Development 118, 241-253 (1993)
Printed in Great Britain © The Company of Biologists Limited 1993

Genomic organization and expression of the planarian homeobox genes *Dth-1* and *Dth-2*

Jordi Garcia-Fernàndez, Jaume Baguñà and Emili Saló

Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Diagonal 645, 08071 Barcelona, Spain

SUMMARY

We have characterized the genomic organization of *Dth-1* and *Dth-2*, planarian homeobox-containing genes, previously described at the cDNA level (J. Garcia-Fernàndez, J. Baguñà and E. Saló (1991), *Proc. Natl. Acad. Sci. USA*, 88, 7338-7342). Genomic analysis shows that *Dth-1* and *Dth-2* genes encode proteins of 533 and 363 amino acids respectively. The open reading frame of *Dth-1* is interrupted by two large introns of 8 kb and 12 kb *Dth-2* also shows two introns, but these are short (42 bp and 44 bp) and the second interrupts helix III at position 44-45, as is the case with other homeobox genes from such divergent animals as *Drosophila*, honeybee, *C. elegans*, ascidians, and mouse, which suggests an ancient evolutionary relationship between these genes.

The spatial distribution of transcripts in adult tissues, determined by in situ hybridization, demonstrates that *Dth-1* is expressed at a high level in the gastrodermal cells, while *Dth-2* is expressed in the peripheral parenchyma, at higher levels in the dorsal than the ventral regions. Their specific spatial distribution suggests a possible role for these homeobox genes in determination and/or differentiation of specific cell types. The expression pattern of both genes is more or less continuous, but in *Dth-1* clustered discontinuous labelling in areas surrounding the gastrodermis may indicate a specific expression of this gene in groups of undifferentiated cells (neoblasts) already committed or determined to gastrodermal cell fates.

In situ hybridization analysis during early regeneration shows expression only in the postblastema (stump) differentiated areas while no expression has been detected in the undifferentiated blastema, indicating that neither gene has a role in pattern formation mechanisms known to occur at the early stages of regeneration (0-3 days). Hence, *Dth-1* and *Dth-2* are planarian homeobox genes presumably involved in specific cell or tissue determination and/or differentiation.

Key words: planarian, homeobox, gene expression, in situ hybridization

INTRODUCTION

The homeobox is a short and ancient DNA sequence widely distributed on the phylogenetic tree, from plants to vertebrates (Gehring, 1987; Vollbrecht et al., 1991; Shashikant et al., 1991). Such a wide distribution of homeoboxes and their extraordinary degree of sequence conservation suggests a fundamental biological function (Gehring, 1985). From *Drosophila* to mammals most of the homeobox genes have key roles in development; the best known are the *Antennapedia (Antp)* type genes, which are involved in the formation of axes and body segments (Affolter et al., 1990; De Robertis et al., 1991; McGinnis and Krumlauf, 1992). Some homeodomains that are highly divergent from the Antp-type have also been isolated and act as tissue-specific transcription factors that define a cell type (Lazzaro et al., 1991).

We have recently isolated the cDNA of two planarian homeobox-containing genes, *Dth-1* and *Dth-2* (Garcia-Fernàndez et al., 1991). Both contain a novel homeodomain closely related to that of *NK-2* of *Drosophila* (Kim and Niremberg, 1989), *TTF-1* of rat (Guazzi et al., 1991), *Nkx-2*, *3* and *4* of mouse (Price et al., 1992), and *Eghbx-3* of the cestode *Echinococcus granulosus* (Oliver et al., 1992), which may define a new homeobox class.

Since planarians belong to the phylum Platyhelminthes, class Turbellaria, considered to be the sister group of all the other Bilateria (Field et al., 1988; Riutort et al., 1992; Adoutte and Philippe, 1993), the structure of the homeobox genes of planarians should shed some light on the origin and evolution of homeobox genes. Furthermore, to gain understanding of the role of homeobox genes in planarians, two research approaches have been followed. First, gene structure has been determined to define the regulatory sequences that allow comparative intra- and interspecies studies. Second, spatiotemporal expression patterns have been studied to provide insights into their function.

In this report, we describe the genomic sequence including the 5′ and 3′ flanking sequences of two planarian homeobox genes: *Dth-1* and *Dth-2*. Both homeobox genes show tissue-specific expression in the adult and during regeneration. *Dth-1* is detected in the intestinal epithelium, while *Dth-2* is expressed in the peripheral parenchyma, with higher values in dorsal than in ventral areas.

MATERIALS AND METHODS

Species

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

Regenerating organisms

Planarians, 9-10 mm long, were cut postpharyngeally according to Saló and Baguñà (1984) and left to regenerate in Petri dishes with spring water in the dark at 17°C.

