Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 7338–7342, August 1991
Developmental Biology

Planarian homeobox genes: Cloning, sequence analysis, and expression

(homeodomain/pattern formation/regeneration)

JORDI GARCIA-FERNÀNDEZ, JAUME BAGUÑÀ, AND EMILI SALÓ*

Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain

Communicated by Walter J. Gehring, April 26, 1991

ABSTRACT **Freshwater planarians (Platyhelminthes, Turbellaria, and Tricladida) are acoelomate, triploblastic, unsegmented, and bilaterally symmetrical organisms that are mainly known for their ample power to regenerate a complete organism from a small piece of their body. To identify potential pattern-control genes in planarian regeneration, we have isolated two homeobox-containing genes, *Dth-1* and *Dth-2* [*Dugesia (Girardia) tigrina* homeobox], by using degenerate oligonucleotides corresponding to the most conserved amino acid sequence from helix-3 of the homeodomain. *Dth-1* and *Dth-2* homeodomains are closely related (68% at the nucleotide level and 78% at the protein level) and show the conserved residues characteristic of the homeodomains identified to date. Similarity with most homeobox sequences is low (30–50%), except with *Drosophila NK* homeodomains (80–82% with *NK-2*) and the rodent *TTF-1* homeodomain (77–87%). Some unusual amino acid residues specific to NK-2, TTF-1, Dth-1, and Dth-2 can be observed in the recognition helix (helix-3) and may define a family of homeodomains. The deduced amino acid sequences from the cDNAs contain, in addition to the homeodomain, other domains also present in various homeobox-containing genes. The expression of both genes, detected by Northern blot analysis, appear slightly higher in cephalic regions than in the rest of the intact organism, while a slight increase is detected in the central period (5 days) of regeneration.**

The homeobox, a 180-base-pair (bp) DNA sequence, was first discovered as a region of sequence similarity between genes involved in the control of *Drosophila* development (for review, see refs. 1 and 2). It was later found in several higher metazoans, but in lower metazoans it has only been found in nematodes (3). Proteins coded by homeobox-containing genes act as transcription factors in which the homeodomain is partially or fully responsible for sequence-specific recognition of DNA (4). These genes seem to be involved in a number of regulative processes, such as cell determination and pattern formation during development, and in cell differentiation (for review, see ref. 5).

Freshwater planarians are widely known for their ability to regenerate and for their ability to grow and degrow (to shrink in volume and length), depending on body size, temperature, and food availability (6). Both phenomena, which involve continuous processes of cell proliferation, cell determination and differentiation, and pattern formation confer an extreme morphological plasticity on planarians. Regeneration and growth/degrowth seem to have a common cellular basis: the presence of a population of small undifferentiated self-renewing cells (called neoblasts) that are scattered throughout the body, amount to 25–30% of total cells, and give rise to all differentiated cell types in the intact organism and to blastema cells during regeneration (7–9).

Regeneration in planarians is fairly well understood at the tissue and cell levels. Moreover, some molecular details [e.g., changes in the pattern of protein expression (10) and action of mitogenic factors (11, 12)] are known, though no genetic data have so far been produced. To further our understanding of regeneration and to detect, isolate, and study genes that play a controlling role in this phenomenon, planarian genomic and cDNA libraries were screened with probes of the *Drosophila* homeobox or with degenerate oligonucleotides corresponding to the most conserved amino acid sequence of the homeodomain (3).

This work reports the detection, isolation, and sequencing of two homeobox-containing genes† from a cDNA library of the planarian *Dugesia (Girardia) tigrina*. Sequence analysis of both cDNAs shows the presence of a potential family of homeodomains shared by the *Drosophila NK-2* gene (13) and the rodent *TTF-1* gene (14). In addition to the homeodomain, other domains are also present in these genes. Their expression was studied by Northern blot analysis of tissue from different regions of intact organisms and in different stages of regeneration.

MATERIAL AND METHODS

Species. Planarians used in this study belong to the species *Dugesia (G.) tigrina* (Girard) collected near Barcelona. Two-week-starved organisms were used in all experiments.

Isolation of cDNA Clones. A cDNA library was prepared in λgt10 from intact and regenerating (1 hr and 5, 7, and 10 days) planarians as described (15). The library contains oligo(dT)-primed cDNAs and was prepared from RNAs of >0.5 kilobases (kb).

