

HOX genes in the sepiolid squid *Euprymna scolopes*: Implications for the evolution of complex body plans

Patrick Callaerts^{*†}, Patricia N. Lee^{**‡}, Britta Hartmann^{*§}, Claudia Farfan[‡], Darrett W. Y. Choy[‡], Kazuho Ikeo[¶], Karl-Friedrich Fischbach[§], Walter J. Gehring^{||}, and H. Gert de Couet^{*§***}

^{*}Department of Biology and Biochemistry, University of Houston, Houston, TX 77204-5513; [§]Institut fuer Biologie III, Schaeenzlestrasse 1, 76104 Freiburg, Germany; ^{||}Department of Cell Biology, Biozentrum, Klingelbergstrasse 70, CH-4056 Basel, Switzerland; [‡]Department of Zoology, University of Hawaii at Manoa, 2538 McCarthy Mall, Honolulu, HI 96822; and [¶]National Institute of Genetics, Center for Information Biology and DNA Data Base in Japan, Yata 1111, Mishima, Japan

Contributed by Walter J. Gehring, December 19, 2001

Molluscs display a rich diversity of body plans ranging from the wormlike appearance of aplacophorans to the complex body plan of the cephalopods with highly developed sensory organs, a complex central nervous system, and cognitive abilities unrivaled among the invertebrates. The aim of the current study is to define molecular parameters relevant to the developmental evolution of cephalopods by using the sepiolid squid *Euprymna scolopes* as a model system. Using PCR-based approaches, we identified one anterior, one paralog group 3, five central, and two posterior group *Hox* genes. The deduced homeodomain sequences of the *E. scolopes* *Hox* cluster genes are most similar to known annelid, brachiopod, and nemertean *Hox* gene homeodomain sequences. Our results are consistent with the presence of a single *Hox* gene cluster in cephalopods. Our data also corroborate the proposed existence of a differentiated *Hox* gene cluster in the last common ancestor of Bilaterians. Furthermore, our phylogenetic analysis and in particular the identification of *Post-1* and *Post-2* homologs support the Lophotrochozoan clade.

The molluscs constitute one of the most successful animal phyla. The earliest fossils appear in the Tommotian and Atadabanian stages of the early Cambrian about 530 million years ago (1, 2). Between 50,000 and 100,000 extant mollusc species are divided into seven classes: Aplacophora, Polyplacophora, Monoplacophora, Scaphopoda, Bivalvia, Gastropoda, and Cephalopoda (3). Species of the various classes display a rich diversity of body plans ranging from a wormlike appearance in the aplacophorans to the complex and specialized anatomy found in the cephalopods. Despite the very diverse mollusc adult morphologies, the following synapomorphies have been proposed for the average mollusc: a ventral locomotory foot, a dorsal shell secreted by the mantle that also defines a cavity that houses the gills, a chitinous radula for feeding, a hemocoel, a dorsal heart, and a coelom mostly restricted to the pericardial and gonadal domains (4, 5). However, the very different forms observed in bivalves, gastropods, and cephalopods suggest that the basic molluscan body plan can be modified to the extent that all but the most fundamental characteristics are obscured. Contrary to the situation with the arthropods, nematodes, echinoderms, and chordates, where the mechanisms that underlie embryonic development have been well studied (6–10), surprisingly little is known about developmental mechanisms in molluscs.

Key determinants for anteroposterior body axis formation are a subset of homeobox-containing genes, the *Hox* genes. Homeobox genes encode transcription factors containing a highly conserved DNA-binding motif, the homeodomain (11). *Hox* genes initially were discovered in *Drosophila* but have since been identified in all animal phyla where they were studied (reviewed in refs. 11–14). Clustered organization of *Hox* genes has been found in a number of chordates including human, mouse (15), zebrafish (16), and amphioxus (17). Nonchordate *Hox* clusters have been described for *Drosophila* (18), ribbonworm (19), sea urchin (20), and *Caenorhabditis elegans* (21–23). Phylogenetic

analysis of the available *Hox* gene information indicates that the last common ancestor of protostomes and deuterostomes had a cluster of at least six or seven *Hox* genes (14, 24). It appears contradictory that, despite their remarkable conservation, *Hox* genes could be responsible for the dramatic differences in body plans within and between phyla. Several hypotheses have been put forward to resolve this question. It has been proposed that the divergence between classes or phyla is accompanied by molecular changes in the *Hox* gene cluster or in the deployment of individual members. These changes may encompass duplications of the *Hox* gene cluster itself, selective gain or loss of *Hox* genes, changes in *Hox* gene expression levels, and changes in regulatory interactions between *Hox* proteins and their targets. Comparative genome analyses support the hypothesis that seemingly drastic developmental differences between phyla are brought about by differential deployment of functionally and structurally conserved *Hox* gene products in development rather than by the *de novo* invention of unique proteins (6, 10, 24). The molecular mechanisms contributing to this evolutionary “tinkering” (25, 26) are thought to have been established during the pre-Cambrian period, ultimately contributing to the “Cambrian explosion” of morphological novelties (1, 24).

We have analyzed the *Hox* genes of the Hawaiian sepiolid squid *Euprymna scolopes* (Cephalopoda, Dibranchiata, Sepioloidea), of which adults and embryos can be reared readily under laboratory conditions, providing the basis for systematic developmental studies (27) (Fig. 1). We have identified one anterior group gene, one ortholog of paralog group 3 (PG3), five central group genes, and two posterior group genes. Our data support the idea of a single putative *Hox* cluster in *E. scolopes* encoding a minimum of nine *Hox* paralogs, and support earlier hypotheses about the ancient origin of the *Hox* gene cluster (14, 28). The presence of the diverged posterior group genes provides additional support for the Lophotrochozoan assemblage (29).

