-----	-----	-----	-----	----c-----	-----
Psc064	-----	-----	-----	-----	----c-----	-----
Psc098	-----	-----	-----	-----	-----	-----
Psc022	-----	-----	-----	-----	-----	-----
Psc051	-----	-----	-----	-----	-----	-----
Psc053	-----	-----	-----	-----c-----	-----	-----
Psc263	-----	-----	-----	-----	----c-----	-----
Psc112	-----	-----	-----	-----	----c-----	-----
Psc215	-----	-c-----	-----	-----	----c-----	-----
Psc228	-----	-c-----	-----	-----	----c-----	-----
Psc087	-----	-c-----	-----	-----	----c-----	-----
Psc211	-----	-c-----	-----	-----	----c-----	-----
Psc059	-----	-c-----	-----	-----	----c-----	-----
Psc073	-----	-c-----	-----	-----	----c-----	-----
Psc235	-----	-c-----	-----	-----	----c-----	-----

	61					114
<i>Droso.</i>	GCCCTGTGCC	TCACGGAGCG	CCAGATAAAG	ATTTGGTTCC	AGAATCGGCG	CATG
Psc212	-----	-----	-----	-----	----c-----	----
Psc064	-----	-----	-----	-----	----c-----	----
Psc098	-----	-----	-----	-----	----c-----	----
Psc022	-----	-----	-----	-----	----c-----	----
Psc051	-----	-----	-----	-----	----c-----	----
Psc053	-----	-----	-----	-----	----c-----	----
Psc263	-----	-----	-----	-----	-----	----
Psc112	-----	-----	-----	-----	-----	----
Psc215	-----	-----	-----	-----	-----	----
Psc228	-----	-----	-----	-----	-----	----
Psc087	-----	-----	-----	-----	-----	----
Psc211	-----	-----	-----	-----	----c-----	----
Psc059	-----	-----	-----	-----	----c-----	----
Psc073	-----	-----	-----	-----	----c-----	----
Psc235	-----	-----	-----	-----	-t--c-----	----

Figure 2: Multiple sequence alignment of *Morone Hoxa-4*-like sequence. All *Morone* sequences are shown in reference to the *Drosophila Antennapedia* homeobox sequence. Dashes signify identity with the *Drosophila Antennapedia* sequence.

1 60
Droso. GAGAAGGAGT TTCACTTCAA TCGCTACTTG ACCCGTCGGC GAAGGATCGA GATCGCCCAC
Psc229 -----a----- ----t----- c-----c-- ---a-g--ca --c-----t-- -----aa-t

51 114
Droso. GCCCTGTGCC TCACGGAGCG CCAGATAAAG ATTTGGTTCC AGAATCGGCG CATG
Psc229 a-g--t--t- -at-a----- g-----c-- --c----- --c--c-- ----

Figure 3: Multiple sequence alignment of *Morone Hoxa-6*-like sequences. All *Morone* sequences are shown in reference to the *Drosophila Antennapedia* homeobox sequence. Dashes signify identity with the *Drosophila Antennapedia* sequence.

	1					60
Droso.	GAGAAGGAGT	TTCACTTCAA	TCGCTACTTG	ACCCGTCGGC	GAAGGATCGA	GATCGCCCAC
Psc231	-----a----	----t-----	c--g---c-c	--g--g----	-gc-----	-----g---
Psc039	-----a----	----t-----	c--g---c-c	--g--g--tg	-gc-----	-----g---
	61					114
Droso.	GCCCTGTGCC	TCACGGAGCG	CCAGATAAAG	ATTTGGTTCC	AGAATCGGCG	CATG
Psc231	--g--c----	----c-----	g-----c--a	--c-----	-----c--	----
Psc039	--g--c----	----c-----	g-----c--	--c-----	-c--c--c--	----

Figure 4: Multiple sequence alignment of *Morone Hoxa-7*-like sequences. All *Morone* sequences are shown in reference to the *Drosophila Antennapedia* homeobox sequence. Dashes signify identity with the *Drosophila Antennapedia* sequence.