Approximately 4×10^5 recombinant clones in λgt10 were probed with degenerate oligonucleotides corresponding to the most conserved sequence found in helix-3 of the homeodomain *HB-1* (3): AA(A/G) ATX TGG TT(T/C) CA(A/G) AA(C/T) (A/C)GX (A/C)G. Synthetized oligonucleotides were 5′-end-labeled with T4 kinase according to Maniatis *et al.* (16) using [γ-^{32}P]ATP (6000 Ci/mmol; 1 Ci = 37 GBq; New England Nuclear). Hybridizations were performed as described by Bürglin *et al.* (3).

Nucleotide Sequencing and Analysis. Homeobox-hybridizing *Eco*RI fragments were subcloned into pBluescript (Stratagene) using standard techniques (16) and sequencing was carried out on both strands using T7 DNA polymerase (Pharmacia) according to the manufacturer's instructions. Some initial sequences were obtained using degenerate oligonucleotides as primers and sequencing strategy was performed by unidirectional digestion with exonuclease III (17). Nucleotide sequence data were analyzed with the sequence analysis software package of ref. 18 from the University of Wisconsin Genetics Computer Group.

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.

*To whom reprint requests should be addressed.

†The sequences reported in this paper have been deposited in the GenBank data base (accession nos. X56499 and X56500).

Northern Blot Analysis. Total cellular RNA was isolated by the guanidinium thiocyanate method (19) from heads and bodies of intact planarians and in various regenerative stages. Poly(A)$^+$ RNAs were prepared by retention on oligo(dT) columns, fractionated on a 1.2% agarose/formaldehyde gel, and transferred to a nylon membrane (Amersham) by standard procedures (16). DNA probes were ^{32}P-labeled by random priming and hybridized to filters at high stringency [42°C in 50% (vol/vol) formamide/5× SSPE/5× Denhardt's solution/0.1% SDS/tRNA (100 μg/ml) (1× SSPE = 0.18 M NaCl/10 mM sodium phosphate, pH 7.4/1 mM EDTA)]. Filters were washed (65°C in 0.5× SSPE/0.1% SDS) and exposed to Kodak XAR-5 film with two intensifying screens. The resulting autoradiographs were measured with an Ultra-Scan-XL enhanced laser densitometer (Pharmacia LKB).

RESULTS

Isolation and Sequence Analysis of Two cDNAs Containing a Homeobox. A planarian cDNA library of 4 × 10^5 plaques was screened using replica filters (Schleicher & Schuell) with a 5'-end-labeled mixture of 1024-fold degenerate oligonucleotides (HB-1) (3). We isolated seven λ clones of various sizes. All phage insert DNAs were subcloned into pBluescript (Stratagene) and mapped with restriction endonucleases. All seven clones had inserts between 0.7 and 1.7 kb long and fell into two groups according to their restriction map (Fig. 1).

The nucleotide sequence shown in Fig. 2 is from five (*Dth-1*) and two (*Dth-2*) cDNA clones. The longest open reading frame of *Dth-1* starts with the ATG at position 131 and codes for a putative protein of 451 amino acids that terminates with an ochre codon at position 1484. The termination signal is followed by 189 bases of untranslated trailer sequence comprising two canonical poly(A) addition sequences (AATAAA), 45 and 12 nucleotides upstream of the insertion site, respectively. The *Dth-2* coding sequence starts with the ATG at position 129 and potentially codes for a protein of 363 amino acids. An amber codon appears at position 1216. The termination signal is followed by only 76 bases, and a putative polyadenylylation signal (AATAGA) is located 12 nucleotides from its end. A Dth-1-specific transcript of 1.7 kb and a Dth-2-specific transcript of 1.4 kb were detected by Northern blot analysis, showing similar size to the corresponding cDNAs clones isolated.

The base composition analysis from *Dth-1* and *Dth-2* shows a high percentage of A+T bases, 64% in the coding regions of both genes and 70–80% in noncoding regions. There is a strong bias toward using codons with an adenine or thymine in the third position. In *Dth-1* and *Dth-2*, 65% and 83% of the redundant codons, respectively, utilize an adenine or thymine in the third position.