Materials and Methods

DNA and RNA Extraction. Genomic DNA was prepared from brains and arms of two medium-sized adult *E. scolopes*. Only this tissue was used to prevent contamination of the preparation with DNA from ingested prey. Genomic DNA was isolated by equilibrium density centrifugation in cesium chloride gradients (30).

Total RNA was prepared from 25 embryos each from Arnold stages 19–20 and 21–23 (6 and 8 days old, respectively) (27) by

Abbreviation: PG3, paralog group 3.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF127334–AF127342, AF325504, and AY052753–AY052761).

[†]P.C. and P.N.L. contributed equally to this work.

^{**}To whom reprint requests should be addressed at: Department of Zoology, University of Hawaii, 2538 McCarthy Mall, Honolulu, HI 96822. E-mail: couet@hawaii.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

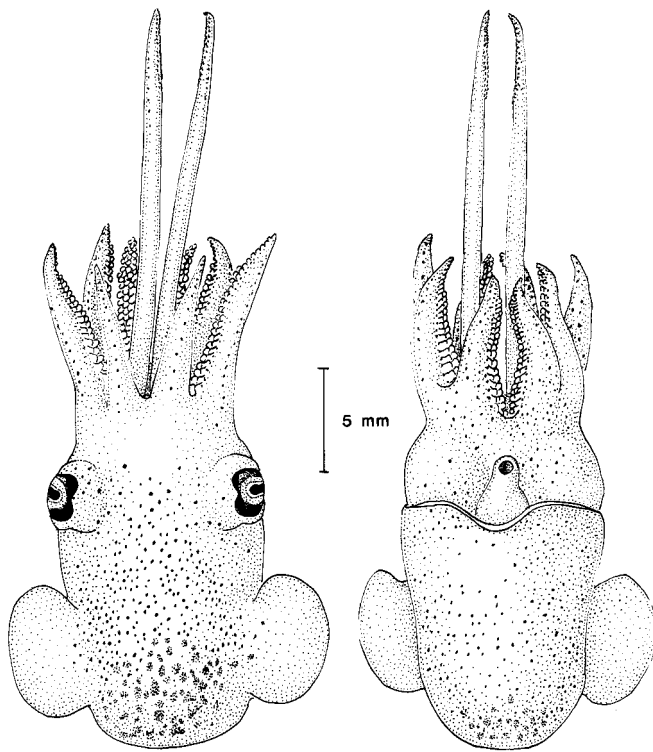


Fig. 1. Dorsal and ventral views of *E. scolopes*, subadult specimen.

using TriReagent following the manufacturer's protocols (Molecular Research Center, Cincinnati). Poly(A)⁺ mRNA was isolated by using an Oligotex Direct mRNA Isolation Kit (Qiagen, Valencia, CA) following the manufacturer's protocols.

Cloning Homeobox Fragments and Hox Genes. PCR primers. We used the previously described primer pairs A/B (31) and E/F (32). For the amplification of the *Post-1* and *Post-2* homeobox fragments, specific primer pairs were designed that match the consensus sequence of the amino terminus of the homeodomain of Lophotrochozoan *Post* genes (14) and have the following sequence: *Post-1* (5') AA(A/G)TA(T/C)CA(A/G)AT(T/C/A)GCNGA(A/G)(C/T)TNGA(A/G)(C/A)GNGA(A/G)TA(T/C)G corresponding to amino acid sequence KYQIAELEREY; *Post-1* (3') (T/C)TT(T/C)TT(T/C)TC(T/C)-TTCATNC(G/T)NC(G/T)(A/G)TT(T/C)TG(A/G)AA-CCA corresponding to amino acid sequence WFQNR-RMKEKK; *Post-2* (5') (C/A)GNTA(T/C)CA(A/G)ACNATG-GTN(C/T)TNGA(A/G)AA(T/C)GA(A/G)TT(T/C) corresponding to amino acid sequence RYQTMVLENEF; *Post-2* (3') (T/C)TT(T/C)TTNC(G/T)(T/C)TTCATNC(G/T)NC(G/T)(A/G)TT(T/C)TG(A/G)AACCA corresponding to amino acid sequence WFQNRMRKRRK.

PCR conditions. Amplification with the A/B primer set in a reaction volume of 100 μ l used 0.45 μ g of genomic DNA, 0.5 μ M primers, 1.5 mM Mg²⁺, 10 mM dNTPs, and 2.5 units *Taq* polymerase. Cycling conditions were one cycle (94°C for 3 min), 35 cycles (94°C for 1 min; 37°C for 1 min; 72°C for 30 sec), and one cycle (72°C for 5 min). Amplification with the primer E/F primer pair in a reaction volume of 100 μ l used 0.45 μ g of genomic DNA, 0.5 μ M primers, 1.5 or 5 mM Mg²⁺, 10 mM dNTPs, and 2.5 units *Taq* polymerase. Cycling conditions were one cycle (94°C for 3 min), 35 cycles (94°C for 1 min; 37°C for 1 min, ramping time 0.5°C per sec, or 40°C for 1 min with maximum ramping; 72°C for 30 sec), and one cycle (72°C for 5 min). Amplification of a *Post-1* homeobox in a reaction volume

of 50 μ l used 0.38 or 0.57 μ g genomic DNA, 5 mM Mg²⁺, 0.2 mM dNTPs, 0.2 μ M primers, and 2.5 units *Taq* polymerase. Cycling conditions were one cycle (94°C for 3 min), 30 cycles (94°C for 1 min; 57°C or 60°C for 1 min; 72°C for 30 sec), and one cycle (72°C for 5 min). Amplification of a *Post-2* homeobox in a reaction volume of 50 μ l used 0.1 or 0.38 μ g genomic DNA, 5 mM Mg²⁺, 0.2 mM dNTPs, 0.2 μ M primers, and 2.5 units *Taq* polymerase. Cycling conditions were one cycle (94°C for 3 min), 30 cycles (94°C for 1 min; 55°C or 56°C for 1 min; 72°C for 30 sec), and one cycle (72°C for 5 min).