	1					60
<i>Droso.</i>	GAGAAGGAGT	TTCACTTCAA	TCGCTACTTG	ACCCGTCGGC	GAAGGATCGA	GATCGCCCAC
Psc047	-----a----	----t-----	c--g---c--	tg---g-c--	-cc-tg-g--	---g---a--
Psc056	-----a----	----t-----	c--g---c--	tg---g-c--	-cc-tg-g--	---g---a--
Psc001	-----a----	----t-----	c--g---c--	tg---g-c--	-cc-tg-g--	---g---a--
Psc222	-----a----	----t-----	c--g---c--	tg---g-c--	-cc-tg-g--	---g---a--
Psc232	-----a----	----t-----	c--g---c--	tg---g-c--	-cc-tg-g-g	---g---a--

	61					114
<i>Droso.</i>	GCCCTGTGCC	TCACGGAGCG	CCAGATAAAG	ATTTGGTTCC	AGAATCGGCG	CATG
Psc047	ctg--taa--	-gt-c--a--	-----c---	--c-----	-----c--c--	----
Psc056	ctg--taa--	-gt-c--a--	-----c---	--c-----	-----c--c--	----
Psc001	ctg--taa--	-gt-c--a--	-----c---	--c-----	-----c--c--	----
Psc222	ctg--taa--	-gt-c--a--	-----c---	--c-----	-----c--c--	----
Psc232	ctg--taa--	-gt-c--a--	-----c---	--c-----	-----c--c--	----

Figure 5: Multiple sequence alignment of *Morone Hoxb-3*-like sequences. All *Morone* sequences are shown in reference to the *Drosophila Antennapedia homeobox* sequence. Dashes signify identity with the *Drosophila Antennapedia* sequence.

	1					60
<i>Droso.</i>	GAGAAGGAGT	TTCACTTCAA	TCGCTACTTG	ACCCGTCGGC	GAAGGATCGA	GATCGCCCAC
Psc224	-----	-c-----	-----	-----	-----	-----
Psc104	-----	-c-----	-----	-----	-----	-----
Psc205	-----	-----	-----	-----	-----	-----
Psc012	-----	-----	-----	-----	-----	-----
Psc086	-----	-----	-----	-----	-----	-----
Psc045	-----	-----	-----	-----	-----	-----
Psc063	-----	-----	-----	-----	-----	-----
Psc226	-----	-----	-----	-----	-----	-----
Psc204	-----	-----	-----	-----	-----	-----
Psc101	-----	-----	-----	-----	-----	-----
Psc114	-----	-----	-----	-----	-----	-----
Psc052	-----	-----	-----	-----	-----	-----
Psc019	-----	-----	-----	-----	-----	-----
Psc099	-----	-----	-----	-----	-----	-----
Psc100	-----	-----	-----	-----	-----	-----
Psc023	-----	-----	-----	-----	-----	-----
Psc115	-----	g-----	-----	-----	-----	-----
Psc057	-----	-----	-----	-a-----	-----	-----
Psc203	-----	-----	-----	t-----t	-----	-----
Psc092	-----	-c-----	-----	-----	-----	-----

	61					114
<i>Droso.</i>	GCCCTGTGCC	TCACGGAGCG	CCAGATAAAG	ATTTGGTTCC	AGAATCGGCG	CATG
Psc224	-----	-----	-----	-----	-----c-----	-----
Psc104	-----	-----	-----	-----	-----	-----
Psc205	-----	-----	-----	-----	-----	-----
Psc012	-----	-----	-----	-----	-----	-----
Psc086	-----	-----	-----	-----	-----	-----
Psc045	-----	-----	-----	-----	-----	-----
Psc063	-----	-----	-----g-----	-----	-----	-----
Psc226	-----	-----	-----	-----	-----c-----	-----
Psc204	-----	-----	-----	-----	-----c-----	-----
Psc101	-----	-----	-----	-----	-----c-----	-----
Psc114	-----	-----	-----	-----	-----c-----	-----
Psc052	-----	-----	-----	-----	-----c-----	-----
Psc019	-----	-----	-----	-----	-----c-----	-----
Psc099	-----	-----	-----	-----	-----c-----	-----
Psc100	-----	-----	-----	-----	-----c-----	-----
Psc023	-----	-----	-----	-----	-----c-----	-----
Psc115	-----	-----	-----	-----	-----c-----	-----
Psc057	-----	-----t-----	-----	-----	-----c-----	-----
Psc203	-----	-----	-----	-----c-----	-----c-----	-----
Psc092	-----	-----	-----	-----	-----g-----	-----