Features of Deduced Amino Acid Sequences. The amino acid sequences of Dth-1 and Dth-2 reveal a 60-amino acid region similar to the homeodomain (Fig. 2). Dth-1 and Dth-2 homeodomains conserve the hydrophobic amino acids known to be important in folding and stabilizing the predicted four α-helices (20, 21) and two clusters of basic residues in the homeodomain boundaries (22). The homology between *Dth-1* and *Dth-2* is restricted to the homeodomain, with values of 78% at the protein level and 68% at the nucleotide level. Helix-2 and helix-3 are identical, and the 13 substitutions observed are mainly conservative (23) and scattered in the remaining homeodomain region.

The presumptive Dth-1 and Dth-2 proteins presented some salient features in their primary structure (Fig. 2). In Dth-1 the regions outside the homeodomain are rich in asparagine (14%) and serine (12%) and three repeats rich in asparagine and serine are at positions 3–8, 201–289, and 428–451. Furthermore, an acidic region (positions 260–269) and a proline-rich domain interspersed with acidic amino acids (positions 18–188) are found. Such domains are thought to act as transcription activator domains. Meanwhile, the regions outside the homeodomain in Dth-2 are rich in serine (16%) and alanine (9%), showing the presence of two serine-rich regions (positions 29–42 and 319–331) and two alanine- and serine-rich regions also located at the N-terminal and C-terminal regions (positions 62–94 and 348–363). The amino-terminal domain (positions 1–151) is also typified by the low number (5%) of charged amino acids. In contrast, residues 193–364 contain 22% charged residues. Such regions rich in particular amino acids have been found in many other homeodomain proteins.

Homeodomain Comparison. Comparison of Dth-1 and Dth-2 homeodomains with representative members of classes of homeodomains previously described (Fig. 3) shows a relatively low percentage of identity of 30–50%. The best similarity is observed with the homeoboxes of the *Drosophila* genes *NK-2* (80–82%), *NK-3*, and *NK-4* (13) and the rodent gene *TTF-1* (77–87%) (14). Residues 42 and 52 in the homeodomains of Dth-1, Dth-2, NK-2, and TTF-1 are proline and histidine, respectively, and belong to the most variable positions in the recognition helix (2, 25). Such residues have been found only in the yeast mating type factor a1 (Pro-42) and the *Drosophila* gene cut (His-52). In helix-2 and helix-4, we also found some other unusual residues that are shared with Dth-1, Dth-2, NK-2, and TTF-1, histidine (residue 33), isoleucine [residue 38 also present in yeast *PHO* gene product (26)], and tyrosine (residue 54). Such a combination of unusual residues, to our knowledge, has never been reported in any homeodomain and would define a family of homeodomains. Moreover, the presence of some other unusual amino acid replacements shared with some Dth, NK, and TTF-1 homeodomains scattered in the helix-1 and in the hinge region (Fig. 3) provides additional evidence that the planarian *Dth-1* and *Dth-2* homeobox genes are related to *NK Drosophila* and *TTF-1* rodent genes.

Dth-1 and Dth-2 mRNA Expression. Northern blot analysis identified a single Dth-1 transcript of 1.7 kb and a single Dth-2

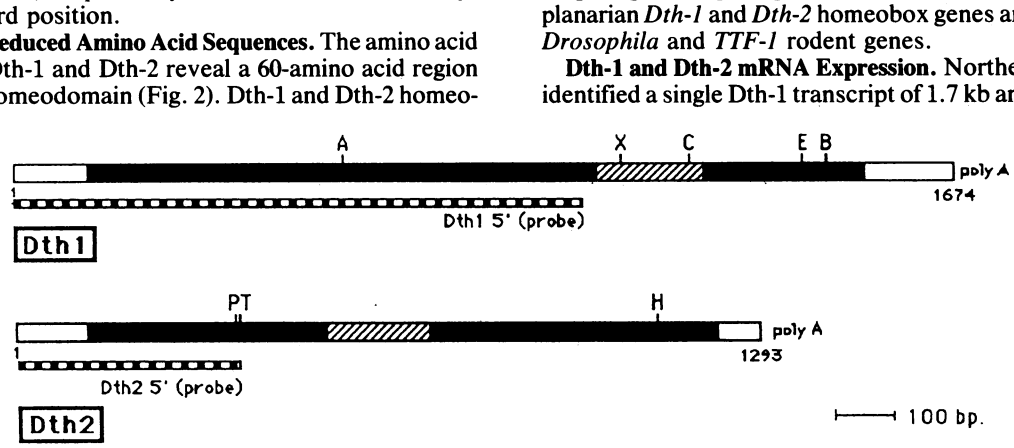

FIG. 1. Restriction map of cDNA clones for *Dth-1* and *Dth-2*. The open reading frames are solid boxes and the homeobox regions are hatched boxes. The probes used in Northern blot analysis (Dth-1 5' and Dth-2 5') are also indicated. Relevant restriction sites are *Apa* I (A), *Bam*HI (B), *Cla* I (C), *Eco*RI (E), *Hin*dIII (H), *Pvu* II (P), *Pst* I (T), and *Xba* I (X).