Cloning and sequencing of PCR products. Genomic PCR products were gel-purified by using either the Mermaid or GeneClean procedure (Bio 101) and cloned into the pCR2.1 vector (Invitrogen). PCR fragments obtained by rapid amplification of cDNA ends were either gel-purified or directly ligated into the pCRII-TOPO vector (Invitrogen). Cloned PCR products were sequenced by dideoxy chain termination (33) with SEQUENASE 2.0 (United States Biochemical) or by using an automated sequencer (Applied Biosystems). Forty-six clones derived from the Hox A/B primer set were analyzed, and another 75 were selected from the Hox E/F experiment. Four *Post-1* clones, and 15 clones obtained with the *Post-2* gene-specific primer set, were sequenced.

Rapid amplification of cDNA ends-PCR. PCR-mediated rapid amplification of cDNA ends (Marathon cDNA Amplification Kit, CLONTECH) was used to clone the 5' and 3' sequences of the *E. scolopes* Hox cDNAs starting from *E. scolopes* embryonic poly(A)⁺ mRNA following the manufacturer's protocols. Gene-specific nested primers were synthesized for each *Hox* paralog (primer sequences are available on request) and used in conjunction with the nested adaptor-specific primers provided in the Marathon cDNA amplification kit. Touch-down PCR cycling parameters were: one cycle (94°C for 1 min), five cycles (94°C for 30 sec; 72°C for 4 min), five cycles (94°C for 30 sec; 70°C for 2.5 min), and 20 cycles (94°C for 20 sec; 68°C for 2.5 min). A second 50- μ l reaction with the nested primers and 1 μ l of first-round PCR products as the template was performed by using the same cycling parameters as described above.

DNA and Protein Sequence Analysis. Nucleotide sequences and their conceptual translation products were entered into the BLASTX and BLASTP programs of the National Center for Biotechnology Information to identify putative homologs and paralog groups and to determine the similarity between cognate genes (34). The selection of posterior group gene sequences for amino acid substitution calculations and phylogenetic tree construction reflected available representatives from the major phyla and similarity scores obtained from BLASTX and BLASTP analyses. We further concentrated on phyla for which comprehensive data sets are available to increase the confidence in paralogy assignments, and the sequences of the entire homeodomains were applied to the analysis. The sequences were aligned by using CLUSTAL X 1.8.1 (35–37) with manual modifications performed to optimize matches. Alignments were used to calculate amino acid substitution numbers between sequence pairs with the correction of multiple substitutions by using Kimura's method (38, 39). For distance-based phylogenetic analyses and tree construction, programs from the ODEN package were applied. Reconstruction of phylogenetic trees by the maximum-likelihood method was performed by using PUZZLE 4.0.2 (40, 41), applying the BLOSUM 62 substitution model (42) and 1,000 puzzling steps. Final editing of the resulting trees was performed with the program TREEVIEW (43).

Nomenclature. The *E. scolopes* Hox genes were named in accordance with previous suggestions and in keeping with the current literature (12, 14). The gene name is prefixed with the letters *Esc*

denoting the species initials for *E. scolopes*, followed by the paralogy assignment.

Results

Using PCR amplification with degenerate homeobox-specific primers we were able to recover 11 unique homeobox gene fragments from genomic DNA of *E. scolopes* in a total of 116 informative sequences. Two of these sequence classes represented the non-*Hox* cluster genes *caudal* and *gbx* (results not shown; GenBank accession numbers AF127341–2). A presumed *labial* (PG1) ortholog was recovered with the HOX E/F primer set only, whereas PG3 and central homeoboxes were identified at various frequencies with both primer sets (data not shown). The *E. scolopes* *Post-1* and *Post-2* homeoboxes were recovered with the *Post*-specific primers. Subsequently, complete homeobox sequences and flanking sequences were obtained for all *Hox* genes except for the *Deformed* gene.

Anterior Group Orthologs. The putative homeodomain sequence of the *E. scolopes* *labial* ortholog (*Esc-lab*) is very similar to that of the nemertean *Lineus* (19) and polychaetes [*Ctenodrilus* *CTs-lab* (44); *Chaetopterus* *CHv-Hb3* (45); *Nereis virens* *Nvi-lab* (14)] and brachiopods (*Lingula anatina*, ref. 14). We were unable to identify a sequence with significant homology to paralog group 2 (*proboscipedia* orthologs).

PG3 Group Orthologs. *Esc-Hox3* is a member of PG3 (Table 1 and Fig. 2), which is identified by the *Drosophila* gene *zerknüllt* (*zen*). The closest relatives of the *E. scolopes* sequence are the *Hox3* paralogs of representative species of the Lophotrochozoa clade. These include the mollusc *Patella vulgata* (*Pvu-Hox3*; ref. 14), the nemertean *Lineus* (*LsHox-3*; ref. 19), the polychaetes *Chaetopterus* (*ChvHb-5*; ref. 45), *Ctenodrilus* (*CTs-Hox3*; ref. 44), *Nereis* (*Nvi-Hox3*; ref. 14), and the brachiopod *Lingula* (*Lan-Hox3*; ref. 14).

Central Group Orthologs. We recovered five distinct sequences that are representative for *Antennapedia*-like homeoboxes of the central group. The central group orthologs were assigned to paralogy groups based on conserved residues within the homeodomain and on signature motifs for particular paralogs found in regions flanking the homeodomain (14).

Esc-Dfd displayed the highest sequence similarity with paralog group 4 genes from both Ecdysozoan and Lophotrochozoan phyla. The *Drosophila* *Deformed* gene is the namesake for this paralog group. Although we were unable to isolate 3' flanking sequences of this homeobox, which may contain *Dfd*-specific signature motifs (see Fig. 2) the recovered sequence still allowed a confident paralogy assignment.