Genomic library construction and screening

Standard protocols to extract DNA were not suitable for planarian tissues because of its high nuclease and mucopolysaccharide content. Thus, we had to adapt the guanidine isothiocyanate method initially described for RNA extraction (Chirgwin et al., 1979) and modify it as follows: planarians were homogenized in guanidine thiocyanate, phenol/chloroform treated, and dialized against several changes of 10 mM Tris-HCl pH 8.0, 50 mM EDTA, pH 8.0, 100 mM NaCl. After digestion of RNAs (1 hour at 37°C with 100 μg/ml RNAse A), deproteination (3 hours at 50°C with 200 μg/ml proteinase K in 1% SDS, 0.1 M NaCl) and phenol/chloroform treatment, CsCl was added to a final density of 1.55 g/ml. Samples were then ultracentrifuged for 72 hours at 168,000 ***g*** (rotor Kontron TFT50), and 0.5 ml fractions of the gradient were tested for the presence of DNA. Positive fractions were finally dialized against TE and kept at 4°C.

High relative molecular mass DNA was partially digested with *Sau*3A, then fractionated in a sucrose gradient to obtain DNA fragments of 15-20 kb, which were then dephosphorylated. The genomic library was constructed by cloning this DNA into the *Bam*HI site of λ-Charon35 vector (Wilhelmine and Blattner, 1983). The total number of recombinant clones obtained was approximately 8×10^5. This library was amplified and 2.5×10^6 recombinant clones were screened for sequences homologous to *Dth-1* and *Dth-2* cDNAs (Garcia-Fernàndez et al., 1991) using the method of Benton and Davies (1977). Probes (complete cDNAs) were ^{32}P-labelled by random priming and hybridized to phage DNA on nitrocellulose filters (Schleicher & Schuell) at high stringency.

Restriction analysis and sequencing

Positive clones were characterized by restriction mapping, and selected fragments were subcloned into pBluescript +SK (Stratagene) using standard techniques (Maniatis et al., 1982). Exons were located on the restriction map by hybridization of digested phage DNA with different fragments of the cDNAs. Sequencing was carried out using T7 DNA polymerase (Pharmacia) according to the manufacturer's instructions. Sequencing strategy was performed by unidirectional digestion with exonuclease III (Henikoff, 1987). Some internal primers were also synthesized and used where necessary. All predicted restriction sites were sequenced across to ensure nucleotides were not lost in subcloning. Nucleotide sequence data were analyzed with the Sequence Analysis Software Package of the Genetics Computer Group of the University of Wisconsin (Deveraux et al., 1984).

Primer extension

Primer extension analysis was carried out essentially as described in Ausubel et al. (1987), with primers 5′-end-labelled with T4 kinase at a specific activity of $2\text{-}4\times10^6$ cts/minute per pmol, using [γ-^{32}P]ATP (6000 Ci/mmol, New England Nuclear). Primers used were as follows: OD-2, 5′GAGGATCATCGATTTGG3′, complementary to nt 171-150 of *Dth-1* genomic sequence; ES-4, 5′GGGTGTCTCAGTAGATTGAGTAGTC3′, complementary to nt 110-81 of published *Dth-2* cDNA (Garcia-Fernàndez et al., 1991).

Southern blotting

15 μg of total genomic DNA digested with *Eco*RI or *Hin*dIII was separated on a 0.8% agarose gel and transferred to nylon membranes (Amersham) by capillary blotting. Hybridization was carried out according to the standard procedure of Maniatis et al,. 1982. The high-stringency conditions were exactly as described by Garcia-Fernàndez et al. (1991). Middle stringency hybridizations were carried out with similar buffer to that used for the high-stringency conditions, only the hybridization temperature and formamide concentrations were raised to 40°C, and 46% (vol/vol). Middle-stringency hybridized filters were washed 2×15 minutes at 65°C in 1× SSC, 0.1% SDS.

In situ hybridization

In situ hybridizations were carried out according to Ingham et al. (1985), with several modifications.

Tissue preparation

Intact and regenerating planarians were fixed in 4% paraformaldehyde in PBS (10 mM phosphate buffer pH 7.4, 150 mM NaCl) or 2% paraformaldehyde plus 0.1% glutaraldehyde in PBS for 1 hour at 4°C, and washed twice with PBS for 15 minutes prior to storage in 70% ethanol at 4°C. They were embedded in Paraplast-plus according to Langdale et al. (1987). 8 μm sagittal and transverse sections were cut in an Anglia Scientific 500 microtome and dried on poly-D-lysine precoated slides at 42°C. To obtain frozen sections, planarians were fixed in 4% paraformaldehyde in PBS for 45 minutes at 4°C, washed for 30 minutes in PBS, cryoprotected by incubation in 20% saccharose in PBS for 12 hours and embedded in OCT (Miles Laboratories) at −70°C. 10 μm sections were cut on a Reichert-Jung 2.800 Frigocut E microtome and postfixed in 4% paraformaldehyde.