Figure 6: Multiple sequence alignment of *Morone Hoxb-4*-like sequences. All *Morone* sequences are shown in reference to the *Drosophila Antennapedia homeobox* sequence. Dashes signify identity with the *Drosophila Antennapedia* sequence.

	1					60
<i>Droso.</i>	GAGAAGGAGT	TTCACTTCAA	TCGCTACTTG	ACCCGTCGGC	GAAGGATCGA	GATCGCCCCAC
Psc259	-----a	-----	-----	-----	-----	-----
Psc221	-----a	-----	-----	-----	-----	-----
Psc054	-----a	-----	-----	-----	-----	-----
Psc032	-----a	-----	-----	-----	-----	-----
Psc207	-----	-----	-----	-----	-----	-----
Psc225	-----	-----	-----	-----	-----	-----
Psc094	-----	-----	-----	-----	-----	-----
Psc024	-----	-----	-----	-----	-----	-----
Psc007	-----a	-----	-----	-----	-----	-----
Psc076	-----a	-g-----	-----	-----	-----	-----
Psc220	-----	-----	-----	-----	-----	-----
Psc037	-----	-----	-----	-----	-----	-----
Psc227	-----a	-----	-----	-----	-----	-----
Psc202	-----c	-----	-----	-----	-----	-----
Psc097	-----	-----	-----	-----	-----	-----
Psc213	-----a	-c-----	-----	-----	-----	-----
Psc021	-----	-c-----	-----	-----	-----	-----
Psc069	-----a	-----	-----	-----	-----	-----
Psc089	-----a	-----	-----	-----	-----	-----
Psc068	-----a	-----	-----	-----	-----	-----
Psc028	-----a	-----	-----	-----	-----	-----
Psc035	-----g	-----	-----	-----	-----	-----
Psc026	-----a	-----	-----	-t-----	-----	-----
Psc210	t-----a	-c-----	-----	-----	-----	-----
Psc090	-----t	-----g--	-----	-----	-----	-----
Psc040	-----g	-----	-----	-----c-t--g	-----	-----
Psc208	-----	-----	-----	-----	-----	-----

	61					114
<i>Droso.</i>	GCCCTGTGCC	TCACGGAGCG	CCAGATAAAG	ATTTGGTTCC	AGAATCGGCG	CATG
Psc259	-----c	-----	-----	-----	-----	-----
Psc221	-----c	-----	-----	-----	-----	-----
Psc054	-----c	-----	-----	-----	-----	-----
Psc032	-----c	-----	-----	-----	-----	-----
Psc207	-----c	-----	-----	-----	-----	-----
Psc225	-----c	-----	-----	-----	-----	-----
Psc094	-----c	-----	-----	-----	-----	-----
Psc024	-----c	-----	-----	-----	-----	-----
Psc007	-----c	-----g-	-----	-----	-----	-----
Psc076	-----c	-----	-----	-----	-----	-----
Psc220	-----	-----	-----	-----	-----	-----
Psc037	-----	-----	-----	-----	-----	-----
Psc227	-----	-----	-----	-----	-----	-----
Psc202	-----	-----	-----	-----	-----	-----
Psc097	-----g	-----	-----	-----	-----	-----
Psc213	-----	-----	-----	-----	-----	-----
Psc021	-----	-----	-----	-----	-----	-----
Psc069	-----	-----	-----	-----	-----	-----
Psc089	-----	-----	-----	-----	-----	-----

Figure 7: Multiple sequence alignment of *Morone Hoxb-5*-like sequences. All *Morone* sequences are shown in reference to the *Drosophila Antennapedia homeobox* sequence. Dashes signify identity with the *Drosophila Antennapedia* sequence.