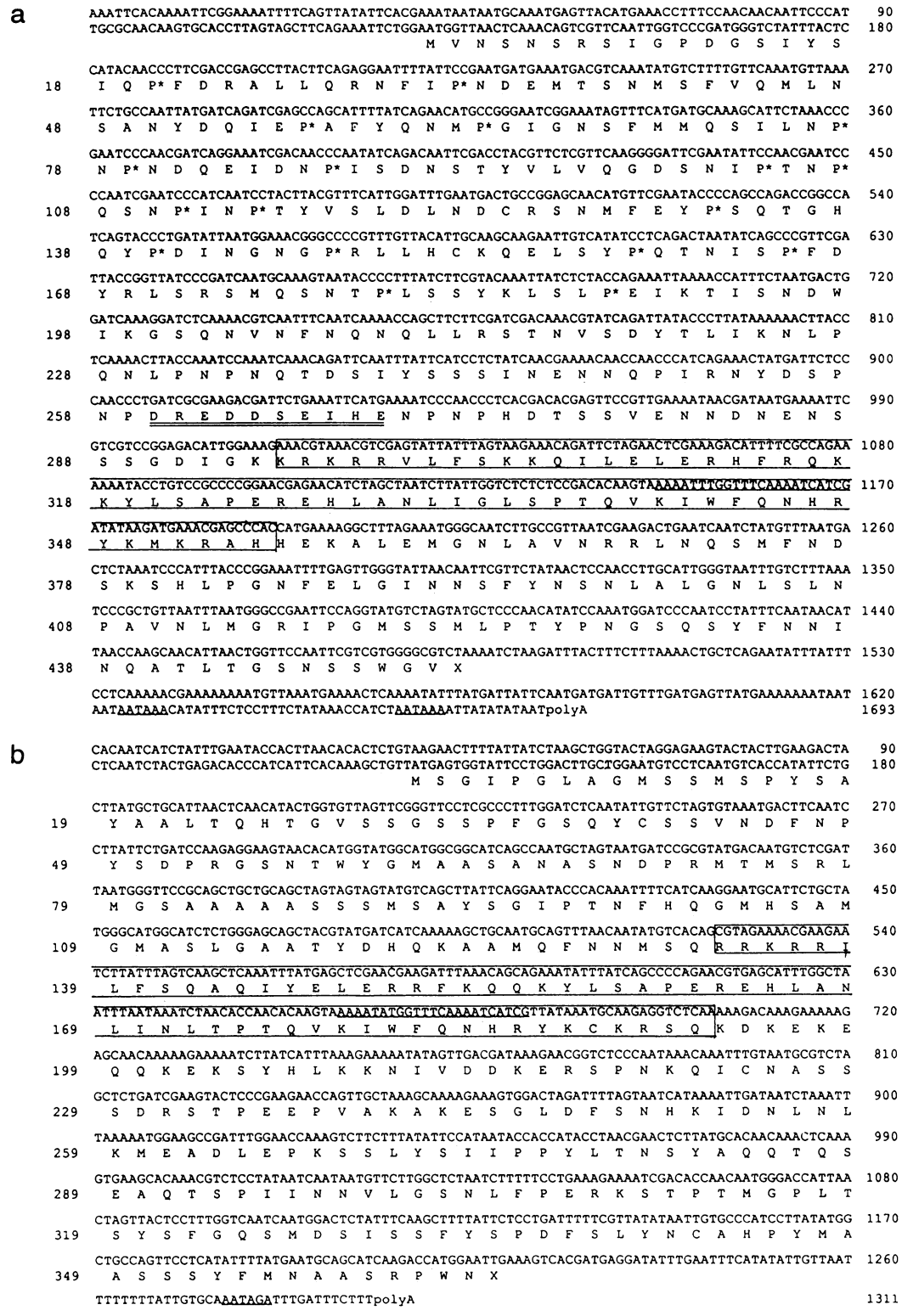

FIG. 2. cDNA sequences and predicted protein sequences of *Dth-1* (*a*) and *Dth-2* (*b*) genes. Nucleotides are numbered at the right margin, and the deduced amino acids are numbered at the left. The homeoboxes are boxed (nucleotides 1013–1192 in *Dth-1* and 525–704 in *Dth-2*). The acidic region is double underlined and the proline residues of the proline-rich region are indicated by asterisks. The region homologous to the synthetic degenerate oligonucleotides HB-1 and the polyadenylylation signals are underlined.

transcript of 1.4 kb by comparison with an RNA ladder (Bethesda Research Laboratories) used as standard. After densitometric scanning analysis of two independent Northern blots, both transcripts were found to be 2 times higher in the head region than in rest of the body (Fig. 4*A*). During regeneration, *Dth-1* and *Dth-2* expression was continuous, with an increase of 2.4 times at 5 days (Fig. 4*B*).

DISCUSSION

We have isolated two homeobox-containing genes, *Dth-1* and *Dth-2*, in *Dugesia (G.) tigrina*. DNA sequence analyses of both cDNAs and Northern blot experiments provide definitive evidence that the genome of the acoelomate nonsegmented planarians includes homeobox-containing genes that are transcribed in the intact and in the regenerating organism.

FIG. 3. Comparison of Dth-1 and Dth-2 homeodomains (single-letter amino acid code) with representative members of the classes of homeodomain-containing proteins defined (2, 13, 24). The percentage homologies are shown on the right. The positions of the four helices are indicated by open boxes and the two basic clusters with thick bars. The invariant amino acids of the homeodomain (●) and the highly conserved amino acids (○) are indicated.

The *Dth-1* and *Dth-2* homeodomains show a high percentage of similarity with each other but a very limited similarity (between 30 and 50%) to most homeodomain sequences, except those of *Drosophila NK* genes (13, 28) and rodent *TTF-1* genes (14). Dth-1, Dth-2, NK-2, and TTF-1 have a glutamine at position 9 of the recognition helix, which would qualify them as belonging to Antp type DNA-binding specificity (25). However, the unusual amino acid replacements at positions 1 and 11 of the same helix added to the other replacements found in helix-2, helix-1, and the hinge region, shared only by *Dth-1, Dth-2, Drosophila NK* genes, and rodent *TTF-1* genes, may determine specific DNA-binding specificities and thus define a family of homeodomains. Moreover, the consensus recognition sequence recognized by the TTF-1 homeodomain shows no similarity to the sequence recognized by antennapedia (14).