Esc-Scr represents a paralog group 5 member, identified by the *Drosophila* *Sex combs reduced* gene. It is most similar to *Hru-Hox5* from the archaegastropod *Haliotis* (46) and additional Lophotrochozoan *Hox* genes, including the brachiopod *Lingula* and the polychaete *Nereis* (14). It also shows significant sequence similarity to *Amphi-Hox5* (17), and the leech *Hox* genes *Lox20* and *Lox6* (47).

Esc-Antp displayed significant similarity with *Antennapedia* orthologs, including *Priapulius Pca-HB2*, the nemertean *Ls-Hox7* homeodomain, and the brachiopod *Lan-HB1* sequence. The inferred sequence of *Esc-Hox7* is identical to *Drosophila* *Antennapedia*.

Esc-Lox5 appears to be a member of the *Antennapedia* group of *Hox* genes, most similar to paralog group 6 homeobox genes from other Lophotrochozoa. Members of this paralog group share conserved amino acids Q1, T4, and G39 of the homeodomain (14), features that separate them from the Ecdysozoan and deuterostome paralog group 6 members. These features were first defined in leech homeobox genes and are therefore fre-

Table 1. Species names and taxonomic affinities used in sequence analyses

| Genus/species | Phylum/class | Abbreviation | References |
|--------------------------------|-----------------------------|--------------|--------------|
| <i>Euprymna scolopes</i> | Mollusca/Cephalopoda | Esc | This article |
| <i>Haliotis rufescens</i> | Mollusca/Gastropoda | Hru | 46 |
| <i>Chaetopterus variegatus</i> | Annelida/Polychaeta | Chv | 45 |
| <i>Nereis virens</i> | Annelida/Polychaeta | Nvi | 14 |
| <i>Helobdella robusta</i> | Annelida/Hirudinea | Hro | 14 |
| <i>Helobdella triserialis</i> | Annelida/Hirudinea | Htr | 14 |
| <i>Priapulius caudatus</i> | Priapulida | Pca | 14 |
| <i>Lineus sanguineus</i> | Nemertea | Lsa | 19 |
| <i>Lingula anatina</i> | Brachiopoda | Lan | 14 |
| <i>Dugesia japonica</i> | Platyhelminthes/Turbellaria | Dja | 48 |
| <i>Drosophila melanogaster</i> | Arthropoda/Hexapoda | Dme | 18 |
| <i>Tripneustes gratilla</i> | Echinodermata | Tgr | 57 |
| <i>Ciona intestinalis</i> | Chordata | Cin | 17 |
| <i>Branchiostoma floridae</i> | Chordata | Bfl | 17 |

quently grouped as *Lox5* genes. In addition, a motif C terminal to the homeodomain (KLTGP) characteristic for *Lox5* orthologs strongly supports the notion that this is a member of paralog group 6/*Lox5*. Because established paralog groups are based on vertebrate gene sequences, and because the central group *Hox* genes have no true orthologs among the invertebrate phyla (see ref. 14), we elected to name this gene *Esc-Lox5*.

Esc-Lox4 is most similar to *Lox4* from the brachiopod *Lingulina anatina*. *Esc-Lox4* shares two of the three conserved residues, Q-19, H-22, and K-27, that identify the *Lox4* homeodomain, in addition to a *Lox4*-specific motif C terminal of the homeodomain. Hence, we identify *Esc-Lox4* as the *Lox4* homolog from *E. scolopes*.

Posterior Group Orthologs. Complete homeodomain sequences for *Post-1* and *Post-2* class genes were obtained through the combination of genomic PCR and rapid amplification of cDNA ends. The *Esc-Post-1* homeodomain contains most of the previously assigned conserved residues (14). Based on the additional sequence information now available, we have designated additional conserved residues (>50% majority)(see Fig. 2). Similarly, the *Esc-Post-2* homeodomain displays very high sequence similarity to the previously identified *Post-2* sequences, in particular to the nemertean *Lineus sanguineus* *Hox9* gene (14, 19, 48). Both neighbor-joining analysis (Fig. 3) and maximum-likelihood analysis (not shown) provided support for the distinction between the posterior genes of the Lophotrochozoan *Post-1* and *Post-2* groups, and an Ecdysozoan *Abdominal-B* group, as was previously observed (14). The neighbor joining raised the interesting possibility that the *Post-2* and *Post-1* groups may be sister groups of the Ecdysozoan *Abd-B* and deuterostome *posterior* genes, respectively. More sequence information will be required to provide this intriguing possibility with the necessary bootstrap support.

Discussion

The identification of nine unique *Hox* genes (one anterior group, one PG3 group, five central group, and two posterior) from the sepiolid squid *E. scolopes*, a cephalopod, provides a systematic analysis of *Hox*-cluster genes in a mollusc. These data will enable us in future experiments to further refine the phylogenetic relationship within the Lophotrochozoa, analyze in detail the

PG-1 Labial Antp: RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRMRKWKKN

| | | | |
|----------------|---------------------------|--|----------------------------|
| Dme lab | GLGSGSLSSCSLSNST | NNS--TNF-NK-LT-----A-----NT-Q-N-T-V-----Q--RV | KEGLIPADILTQHSTS |
| Esc lab | CGNGSVGNLGGGLGAGGG | NST--TNF-NK-LT-----K---A-----A--G-N-T-V-----Q--RL | KEAQTPLGNTGNLENCALV |
| Nvi lab | KPGEFYTPPGQ | PNM--TNF-NK-LT-----K---A-----A--G-N-T-V-----Q--RM | KETNVSPTFVSV |
| Lan lab | TKQTKNYQDYSGNI | PNM--TNFSDK-LT-----K---A-----A--G-N-T-V-----Q--RM | KESQTLPTAAFGE |
| Lsa Hox1 | KNQRLCFAGQ | PNT--TNF-NK-LT-----K---A-----A--G-N-T-V-----Q--RM | KEGLVQNNTTGPRIYMDNV |
| Pca lab | | ---K---A-----AS-G-N-T-V-- | |
| Bfl Hox1 | KRNPRRTGKPEGYFTTSG | PNN--TNF-TK-LT-----Y-K---A--V---A--N-N-T-V-----Q--RE | KENGFSTPGSGGSPAGE |