Probes

The *Dth-1* and *Dth-2* cDNA clones (Garcia-Fernàndez et al., 1991) were cloned in pBluescript SK+ (Stratagene) and used as templates for synthesis of sense and antisense RNA probes following the suppliers instructions (RNA transcription kit, Stratagene), using uridine 5′-(α-[^{35}S]thio)triphosphate (1000 Ci/mmol, Amersham) diluted 1:2 with nonlabelled rUTP. Probe length was reduced to an average of 200 nucleotides by limited alkaline hydrolysis, as described by Cox et al. (1984). Final specific activity was approximately 7×10^9 cts/minute per μg .

Hybridization

After acid treatment (20 minutes in 0.2 M HCl) and neutralization (10 minutes in 2× SSC), sections were deproteinized for 30 minutes at 37°C in 1 μg/ml proteinase K, postfixed in 4% paraformaldehyde, treated for 10 minutes with freshly prepared 0.1% acetic anhydride in 0.1 M triethanolamine buffer (pH 8.0), dehydrated in ethanol, and air dried. Approximately 8×10^6 cpm of ^{35}S-labelled probe was applied in 40 μl of hybridization buffer (50% formamide, 1 μg/ml yeast tRNA, 50 mM DTT, 500 μg/ml polyadenylic acid, 10% dextran sulphate, 0.3 M NaCl, 10 mM Tris-HCl pH 6.8, 10 mM phosphate buffer pH 6.8, 5 mM EDTA). Hybridization was carried out at 50°C for 12 hours, and the slides were then washed for 5 hours in buffer (50% formamide, 10 mM DTT, 0.3 M NaCl, 10 mM Tris-HCl pH 6.8, 10 mM phosphate buffer pH 6.8, 5 mM EDTA), rinsed for 5 minutes at 37°C in NTE solution (0.5 M NaCl, 10 mM Tris-HCl pH 8, 1 mM EDTA, 10 mM DTT), treated for 30 minutes at 37°C with 20 μg/ml

RNaseA in NTE solution, and washed in NTE for 1 hour with several changes of solution. The slides were then incubated in washing buffer for 12 hours at 50°C, dehydrated, air dried coated with Kodak NTB-2 emulsion and exposed for 15 days. After development, and occasional counterstaining with methylene blue, the sections were observed and photographed under a compound microscope equipped with Nomarski interference optics.

RESULTS

Isolation of genomic sequences

An amplified genomic library of *Dugesia (Girardia) tigrina* was screened at high stringency conditions with completed *Dth-1* and *Dth-2* cDNAs as probes (Garcia-Fernàndez et al., 1991). The structure and restriction map of five overlapping lambda DNA clones spanning more than 45 kb for *Dth-1*, and one clone spanning more than 15 kb for *Dth-2* are shown in Fig. 1 and Fig. 2 respectively. We have confirmed that these genomic clones exactly match cDNA sequences previously reported (data not shown).

A Southern blot of *D. tigrina* genomic DNA digested with *Eco*RI or *Hin*dIII was probed under high stringency conditions with a *Dth-1* cDNA fragment (*Dth-1* 5′ probes, nucleotides 1-995 of the cDNA), and with the whole cDNA of *Dth-2*. As shown in Fig. 3A, three strong hybridizing bands were detected in each digestion using *Dth-1* 5′ as a probe (6 kb, 5.5 kb and 4 kb bands in the *Eco*RI digestion, and 10 kb, 9 kb and 6 kb bands in the *Hin*dIII digestion), corresponding to the three exons deduced in the genomic map. A single 5.5 kb band in the *Eco*RI digestion, and 20 kb, 3.5 kb and 2.2 kb bands in the *Hin*dIII digestion were detected in the hybridization with the *Dth-2* cDNA probe (Fig. 3B). The weak extra bands of 7 kb (*Eco*RI) and 20 kb (*Hin*dIII) observed in the *Dth-1* hybridization (Fig. 3A) were considered genomic partial digestions or polymorphic sites. For *Dth-2*, the *Hin*dIII bands of 20 kb and 3.5 kb (Fig. 3B) both cover the 5′ end of the *Dth-2* gene, and are due to a polymorphic *Hin*dIII site produced by the insertion of a *mariner*-like transposon (work in progress) found in some but not all the genomic clones isolated from this region. We conclude that *Dth-1* and *Dth-2* are single copy genes in the *D. tigrina* genome.