The codon bias observed in *Dth-1* and *Dth-2*, also found in several genes from lower metazoans (29), may explain the unsuccessful earlier attempts to isolate, even at low stringency, planarian homeobox genes using heterologous probes from *Drosophila* and *Xenopus* (unpublished data). Thus, *Dth-1* and *Dth-2* homeodomains are 80 and 82% similar to the *Drosophila NK-2* homeodomain at the protein level but only 66 and 65%, respectively, similar at the nucleotide level. Moreover, the codon bias introduces an additional hindrance, reducing the uninterrupted nucleotide sequence matches of *Dth-1* and *Dth-2* and the *NK-2* homeodomain to a maximum length of 11 bp, the average region of perfect match being only 2.8 bp long.

Although *Dth-1* and *Dth-2* bear a strong similarity to *Drosophila NK* genes, namely *NK-2*, the lack of functional data for the latter precludes the drawing of any parallel with regard to the function of the former. The striking homology of *Dth-1, Dth-2*, and rodent *TTF-1* could not correspond in developmental functions, since the tissues of expression in rat (thyroid and lung) do not have counterparts in invertebrates. *Dth-1* and *Dth-2* are expressed in the adult intact organism, with slightly higher levels in the head. The main feature of the head region, compared to other body regions, as expected, is its high density in nerve cells (6). Thus, both genes could be involved in the determination and/or differentiation of nerve cells. In the intact organism, most nerve cells have a finite life-span, being continuously replaced by new nerve cells produced by differentiation from undifferentiated cells or neoblasts (30). So, perhaps the continuous expression of *Dth-1* and *Dth-2* is somehow connected to the continuous replacement of cells—namely, nerve cells. The possibility remains, however, that such activity may be involved in some as yet unrecognized physiological regulation that differs from cell replacement and that has its maximum expression in the cephalic region.