	1					60
<i>Droso.</i>	GAGAAGGAGT	TTCACTTCAA	TCGCTACTTG	ACCCGTCGGC	GAAGGATCGA	GATCGCCCAC
Psc262	-----c-----	-----	-----	-----	-----	-----
Psc234	-----c-----	-----	-----	-----	-----	-----
Psc075	-----c-----	-----	-----	-----	-----	-----
Psc003	-----c-----	-----	-----	-----	-----	-----
Psc217	-----c-----	-----	-----	-----	-----	-----
Psc113	-----c-----	-----	-----	-----	-----	-----
Psc264	-----	-----	-----	-----	-----	-----
Psc230	-----	-----	-----	-----	-----	-----
Psc206	-----	-----	-----	-----	-----	-----
Psc020	-----	-----	-----	-----	-----	-----
Psc116	-----	-----	-----	-----	-----	-----
Psc096	-----	-----	-----	-----	-----	-----
Psc014	-----	-----	-----	-----	-----	-----
Psc080	-----	-----	-----	-----	-----	-----
Psc031	-----	-----	-----	-----	-----	-----
Psc011	-----	-----	-----	-----	-----	-----
Psc081	-----	-----	-----	-----	-----	-----
Psc029	-----	-----	-----	-----	-----	-----
Psc066	-----	-----	-----	-----	-----	-----
Psc038	-----	-----	-----	-----	-----	-----
Psc041	-----	-----	-----	-----c-----	-----	-----
Psc219	-----	-----	-----	-----c-----	-----c-----	-----
Psc044	-----	-----	-----	-----	-----	-----
Psc042	-----c-----	-----	-----	-----	-----	-----
Psc071	-----	-----	-----g-----	-----	-----	-----

	61					114
<i>Droso.</i>	GCCCTGTGCC	TCACGGAGCG	CCAGATAAAG	ATTTGGTTCC	AGAATCGGCG	CATG
Psc262	-----	-----	-----	-----	-----c-----	-----
Psc234	-----	-----	-----	-----	-----c-----	-----
Psc075	-----	-----	-----	-----	-----c-----	-----
Psc003	-----	-----	-----	-----	-----c-----	-----
Psc217	-----	-----	-----	-----	-----	-----
Psc113	-----	-----	-----	-----	-----	-----
Psc264	-----	-----	-----	-----	-----c-----	-----
Psc230	-----	-----	-----	-----	-----c-----	-----
Psc206	-----	-----	-----	-----	-----c-----	-----
Psc020	-----	-----	-----	-----	-----c-----	-----
Psc116	-----	-----	-----	-----	-----c-----	-----
Psc096	-----	-----	-----	-----	-----c-----	-----
Psc014	-----	-----	-----	-----	-----c-----	-----
Psc080	-----	-----	-----	-----	-----c-----	-----
Psc031	-----	-----	-----	-----	-----c-----	-----
Psc011	-----	-----	-----	-----	-----	-----
Psc081	-----	-----	-----	-----	-----	-----
Psc029	-----	-----	-----	-----	-----	-----
Psc066	-----	-----c-----	-----	-----	-----c-----	-----
Psc038	-----	-----	-----	-----	-----tc-----	-----
Psc041	-----	-----	-----	-----	-----c-----	-----
Psc219	-----	-----	-----	-----	-----	-----

Figure 7 continued: Multiple sequence alignment of *Morone* *Hoxb*-5-like sequences. All *Morone* sequences are shown in reference to the *Drosophila* Antennapedia homeobox sequence. Dashes signify identity with the *Drosophila* Antennapedia sequence.