PG-3 Zerknüllt

| | | | |
|-----------------|---------------------------|--|------------------------------|
| Dme zen | NLPSQFNHDSQRVK | L--S-TAF-SV-LV---N--KS-M--Y-T-----QR-S-C--V-----F--DI | QGHREPKSNAKLAQPQAEQS |
| Esc Hox3 | NLNSSRDVIPRQNIPEQP | A--A-TA--SA-LV-----Q--C-P---M-AL-N-S-----RF---K | KLKVMMDKSGCGKSSEVLNPN |
| Nvi Hox3 | GGGSGANGDSTNGEGEK | S--A-TA-NSA-LV-----C-P---M-AL-S-S-----Y--DQ | RMKPNSEKELSEDGDQYGRGN |
| Lan Hox3 | FTNTEGQP | T--A-TA--SA-LV-----C-P---M-AL-N-----HA--R | RSKKSCKSNSSN |
| Lsa Hox3 | SELTGP | P--S-TA--SA-LV-----C-P---M-AL-N-S-----Y--DQ | KQKNLMKIQLEIKTEETRP |
| Pca Hox3 | | A-LV-----C-P---M-AM-N----- | |
| Cin Hox3 | SESNSSSGDGDKCYSA | S--E-TA--NS-LV-----SH--C-P---L-QG-G-----F--Q | KQKAVLQRNNGD |
| Bfl Hox3 | NLSVGTTEPGESDGLGAA | G--A-TA--SA-LV-----Y-K---A--V---A--N-N-T-V-----Y--Q | KVKGGSGGSGGMNSPSP |

PG-4 Deformed

| | | | |
|----------------|---------------------------|---|----------------------|
| Dme Dfd | HVAGVANGSYQPME | P--Q-TA--H-I-----Y-----T-V-S-----DN | KLPTNTRVRRKTVDANGNPT |
| Esc Dfd | WMRKSHSNTGKVPVYCKN | S--S-TA--H-I-----K-----T--S----- | |
| Nvi Dfd | GANGAYGTD | S--T-TA--H-V----- | KLPTNTRLS |
| Hro Lox6 | NLSSSSSSSLIA | D--A-TS-S-----NG-----S-G-----D- | RLPNSKSGKIGSANLSSGTG |
| Lsa Hox4 | AANGNFGSE | S--S-TA--H-I-----D-S-----H | KLPTNTRLP |
| Pca Dfd | | A--H-V-----S----- | |
| Bfl Hox4 | PWMKKVHSNTGSTSYNGQD | T--S-TA--Q-V-----S-G-----D- | RLPTNTRSSSASGGSGARV |

PG-5 Sex-combs Reduced

| | | | |
|----------------|---------------------------|---------------------------------------|------------------------------|
| Dme Scr | VHLGTSTVNANGE | T--Q-TS-----H----- | KMASMNIYPYHMGYPYHQ |
| Esc Scr | PQIYPWMRRMQYSDPDGE | S--S-TS--H-----Y-K-----N-----H | KLSHIAKNMNVCAPLDKGLDI |
| Hru Hox5 | YWMRRMQYSSDGDNDGE | T--S-TS--H-----Y-K-----N-----H | KLSHIAKNMNL |
| Nvi scr | FLSKEANGVE | S--T-TS--H-----N-----H | KLGLHAKSQGTLKDPGGPID |
| Hro Hox6 | HNNNLSSSSSSSLIA | D--A-TS-S-----NG-----S-G-----D- | RLPNSKSGKIGS |
| HtrLox20 | MKRMLHCEENGEQVD | N--T-TS--H-----S-----V-N-S-----DH | PNHPGSNGLHSP |
| Lan Scr | GMNGIE | S--T-TS--H-----N-----Q | KLHLTKTEAMKLSIDGKPTV |
| Pca HB1 | | -----NV-R-----H | RLPKPVPFPFGDFGH |
| Bfl Hox5 | AGTGD | N--T-TA----- | KLKLSLSCQQTQGLNPN |
| Cin Hox5 | IYPWMKRHHGGTDPDP | S--T-TA-----Y-----V--T----- | KLSSLNSVTHVS |

Antennapedia

| | | | |
|-----------------|---------------------------|----------------|----------------------|
| Dme Antp | | | KTKGEPGSGGEG |
| Esc Antp | NSATTLPIYPWMRSQYGP | | KAEMPGTENGKQP |
| Lsa Hox7 | RIDYPIIPDVCIFVIGPD | | KQPVGIVCKEEEEP |
| Lan HB1 | PD | ---H--S-H----- | KGIELSRDPPQGV |
| Pca HB2 | D | -----NM-----C | KARLIPNGDFPKELTK |

Lox 5

| | | | |
|-----------------|---------------------------|----------------------------------|------------------------------|
| Esc Lox5 | MRPVNGDPLILEEITAYE | Q--T-----F-----S-G-S----- | NVSKLTGPDKSKMDDHRPNKN |
| Nvi Lox5 | EFGFE | Q--T-----Y-----G----- | NLSK |
| Hro Lox5 | NVPIYPWMRSVFGDPGFD | Q--T-----YS-----S-A-S----- | NVQKLTGPGGVGDIPPVSSQN |
| Lan Lox5 | IGYE | Q--T-----Y-----H-G----- | NIPKLTGPNQIDETSTKT |
| Lsa Lox5 | YRFRNANRE | Q--T-----K-----G----- | NLQKLTGENTKLEPQDPTK |