The changes in activity of *Dth-1* and *Dth-2* during regeneration were unexpectedly small given their presumed regulatory role. To account for this, two main situations could be envisaged. (*i*) *Dth-1* and *Dth-2* may be involved in physiological processes not bearing directly on developmental, regenerative, or cell differentiation processes. (*ii*) Regeneration is a local phenomenon whose main processes occur in a narrow strip (300–400 μm) of tissue near the wound (9). Since poly(A)$^+$ RNA used for Northern blots was obtained from the whole regenerant, actual variations in the expression of both genes in the wound area may have been blurred by bulk RNA.

Although variations in *Dth-1* and *Dth-2* expression during regeneration are small, both show a maximum at 5 days. This is the period when nerve cell differentiation is at its height (7), thus supporting the presumed role of these genes in nerve-cell specification. In any case, the fact that *Dth-1* and *Dth-2* are expressed in the intact organism, in which cell turnover but not regeneration is occurring, may indicate that their increase during regeneration is linked to the enhancement of processes such as cell determination and differentiation, occurring continuously in intact organisms.

Are *Dth-1* and *Dth-2* the only representatives of homeobox containing genes in freshwater planarians? By screening the same cDNA library with various degenerate oligonucleotides that code for the helix-3 region of the classes paired and *POU* homeodomains, we have obtained several positive clones

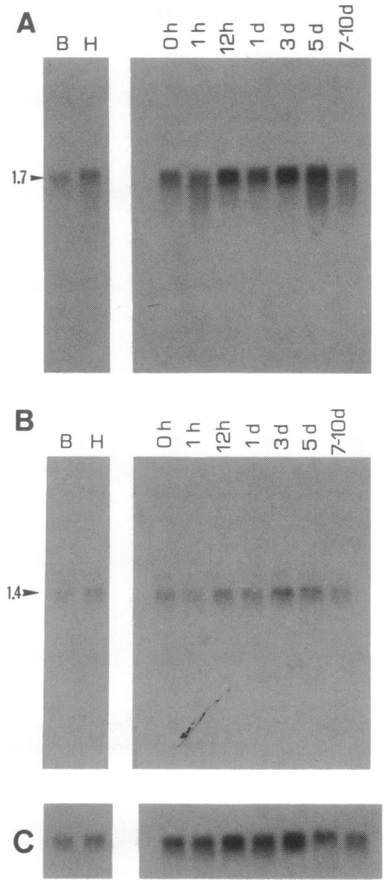

FIG. 4. Northern blot analysis of poly(A)$^+$ RNA, 6 μg per lane, from heads (lanes H) and bodies (lanes B) and poly(A)$^+$ RNA, 10 μg per lane, from intact animals (lanes 0h) and animals in various stages of regeneration [lanes 1h, 12h, 1d, 3d, 5d, and 7–10d; where h is hour(s) and d is day(s) of regeneration]. RNAs were electrophoretically separated on 1.2% agarose/formaldehyde gel, transferred to nylon membrane (Amersham) by capillary blotting, and probed with random-primed ^{32}P-labeled DNA from the 5′ end (without the homeobox sequence) of *Dth-1* (994-bp fragment obtained by unidirectional digestion with exonuclease III) (*A*) and of *Dth-2* (an *Eco*RI(artificial)–*Pst* I 371-bp fragment) (*B*). Blot was stripped and rehybridized with a *Drosophila* 5C actin gene probe (27) to control for levels of RNA.

with specific inserts of 1.5–3 kb (unpublished data). From these preliminary results we can suggest the presence of homeobox genes similar to those found in *Caenorhabditis elegans*.

Planarians belong to the phylum Plathyhelminthes, a group that separated very early from the rest of metazoans (31). The key phyletic position of the Platyhelminthes and the possible roles of homeobox sequences in developmental and regenerative processes in these organisms may be crucial to an understanding of how certain innovations in development and evolution (e.g., bilateral symmetry, anteroposterior organization, cephalization, etc.) have appeared and been transformed during ontogeny and evolution in the major spiralian groups.