To determine whether the *D. tigrina* genome contains other genes possessing a similar homeobox to *Dth-1* and *Dth-2*, Southern genomic hybridization was carried out under middle stringency conditions using the *Dth-1* cDNA homeobox sequence (hb probe Fig. 1). Only the 5.5 kb in the *Eco*RI digestion and the 9.0 kb in *Hin*dIII digestion corresponding to the third exon, deduced in the genomic map, was detected (Fig. 3C), the small weak bands are artefactual, as deduced from the control hybridization with plasmid DNA (results not shown). No extra band corresponding to the *Dth-1* or *Dth-2* homeobox was detected. These results suggest that it is unlikely that a gene containing a similar homeobox to that of *Dth-1* or *Dth-2* is present in the *D. tigrina* genome. The same experiment performed with another planarian species *Dugesia (Schmidtea) mediterranea* at middle stringency shows weak bands of 7 and 5 kb (*Eco*RI) and 5, 3 and 2.5 kb (*Hin*dIII; Fig. 3C) which suggests the presence of similar homeobox genes in *D. mediterranea*.

Structure of the planarian homeobox genes *Dth-1* and *Dth-2*

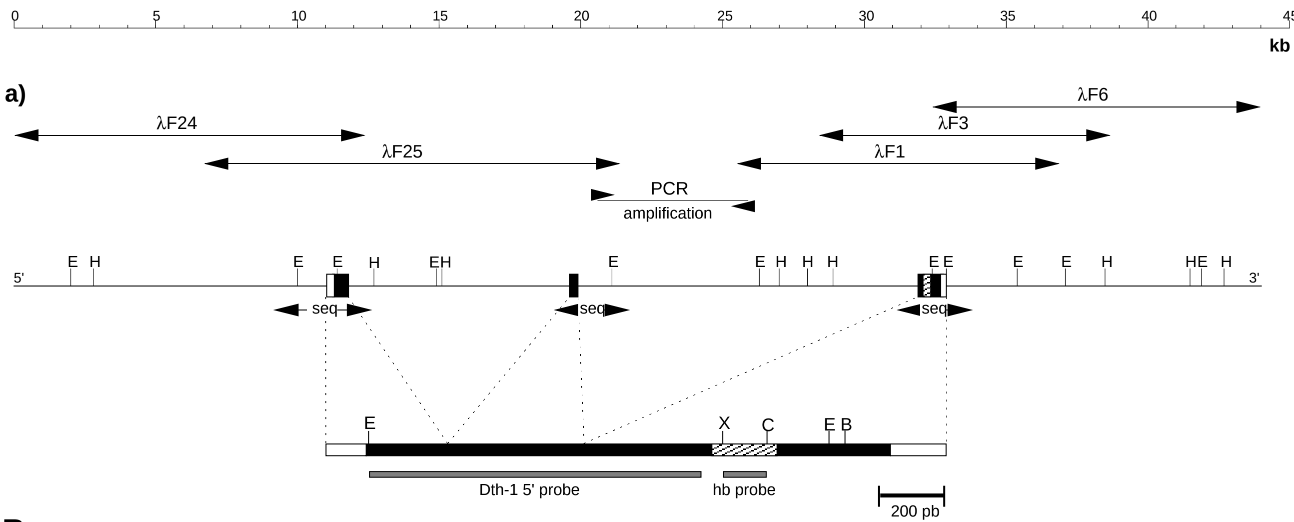
Figs 1 and 2 show the genomic nucleotide sequences of 7.2 kb and 3 kb of the *Dth-1* and *Dth-2* genes, respectively. Taking into account that *Dth-1* and *Dth-2* are the first planarian genes in which the genomic sequences have been characterized, an analysis of the region upstream of the cap site was performed to identify possible promoter and regulatory elements in the DNA. The putative TATA-boxes are not completely adjusted to the consensus (Breathnach and Chambon, 1981) although they are quite similar in the relative position with respect to the startpoint of transcription (Fig. 4A). The CAAT-boxes perfectly match the consensus (Latchman, 1991), the homology between *Dth-1* and *Dth-2* sequences increasing in some extranucleotides (Fig. 4B). The transcriptional startpoint determined by primer extension showed a high degree of conservation compared to the consensus (Breathnach and Chambon, 1981; Hultmark et al., 1986; Fig. 4C). Finally, the exon/intron boundary sequence conservation (Watson et al., 1992) is quite similar. The donor splice junction sequences are extremely conserved, principally in the exon region before the GT consensus, only the short introns show a substitution of G for A at position −1. The acceptor splice junction sequence is conserved, though the sequences immediately preceding this junction are extremely rich in AT rather than CT, and position −3 corresponds to a T (Fig. 4D). *Dth-1* and *Dth-2* genes have a single long open reading frame consisting of 533 and 363 amino acid residues respectively, including a highly diverged homeodomain similar to the *Drosophila NK-2* homeobox.