	61					114
<i>Droso.</i>	GCCCTGTGCC	TCACGGAGCG	CCAGATAAAG	ATTTGGTTCC	AGAATCGGCG	CATG
Psc044	-----	-----	-----	-----	-----	----
Psc042	---a-----	-----	-----	-----	-----	----
Psc071	-----	-----	-----g	-----	----c-----	----

Figure 8: Multiple sequence alignment of *Morone Hoxc-4*-like sequence. All *Morone* sequences are shown in reference to the *Drosophila Antennapedia* homeobox sequence. Dashes signify identity with the *Drosophila Antennapedia* sequence.

Figure 9. Unrooted evolutionary tree for selected striped bass *Hox* sequences. The distance matrix resulted from the calculation of pairwise distance of the nucleotide sequences. The unrooted evolutionary tree was constructed by applying the NJOIN and NTSYS programs to the distance matrix. *Hox* gene designations are provided at the end of the respective branch. Scaled numbers represent relative distance values.

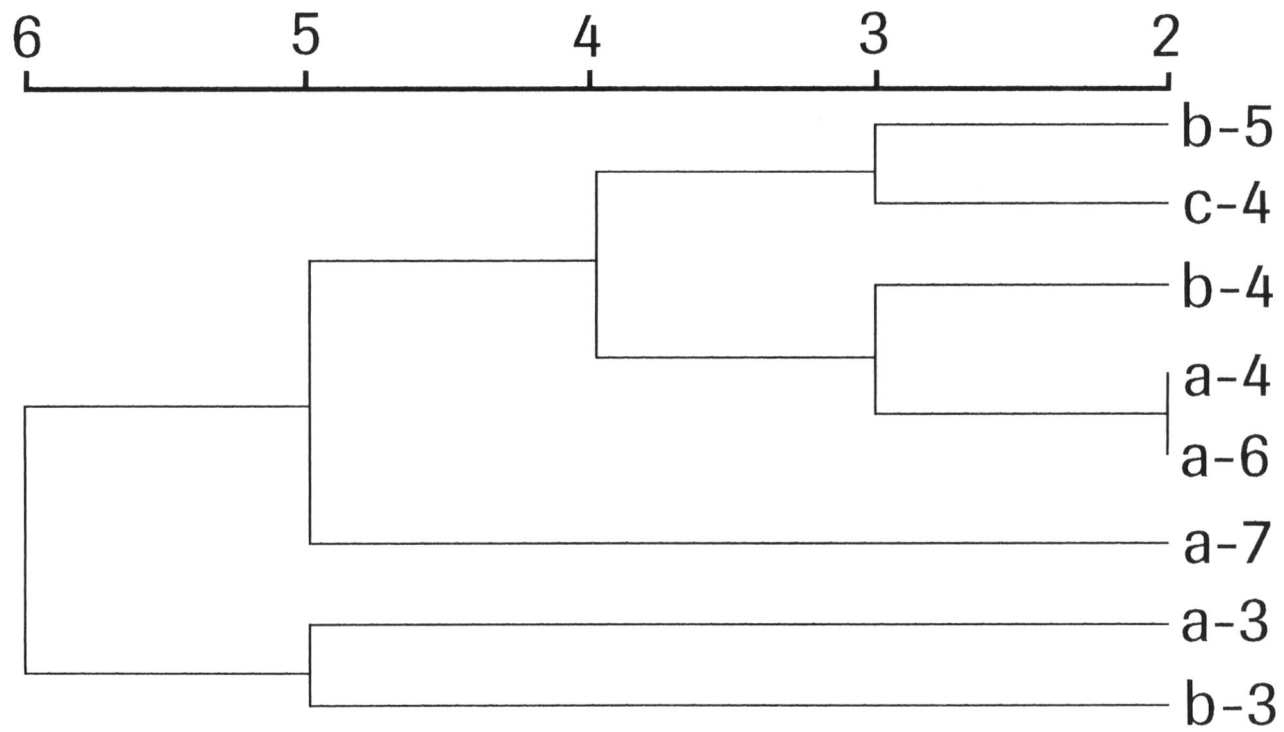


Figure 10. Unrooted evolutionary tree for selected zebrafish *Hox* sequences. The distance matrix resulted from the calculation of pairwise distance of the nucleotide sequences. The unrooted evolutionary tree was constructed by applying the NJOIN and NTSYS programs to the distance matrix. *Hox* gene designations are provided at the end of the respective branch. Scaled numbers represent relative distance values.