We thank the members of the De Robertis laboratory and Gehring laboratory, especially G. Martin, for their teaching molecular biology to E.S. We also thank T. Bürglin, W. Gehring, and E. M. De Robertis for helpful discussions; W. Gehring, T. Bürglin, and M. Affolter for comments on the manuscript; J. M. Palacios and G. Mengod for oligonucleotide synthesis; and R. Rycroft for checking the English. This work was supported by the grants from the Comisión Asesora de Investigación Científica y Técnica (Ministerio de Educación y Ciencia, España; PB85-0094) and the Comissió Interdepartamental de Recerca i Innovació Tecnològica (Generalitat de Catalunya; AR-87; AR-88). J.G.-F. was a recipient of a FPI fellowship.

1. Gehring, W. J. (1987) *Science* **236**, 1245–1252.
2. Scott, M. P., Tamkun, J. W. & Hartzell, G. W. (1989) *Biochim. Biophys. Acta Rev. Cancer* **989**, 25–48.
3. Bürglin, T. R., Finney, M., Coulson, A. & Ruvkun, G. (1989) *Nature (London)* **341**, 239–243.
4. Otting, G., Qian, Y. Q., Billeter, M., Muller, M., Affolter, M., Gehring, W. J. & Wuthrich, K. (1990) *EMBO J.* **9**, 3085–3092.
5. Affolter, M., Schier, A. & Gehring, W. J. (1990) *Curr. Opin. Cell. Biol.* **2**, 485–495.
6. Baguñà, J. & Romero, R. (1981) *Hydrobiologia* **84**, 181–194.
7. Saló, E. & Baguñà, J. (1984) *J. Embryol. Exp. Morphol.* **83**, 63–80.
8. Saló, E. & Baguñà, J. (1985) *J. Embryol. Exp. Morphol.* **89**, 57–60.
9. Saló, E. & Baguñà, J. (1989) *Development* **107**, 69–76.
10. Collet, J. (1990) Ph.D. thesis (Universitat de Barcelona, Barcelona).
11. Saló, E. & Baguñà, J. (1986) *J. Exp. Zool.* **237**, 129–135.
12. Baguñà, J., Saló, E. & Auladell, M. C. (1989) *Development* **107**, 77–86.
13. Kim, Y. & Nirenberg, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7716–7720.
14. Guazzi, S., Price, M., De Felice, M., Damante, G., Mattei, M. G. & Di Lauro, R. (1990) *EMBO J.* **9**, 3631–3639.
15. Gubler, U. & Holfman, B. J. (1983) *Gene* **25**, 263–269.
16. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
17. Henikoff, J. (1984) *Gene* **28**, 351–359.
18. Deveraux, J., Haeberli, R. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
19. Chirgwin, J. M., Przybyla, A. E., Macdonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
20. Quian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J. & Wüthrich, K. (1989) *Cell* **59**, 573–580.
21. Gehring, W. J., Müller, M., Affolter, M., Percival-Smith, A., Billeter, M., Qian, Y. Q., Otting, G. & Wüthrich, K. (1990) *Trends Genet.* **6**, 323–329.
22. Garcia-Blanco, M. A., Clerc, R. G. & Sharp, P. A. (1989) *Genes Dev.* **3**, 739–745.
23. Doolittle, R. F. (1986) in *Of URFS & ORFS* (University Science Books, Mill Valley, CA).
24. Cohen, S. M., Brönner, G., Küttner, F., Jürgens, G. & Jäckle, H. (1989) *Nature (London)* **338**, 432–434.
25. Treisman, J., Gönczy, P., Vashishtha, M., Harris, E. & Desplan, C. (1989) *Cell* **59**, 553–562.
26. Bürglin, T. R. (1988) *Cell* **53**, 339–340.
27. Fyrberg, E. A., Bond, B. J., Hershey, N. D., Mixter, K. S. & Davidson, N. (1981) *Cell* **24**, 107–116.
28. Dohrmann, C., Azpiazu, N. & Frash, M. (1990) *Genes Dev.* **4**, 2098–2110.
29. Fisher, D. A. & Bode, H. R. (1989) *Gene* **84**, 55–64.
30. Baguñà, J. (1981) *Nature (London)* **290**, 14–15.
31. Field, K. G., Olsen, G. J., Lane, D. J., Giovannoni, S. J., Ghiselin, M. T., Raff, E. C., Pace, N. R. & Raff, R. A. (1988) *Science* **239**, 748–753.