The *Dth-1* genomic sequence is organized in three exons of 478, 406 and 970 nucleotides in length and two introns of 8 and 12 kb in length located at positions 473 and 800 (nucleotides) of the mRNA sequence respectively. Primer extension analysis using the OD2 antisense oligonucleotide (Fig. 5A) gave one band of 171 nucleotides in length that defines the transcriptional origin, 178 nucleotides upstream from the origin of the longest cDNA described (Garcia-Fernàndez et al., 1991), and a final size of 1.8 kb for the *Dth-1* cDNA.

The *Dth-2* genomic sequence is organized in three exons of 404, 305 and 537 nucleotides in length and two small introns of 42 and 44 nucleotides in length located at positions 405 and 752 (nucleotides) of the transcribed sequence. The second is located after position 44, inside helix III of the homeodomain. Such intron localization has already been described in several other homeobox genes (Mlodzik et al., 1988; Cribbs et al., 1992; Celniker et al., 1989; Cohen et al., 1989; Allen et al., 1991; Kim and Niremberg, 1989; Walldorf et al., 1989; Saiga et al., 1991; Burglin et al., 1989; Hawkins and McGhee, 1990; Schaller et al., 1990). Primer extension analysis using the ES4 antisense oligonucleotide shows two bands of 160 and 164 nucleotides in length which define two closer transcriptional origins, with more frequent use for the short one (Fig. 5B).

The homeodomain of *Dth-1* is located in the third exon, while the *Dth-2* homeodomain is located between exons