Lox 4

| | | | |
|-----------------|---------------------------|--|------------------------------|
| Dme Ubx | AGSLLPDWLGTNGL | -R-----T-H-----M-----L--I | QAIKELNEQEQAQ |
| Dme AbdA | RVVCGDFNGPNGCP | -R-----F-----H-----L--L | RAVKEINEQARRDREE |
| Esc Lox4 | DHNSVFYPMGIVGPNSSQ | -R-----S-F-----QY-N---K---V---N-S--V-----L--K | QOIREMNGACRSQDSGSEDGD |
| Nvi Lox4 | RSNQE | EE-----S-----R-H---K---V---H-----LR--R | LQIKELNRDITGSLGPKSKD |
| Lan Lox4 | PNSAQ | -R-----S-F-----Q--H---K---V-----L--R | QAIKEINDTLKK |

Post 1

| | | | |
|------------------|----------------------|--|---------------|
| Esc Post1 | IHLPSALPPSAIA | LRKR-RP-SK--IA--R-YALST-ISK--W-LSQL-N-S-----I-A--LQ | KRDETLLKTQTPN |
| Nvi Post1 | PSSMTPRHINIGPTLH | MRKK-KP-SK--IA--YVN-T-I-KPK-W-LSQR-N-S--V-----E--VT | DKKDD |
| Lan Post1 | TTLPAVH | MRKK-KP-SK--IA--R-YVS-T-ISKPK-W-LSQR-Q-S--V-----E--VK | GGKQT |

Post 2

| | | | |
|------------------|-------------------------|---|--------------------------|
| Esc Post2 | TGQGDHLFDSQGSHTK | GRKK-KP-----MV--N--LNSS-I--QK-W--SCK-Q-----V-V-----R--L- | ERAKARLREDRENDRQA |
| Nvi Post2 | DQPR | QRKK-KP-----MV--N--MG-S-I--QK-W--SCK-H-S--V-V-----R--L- | ERAKTLIKSDSDGMT |
| Lsa Hox9 | YTWFSPFSTTEPR | TRKK-KP-----MV--N--LT-S-I--QK-W--SCK-H-----V-V-----R--L- | ARSVKTRMTDANGMTN |
| Lan Post2 | | -KP-----MV--N--LN-A-I--QK-W--SCK-H-S--V-V | |
| Dja abdBa | SGSLTEAEMR | SRKK-KP-----MV--S--TG-A-I--QK-W--SCK-H-S--VKV-----K--LQ | TRKPGSS |
| Tgr Hbox4 | SATS | GRKK-CP--KF-----L--M-L--D--L--RL-S-T--V-----M--Q- | RAQNY |
| Dme abdB | CTPNPGLHEWTGQVS | VRKKRKP-SKF-----L--A-VSKQK-W-L-RN-Q-----V-----N--NS | QRANQNNNNNS |

Fig. 2. Comparisons of known Hox orthologs and homeobox-containing genes with predicted *E. scolopes* homeodomains. Only representative homeodomains and short flanking sequences from individual classes or phyla are shown to simplify the diagrams and to highlight phyletic distances. Emphasis is placed on organisms for which an array of Hox sequences are known to facilitate the orthology comparisons. The *Drosophila* Antennapedia homeodomain sequence is shown at the top of the alignments. Nonidentical amino acids are shown, and dashes in the subject sequences indicate identity. Paralog-specific amino acids (light gray) and paralog-specific signature motifs (dark gray) are shaded. See Table 1 for abbreviations.

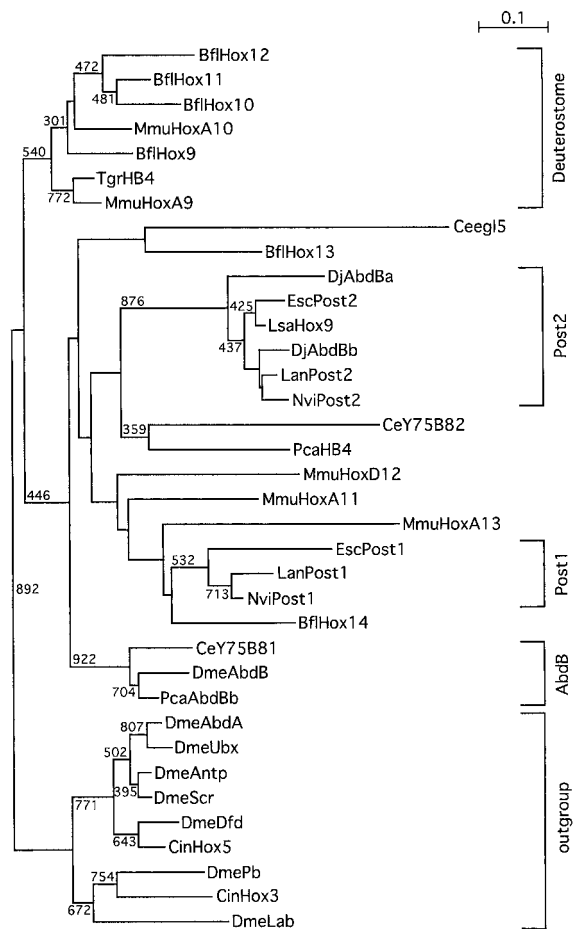


Fig. 3. Phylogram analysis of *E. scolopes* Post-1 and Post-2 homeodomain fragments and other known posterior homeodomain sequences, as revealed by neighbor-joining analysis. Existing representative sequences from Lophotrochozoa, Ecdysozoa, and deuterostomes were selected for the comparison. The *Drosophila* Hox genes were used as an outgroup. One thousand replicate data sets were generated. Internal edge labels refer to the number of bootstraps supporting the partition. Bootstrap values below 300 were not included in the phylogram. High support is found for the Lophotrochozoan Post-1 and Post-2, and Ecdysozoan Abd-B genes. The scale shows the number of amino acid substitutions per site.

genomic organization and the sequence of a Lophotrochozoan *Hox* cluster, and study the spatiotemporal expression pattern and the role of the *Hox* genes in the formation of the very complex morphology in cephalopods.