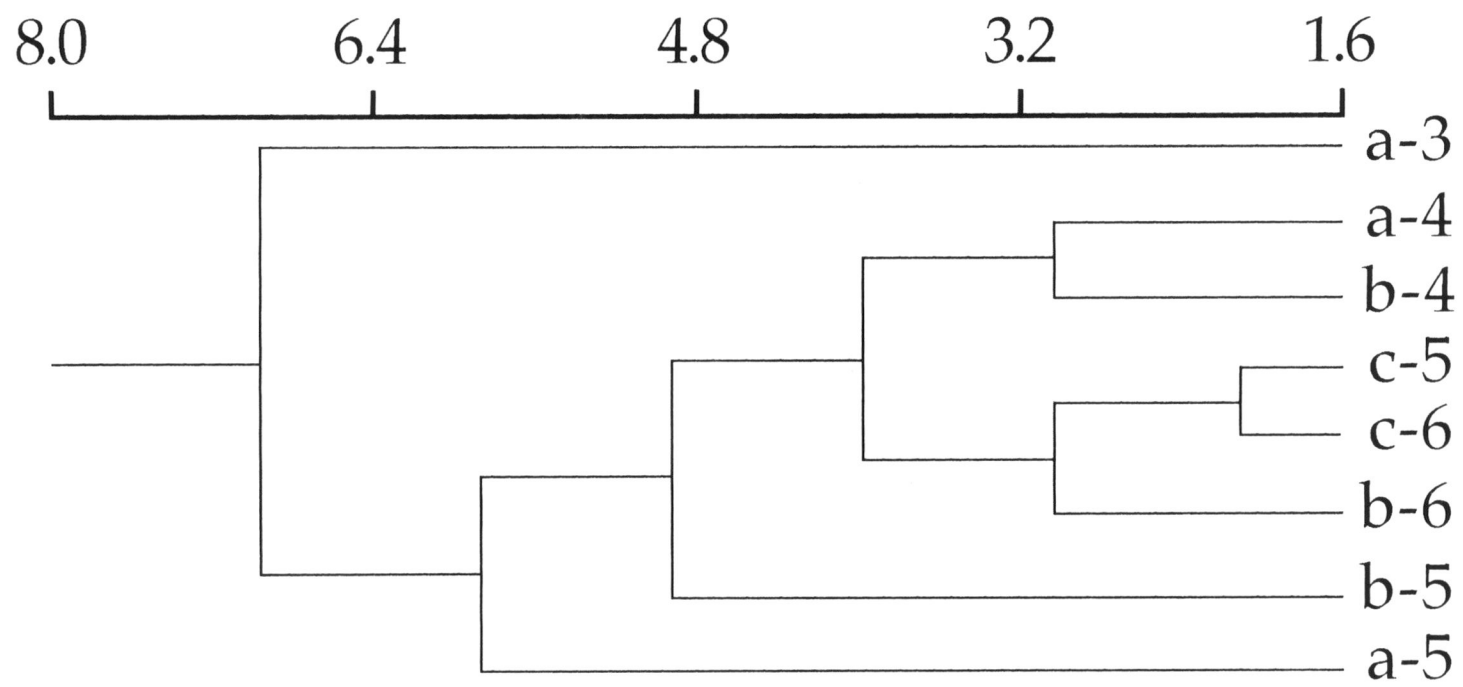


Figure 11. Unrooted evolutionary tree for selected mouse *Hox* sequences. The distance matrix resulted from the calculation of pairwise distance of the nucleotide sequences. The unrooted evolutionary tree was constructed by applying the NJOIN and NTSYS programs to the distance matrix. *Hox* gene designations are provided at the end of the respective branch. Scaled numbers represent relative distance values.