A



B

```
-2343                                                                                     AAGTAAACTCGTAGCAATCAATCGTGATTCGTAAACATCGATT
-2300 GCAACTGTTTCTCGTAAATGTGGCTAATTGAATTATTAATCACTGTCTTCCAAATATTGCAGATTACAATTTCACAATACATTAATCGTGATGTATTAAC
-2200 TTGTATTAACTTTATGTTCGAAACAGAACAAATTTACACTTTCAAATATCAAAAGATTTTGATAATTCGTTTTCATGGAACACAGTTAATTCAAATCAAA
-2100 TAATTTAATTATAATCACACAATTTTTACAAAATTTTCAAACCCAATTTCAACTTTTTTTCACATACACAATTATTTTAAATCGATAAGTAATAAACACA
-2000 AGTAATTCAAGTTTTTTTCAATGATTACATTTAGTTAAACCGATTAAAATTTTTTTCATTTCATTGTAATTTTTTTGAATAATAATTTTCTCAAACTATT
-1900 TTTTTAAATTAATACCTTTTTTTAATATTAATTCTTTCATTTTTTTTTTAATTTTTAAATTTTCTTCTTCCAATTTTTTCTTTTTATTCAGAAATTCGAT
-1800 GTCTTTTGCAATGTGTAAAAAGAATAATGTGTGTATCTATACACAGTACAAATGATTGAGTTAACTAAAATCACAAGGAAATCATTAAATCTATGATATA
-1700 GATTAATCATCATATGCCTTTATACATATACATGGTTTGATATTTTTAACAATTGATTCCACTTTAGATTTTCTAGGTTATATATTCCGGCAGTAACTGG
-1600 CAGATTTTTTTAATAATTTAATTTAAAGTTTTTCAAACGAAAGCACTTCCACTTTCCTGGTGGTAGAATTTATTATCGATATCAAAAAATATCGGATAAT
-1500 ATGACTCCCCCGTTATTGATCAAACGAGTCTCTAAAGTTTTTAGTTTCTTAGATTTAACGACTATTTTTATATTTGTACTTGAAATTACGCAGTTTTTAA
-1400 TTAAAAGCCGTAAAAAAGACGATTCAATAAAAAAATCATTTTAAATTTTGAAAATCAAACAAATAAGCATAATTAAATTAAAAAAAAACTCAATTCAGTC
-1300 TCTAAGAATTCATCAGGATTTTCGAGTACATCATTCTTCGTTAGGGACGATTACTGACCAAGAGTTCCATTATCAAAACTCCAGGAGAAAGTGATCGTAT
-1200 TCAGAGAATTTTATTTGTCTTCGGTCTCTTCTGAGTTATTTTTAATGTAATAAATAAAAACATGAAAAGTAGAATTTTAACAAAAAAGGTTGTAATGTAT
-1100 TGTCGAAAACTTATACGTCAGTAGTTATAGATATGAGTATTTTACCAATCGGTTAGATCAAAATTGTCCTCAAATTCGACCCTTGACCTGGTGTGGTTGG
-1000 TGGTTTTTGGCACCTGGTCCTTCTTATTGTCGCCCATCACCTAAATGATCTTTTTCATTAAAAGTTGATATAAAAGAAATACAAATTCAAGTTCATCGTG
 -900 TATTTTACTTCTTTTGTTCTTTTAAAACTGTTTGATAATATTTTATTTTTGTTTATTGTTTCAACATTTCCAAACATTTTTTGCTTGCATAAACGTATGC
 -800 TTTCGTAACAAAAAAGTCTTTTGTAAGTGCTCACTCTTGATAACCAAGTGTCCATATTAGCTCTACCTCAAACTGGAATGTCAACCTCTCCTCAAAGTGG
 -700 AATGACAAAAGAGAAATATTTCATTTAAAAAGAGGTTGTTCAAACTCATTAAAAGAAAAAAAGGAACGATTTTCTTTGATATTTGTTTATTTTATTGTAT
 -600 AAATATTGAAATAAAATTAAATACAAAGGTGATGATAACTGAAAATCAAACTAGTAGTAGCGAACAAAACAAACAAACCAAACCTTAATGAATTTGTCAA
 -500 AATGATATCGGAGTAGAGATTAGATAATGATGTTACCAGAGTTTTTTTCGGTTTCATGACCGAGCATGATGGTTAACGCGCTCGGTAACATTTATCAGAA
 -400 TCTATTCGTTTTTTTTCGTTATCAGACGACCTGCAAAACCTATTTAGTAAACCATCGGTTTTTGTTTTCCTTTTACATTATCAACTCTCTCTGTTCTGTG
 -300 GTTGATCAAAGTTTTCCAATGAAAATTTTCATCAACAATATTTGATGTCAAATCAGCATCTCTATATTAATAAACGATTAGGTTTTAGTTGTTTTTTGGT
 -200 TAATATTGTATGTGTGAATGTATAGAGTGATGTACAAGTGATAAGGACCCCATGATTATTGTGTTACCTGCTTCTCTTATCGTCAATACAACGTCACATC
                                                        CAAT                                TATA
 -100 GCGATATCGACACGATTTGTCAGTTTTATTAAACCCACTTGACCAATAATCATCATTTTTGACATGAAGTTTTGTTAAAATTGAGAAAAACCTCTTCAGT
      trs
    1 ATTGATATCTAAAACTCAAAATATCTGACGACTTTCCAACCTGTAAAGTTCTAAGTGCCCACTTAAAGATGAGTTCAAATGGTGATAGTGTCAAGTACGA
                                                                                   M  S  S  N  G  D  S  V  K  Y  D
                                                                               OD2        cDNA
  101 CACAAATTTCGATAGAGAAGGATATTCAACAGATTCAAGTAATGAATGCCCTAACCAAATCGATGATCCTCTAAATGGAATTCCAAATTCACAAAATTCG
        T  N  F  D  R  E  G  Y  S  T  D  S  S  N  E  C  P  N  Q  I  D  D  P  L  N  G  I  P  N  S  Q  N  S

  201 GAAAATTTTCAGTTATATTCACGAAATAATAATGCAAATGAGTTACATGAAACCTTTCCAACAACAATTCCCATTGCGCAACAAGTGCACCTTAGTAGCT
       E  N  F  Q  L  Y  S  R  N  N  N  A  N  E  L  H  E  T  F  P  T  T  I  P  I  A  Q  Q  V  H  L  S  S

  301 TCAGAAATTCTGGAATGGTTAACTCAAACAGTCGTTCAATTGGTCCCGATGGGTCTATTTACTCCATACAACCCTTCGACCGAGCCTTACTTCAGAGGAA
      F  R  N  S  G  M  V  N  S  N  S  R  S  I  G  P  D  G  S  I  Y  S  I  Q  P  F  D  R  A  L  L  Q  R  N

  401 TTTTATTCCGAATGATGAAATGACGTCAAATATGTCTTTTGTTCAAATGTTAAATTCTGCCAATTATGATCAGGTTTGTCCTTGTCATTAACACTTAATT
        F  I  P  N  D  E  M  T  S  N  M  S  F  V  Q  M  L  N  S  A  N  Y  D  Q

  474 ATTGTTTTTTTTG          intron 1 (8 kb)           TTATCATTTATTTCCTTTTTATTTTGATTTCATTATCTAGATCGAGCCAG
                                                                                                   I  E  P

  484 CATTTTATCAGAACATGCCGGGAATCGGAAATAGTTTCATGATGCAAAGCATTCTAAACCCGAATCCCAACGATCAGGAAATCGACAACCCAATATCAGA
      A  F  Y  Q  N  M  P  G  I  G  N  S  F  M  M  Q  S  I  L  N  P  N  P  N  D  Q  E  I  D  N  P  I  S  D

  584 CAATTCGACCTACGTTCTCGTTCAAGGGGATTCGAATATTCCAACGAATCCCCAATCGAATCCCATCAATCCTACTTACGTTTCATTGGATTTGAATGAC
        N  S  T  Y  V  L  V  Q  G  D  S  N  I  P  T  N  P  Q  S  N  P  I  N  P  T  Y  V  S  L  D  L  N  D
```