Cephalopod *Hox* Genes and the Relationship with Other Lophotrochozoa. Recent molecular data derived from the comparative analysis of 18S rRNA genes indicate that the bilaterian phylogenetic tree is divided into three clades, the Lophotrochozoa, Ecdysozoa, and Deuterostomia (29, 49). The Lophotrochozoa comprise Annelida, Mollusca, Sipunculida, Echiurida, Plathelminthes, Nemertea, Brachiopoda, and Ectoprocta, whereas the Ecdysozoa include Arthropoda, Nematoda, and other molting phyla (49). Our data on *Hox* cluster genes in the cephalopod *E. scolopes*, and in particular the identification of *Lox4*, *Lox5*, *Post-1* and *Post-2* homologs, are consistent with this subdivision of the protostomes.

A Single *Hox* Cluster in Cephalopods? An archetypic Lophotrochozoan *Hox* gene cluster has been proposed to consist of two anterior genes, one PG3 ortholog, five central genes, and two

posterior class genes (14). This number gives an estimate of the number of *Hox* cluster genes we can expect to find in *E. scolopes*. Of these, we identified all except a paralog group 2 homolog (*proboscipedia*). Previous studies on several Lophotrochozoan species have variably reported the presence of paralog 2 and posterior group genes. A paper on the primitive gastropod *Haliotis* suggested the existence of three members of the anterior group of *Hox* genes, including a *Hox2* paralog, and three members of the central group (50). One anterior group gene and five central group genes have been identified in the gastropod *Patella* (14). *Hox2/proboscipedia* orthologs have also been identified in two polychaetes (14, 44). One explanation for these differing results using PCR methods could be the presence of an intron (19). In addition to *Hox2/proboscipedia* orthologs, earlier studies yielded different results with respect to posterior group genes. Two groups failed to identify *Abd-B* homologs during comprehensive surveys of two polychaetes, *Chaetopterus varipodatus* (45) and *Ctenodrilus serratus* (44). A posterior class gene most similar to *AbdB* was found in oligochaeta (44), and two posterior *Hox* genes were found in the polychaete *Nereis* (14). A member of the posterior group of *Hox* genes in the ribbonworm could not be identified with PCR methods, but eventually was isolated by using low-stringency hybridization methods (19). The identification of posterior group genes in Lophotrochozoan and Ecdysozoan protostomes and in deuterostomes supports the existence of the posterior group of *Hox* genes in the last common ancestor of deuterostomes and protostomes. A number of recent analyses indicate that the posterior class even predates the last common ancestor of Cnidarians and Bilateria. Our data and recent data from other Lophotrochozoa (14) show the existence of divergent posterior group genes in a number of protostome phyla, which helps explain the apparent limitations of the PCR-based approach in identifying this paralog group. The use of specific primers as shown in this article or genomic walking should enable the characterization of all posterior group genes where they are present. Eventually, the identification of additional posterior *Hox* genes from animals of different phyla should help resolve whether the posterior *Hox* genes were inherited from a common ancestor, or whether they are the result of independent duplication events.

In conclusion, based on our results and by extrapolation of the data sets mentioned above, we propose that an archetypal mollusc *Hox* cluster contained at least 10 genes (two anterior group genes, one PG3 ortholog, five central group genes, two posterior group genes), and that a cephalopod *Hox* cluster contains at least nine of those genes. It remains possible that additional, divergent posterior group genes will be discovered. Furthermore, we have no evidence for multiple genes of any given paralogy group. Therefore, our data are consistent with the existence of a single copy of a presumed *Hox* cluster in cephalopods.

***Hox* Genes and Mollusc Body Plan Evolution.** *Hox* genes are one group among many transcriptional regulators that qualify as “master regulator genes” of development (51–53), and they determine regional specializations along the anteroposterior body axis. The *Hox* cluster duplications observed in vertebrates have given rise to a complex interacting network of *Hox* genes controlling many aspects of skeletal, muscle, and neuronal differentiation. On the other hand, there are animals, which possess a single *Hox* gene cluster, yet display a sophisticated morphological organization and complexity, as in *Drosophila*, and the cephalopod *E. scolopes*, as shown in this article. The molluscs comprise classes with species ranging from microscopic wormlike representatives to the largest and behaviorally most sophisticated invertebrates known. Yet, our data and previous reports (14, 50) strongly suggest that molluscs have only one *Hox* cluster. We therefore propose that the rich diversity in body plans observed among molluscs is more likely the consequence

of the recruitment or intercalation of different target genes into existing regulatory networks (intercalary evolution, ref. 54), and that subtle changes in the control regions of both *Hox* genes and their downstream effectors have led to significant alterations of developmental programs (10). In a recent study comparing *Hox* sequences between distantly related metazoans, the authors came to the conclusion that many of the amino acid replacements used as diagnostic criteria for particular paralog groups may represent functionally significant substitutions because they are likely to be localized on the surface of the respective proteins (55). Protein-protein interactions between homeodomains and between homeodomains and paired domains have recently been demonstrated (56). Thus, specific amino acid substitutions may signify the gain or loss of a particular interaction between proteins, and thereby represent indicators of morphological change. The use of *Hox* gene expression pattern analysis to identify homologous structures and the isolation of target genes will allow the characterization of changes in gene regulation that were instrumental in mollusc evolution. In addition, the identification of the *Hox* gene complement in other mollusc species and the identification of differentially regulated target genes will contribute to our understanding of the morphological diversification within the phylum Mollusca, and by extension also across animal phyla.