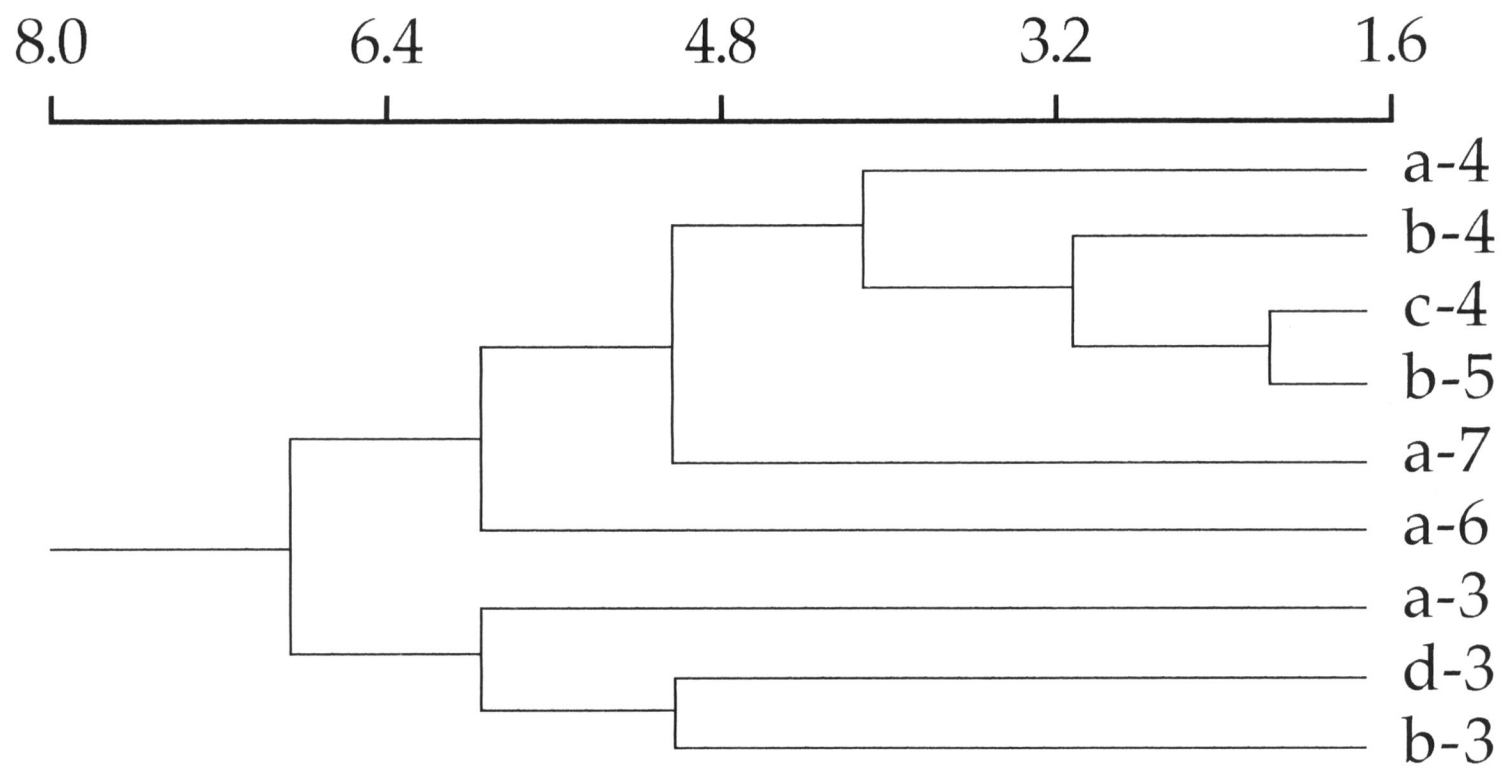


Table 1. Comparison and homology assignments of *Hox* PCR clones from striped bass. All sequences are relative to the *Drosophila* Antennapedia homeobox sequence found at the top of each column with hyphens indicating identical amino acids. ^a The mouse *Hox* genes and cognate groups are indicated in the left column and the *Drosophila* homologs in the right column. ^b The number of individual striped bass clones recovered is indicated to the right of each sequence. ^c ZF-114 has been assigned as an a-3 paralog in this chart (Molven et al., 1992). ^d Based on our analysis ZF-13 has been assigned as a paralog to group a-4, however, it is unclear whether ZF-13 is a paralog of a-4 or b-4 (Njølstad et al., 1988a).

HOX GENE	MOUSE ^b	STRIPED BASS	LAMPREY	ZEBRAFISH
	YLTRRRRIEIAHALCLTERQ	YLTRRRRIEIAHALCLTERQ	YLTRRRRIEIAHALCLTERQ	YLTRRRRIEIAHALCLTERQ
a-3	--M-P--V-M-NL-N-----	-IS-P--V-L-LT-S-----H 15		-C-N--L-M-E--K--D---
b-3	--C-P--V-M-NL-N-S---	--C-P--V-M-NL-N-S--- 5		
d-3	--C-P--V-M-NL-N-----			
a-4	-----T---S---	-----NT---S--- 1	--A---V---S---S---	-----V---T---S---
b-4	-----V-----S---	--S---V---T---N--- 20	-----S---S---	-----V---TM--S---
c-4	-----S---S---	-----S--- 1	-----L---S-A-S---	
d-4	-----T---P---			
a-5	-----S---		-----V-V-N---S---	-----R---S---
b-5	-----S---	-----S--- 29	-----V-N---S---	-----S---
c-5	-----NN---N---			-----NN---N---