Fig. 1

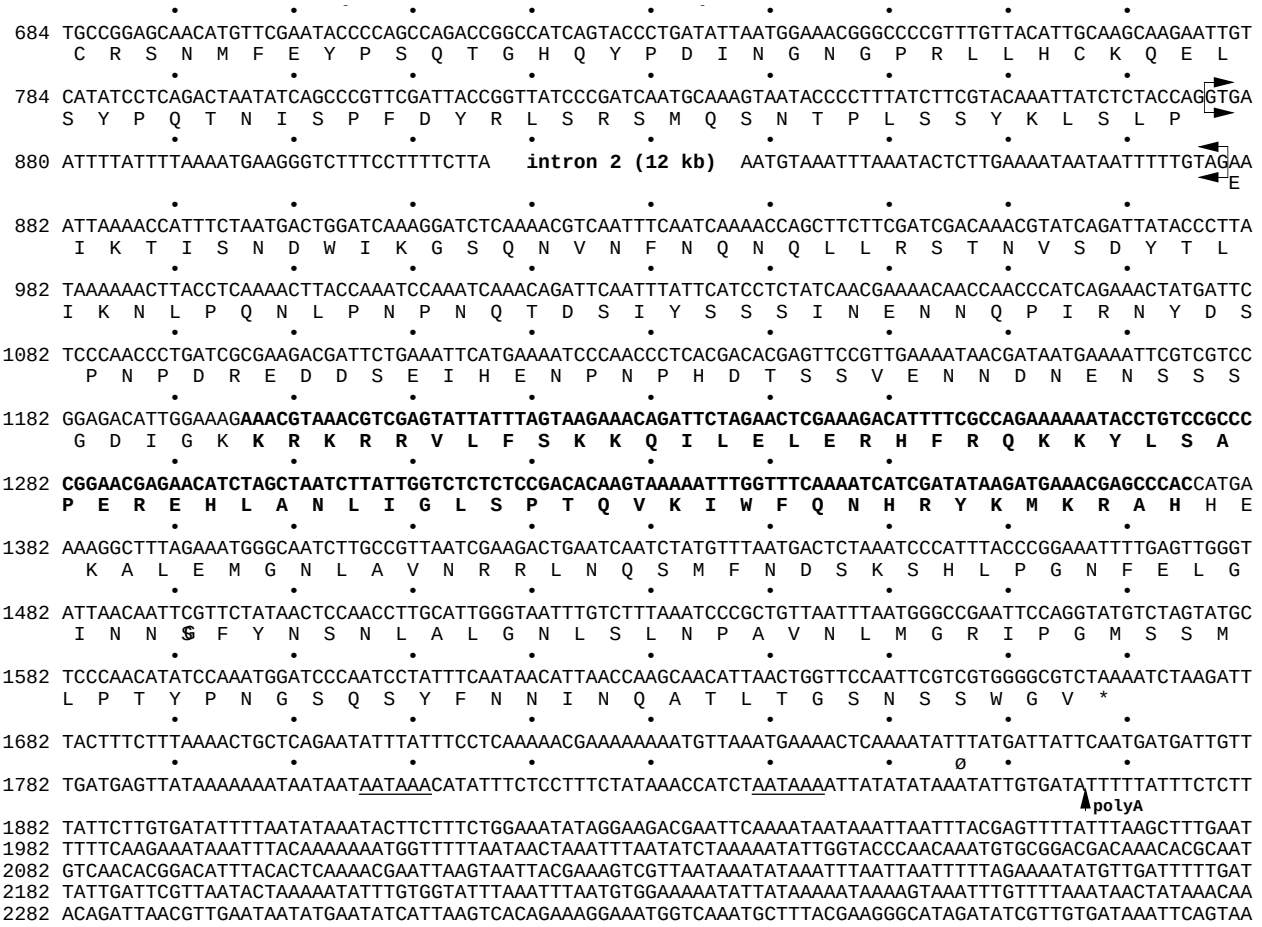


Fig. 1. DNA sequence of the *Dth-1* gene and corresponding amino acid sequence of the putative Dth-1 protein. (A) Restriction map of *Dth-1* genomic clones. Five overlapping clones plus that amplified by PCR to cover the central 4 kb of the second intron are shown at the top. Underneath and enlarged is a 1.8 kb cDNA restriction map. The three exons are boxed; the homeobox is hatched; exon junctions in the cDNA and the sequenced genomic regions are marked. Restriction sites are coded as B, *Bam*HI; C, *Cla*I; E, *Eco*RI; H, *Hin*dIII; X, *Xba*I. The different probes used in this study are also indicated. (B) Sequence of the *Dth-1* genomic DNA illustrated in A. Two nucleotide variations found in one of the sequenced cDNAs are indicated above: a guanine insertion at position 1794, an adenine deletion at position 1854. The amino acid sequence corresponding to the longest open reading frame is shown below the main sequence. The boundaries of an 8 kb and 12 kb intron are indicated by arrowed brackets. The putative CAAT and TATA boxes are boxed and the homeobox sequences are in bold. The polyadenylation signals (AATAAA) are underlined and the position of the OD2 oligonucleotide used for primer extension experiments are overlined. The position of the transcription startpoint (trs); the poly(A)$^+$ site and the 5′ end of the longest cDNA are indicated. The *Dth-1* genomic DNA sequence has been deposited in the GenBank/EMBL database under accession numbers X69203, exon 1; X69200, exon 2; X69201, exon 3.

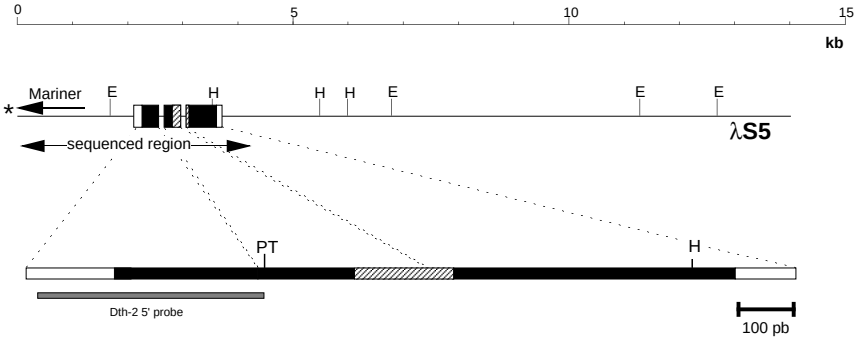