- Conway Morris, S. (1993) *Nature (London)* **361**, 219–225.
- Valentine, J. W. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 6751–6757.
- Adamkewicz, S. L., Harasewych, M. G., Blake, J., Saudek, D. & Bult, C. J. (1997) *Mol. Biol. Evol.* **14**, 619–629.
- Brusca, R. & Brusca, G. (1990) *Invertebrates* (Sinauer, Sunderland, MA).
- Ruppert, E. E. & Barnes, R. D. (1994) *Invertebrate Zoology* (Saunders, Philadelphia).
- Carroll, S. (1995) *Nature (London)* **376**, 479–485.
- Scott, M. P. (1994) *Cell* **79**, 1121–1124.
- Averof, M. & Akam, M. (1995) *Nature (London)* **376**, 420–423.
- Davidson, E. H., Cameron, R. A. & Ransick, A. (1998) *Development (Cambridge, U.K.)* **125**, 3269–3290.
- Gellon, G. & McGinnis, W. (1998) *BioEssays* **20**, 116–125.
- Gehring, W. J., Affolter, M. & Bürglin, T. (1994) *Annu. Rev. Biochem.* **63**, 487–526.
- Bürglin, T. R. (1994) in *Guidebook to the Homeobox Genes*, ed. Duboule, D. (Oxford Univ. Press, London), pp. 25–71.
- Ruddle, F. H., Bartels, J. L., Bentley, K. L., Kappen, C., Murtha, M. T. & Pendleton, J. W. (1994) *Annu. Rev. Genet.* **28**, 423–442.
- de Rosa, R., Grenier, J. K., Andreeva, T., Cook, C. E., Adoutte, A., Akam, M., Carroll, S. B. & Balavoine, G. (1999) *Nature (London)* **399**, 772–776.
- Krumlauf, R. (1994) *Cell* **78**, 191–201.
- Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y. L., et al. (1998) *Science* **282**, 1711–1714.
- García-Fernández, J. & Holland, P. W. H. (1994) *Nature (London)* **370**, 563–566.
- Akam, M. (1987) *Development (Cambridge, U.K.)* **101**, 1–22.
- Kmita-Cunisse, M., Loosli, F. & Gehring, W. J. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 3030–3035.
- Martinez, P., Rast, J. P., Arenas-Mena, C. & Davidson, E. H. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 1469–1474.
- Kenyon, C. & Wang, B. (1991) *Science* **253**, 516–517.
- Bürglin, T. R., Ruvkun, G., Coulson, A., Hawkins, N. C., McGhee, J. D., Schaller, D., Wittmann, C., Muller, F. & Waterston, R. H. (1991) *Nature (London)* **351**, 703.
- Salsler, S. J. & Kenyon, C. (1994) *Trends Genet.* **10**, 159–164.
- Valentine, J., Erwin, D. & Jablonski, D. (1996) *Dev. Biol.* **173**, 373–381.
- Jacob, F. (1977) *Science* **196**, 1161–1166.
- Duboule, D. & Wilkins, A. S. (1998) *Trends Genet.* **14**, 54–59.
- Arnold, J., Singley, C. & Williams-Arnold, L. (1972) *Veliger* **14**, 361–365.
- Schubert, F. R., Nieselt-Struwe, K. & Gruss, P. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 143–147.
- Aguinaldo, A. M., Turbeville, J. M., Linford, L. S., Rivera, M. C., Garey, J. R., Raff, R. A. & Lake, J. A. (1997) *Nature (London)* **387**, 489–493.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
- Murtha, M. T., Leckman, J. F. & Ruddle, F. H. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10711–10715.
- Pendleton, J. W., Nagai, B. K., Murtha, M. T. & Ruddle, F. H. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 6300–6304.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Altschul, S. F., Madden, T. L., Schaeffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997) *Nucleic Acids Res.* **25**, 4876–4882.
- Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. (1998) *Trends Biochem. Sci.* **23**, 403–405.
- Kimura, M. (1980) *J. Mol. Evol.* **16**, 111–120.
- Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, New York).
- Strimmer, K. & von Haeseler, A. (1996) *Mol. Biol. Evol.* **13**, 964–969.
- Strimmer, K. & von Haeseler, A. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 6815–6819.
- Henikoff, S. J. & Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 10915–10919.
- Page, R. D. M. (1996) *Comput. Appl. Biosci.* **12**, 357–358.
- Dick, M. H. & Buss, L. W. (1994) *Mol. Phylogenet. Evol.* **3**, 146–158.
- Irvine, S. Q., Warinner, S. A., Hunter, J. D. & Martindale, M. Q. (1997) *Mol. Phylogenet. Evol.* **7**, 331–345.
- Giusti, A. F., Hinman, V. F., Degnan, S. M., Degnan, B. M. & Morse, D. E. (2000) *Evol. Dev.* **2**, 294–302.
- Kourakis, M. J., Master, V. A., Lokhorst, D. K., Nardelli-Haeffliger, D., Wedeen, C. J., Martindale, M. Q. & Shankland, M. (1997) *Dev. Biol.* **190**, 284–300.
- Nogi, T. & Watanabe, K. (2001) *Dev. Growth Differ.* **43**, 177–184.
- Balavoine, G. & Adoutte, A. (1998) *Science* **280**, 397–398.
- Degnan, B. M. & Morse, D. E. (1993) *Mol. Mar. Biol. Biotechnol.* **2**, 1–9.
- Gehring, W. J. (1996) *Genes Cells* **1**, 185–192.
- Weintraub, H., Tapscott, S. J., Davis, R. L., Thayer, M. J., Adan, M. A., Lassa, A. B. & Miller, A. D. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5434–5438.
- Lewis, E. B. (1992) *J. Am. Med. Assoc.* **267**, 1524–1531.
- Gehring, W. J. & Ikeo, K. (1999) *Trends Genet.* **15**, 371–377.
- Sharkey, M., Graba, Y. & Scott, M. P. (1997) *Trends Genet.* **13**, 145–151.
- Plaza, S., Prince, F., Jaeger, J., Kloter, U., Benassayag, C., Cribbs, D. & Gehring, W. J. (2001) *EMBO J.* **20**, 802–811.
- Dolecki, G. J., Wang, G. & Humphreys, T. (1988) *Nucleic Acids Res.* **16**, 11543–11558.