a-6	-----N-----	-----N----- 27	-----V---S-----	
b-6	-----		-----V---S-----	-----
c-6	-----N-----		-----V---S-----	-----N-----
a-7	-----	----- 2	-----	
b-7	-----T-----		-----G-----	

Table 1 continued. Comparison and homology assignments of *Hox* PCR clones from striped bass. All sequences are relative to the *Drosophila* Antennapedia homeobox sequence found at the top of each column with hyphens indicating identical amino acids.

HOX GENE	MOUSE	GOLDFISH	ATLANTIC SALMON
	YLTRRRRIEIAHALCLTERQ	YLTRRRRIEIAHALCLTERQ	YLTRRRRIEIAHALCLTERQ
a-3	--M-P--V-M-NL-N-----	-IA-P--V-L-LT-S----H	
b-3	--C-P--V-M-NL-N-S----	--C-P--V-M-NL-N-S----	
d-3	--C-P--V-M-NL-N-----		
a-4	-----T---S---	-----T---S---	
b-4	-----V-----S---		
c-4	-----S---S---		
d-4	-----T---P---		
a-5	-----S---		
b-5	-----S---	-----S---	-----M-----S---
c-5	-----NN---N---		
a-6	-----N-----		
b-6	-----	-----S-----	
c-6	-----N-----		
a-7	-----		-----V-----V-----
b-7	-----T-----		

Table 2. Pair-wise comparison of *Hox* and *Hox*-like a-4 and b-5 gene nucleotide sequences. *Hox* and *Hoxa*-4-like nucleotide sequence identities are reported above the dashes and b-5 nucleotide sequence identities are reported below the dashes.

SPECIES	MOUSE	STRIPED BASS	ZEBRAFISH	GOLDFISH	LAMPREY
MOUSE	-	0.23	0.32	0.32	0.20
STRIPED BASS	0.21	-	0.27	0.21	0.32
ZEBRAFISH	0.26	0.19	-	0.34	0.35
GOLDFISH	0.31	0.18	0.16	-	0.34
LAMPREY	0.22	0.34	0.31	0.37	-