two and three. Homeodomain comparisons of *Dth-1* and *Dth-2* with the different sequences previously described shows the highest similarity with the homeodomains of *Drosophila NK-2* (80-82%; Kim and Niremberg, 1989), the rat *TTF-1* (77-87%; Guazzi et al., 1990) the mouse *Nkx-2.2* (83-82%), *Nkx-2.3* (67-78%) *Nkx-2.4* (77-88%; Price et al., 1992) and the platyhelminth cestode *Echinococcus granulosus Eghbx-3* (65-62%; Oliver et al., 1992). All (Fig. 6) show the same helix III with some unusual positions in all three helices of the homeodomain (Garcia-Fernàndez et al., 1991). In addition, the sequence identity also goes beyond the homeobox. In the amino acid sequences, downstream of the different members of the *NK-2* homeobox family, the identity extends to three additional amino acids, two acidic (positions 62 and 66) and one basic (position 63). Moreover, downstream of *Dth-2* and *TTF-1* homeodomain the identity extends to five additional amino acids (Fig. 6A). At the other end of the homeodomain, the similarity predicted between *Dth-1* and the *NKx2.2* amino acid sequence extends for a further five amino acids (Fig. 6A). All these results taken together would define them as members of a new class of homeobox containing genes that could be named *NK-2*, according to the first gene isolated.

Spatial localization of *Dth-1* and *Dth-2* messenger by in situ hybridization

As an aid to understanding the tissue sections of *D. tigrina*, we have included a brief morphological and histological description of intact and regenerating planarian. A more detailed description can be found elsewhere (Baguñà et al., 1988; Rieger et al., 1991).

Fig. 2. DNA sequence of the *Dth-2* gene and corresponding amino acid sequence of the putative Dth-2 protein. (A) Structure of *Dth-2* genomic and cDNA clones. The three exons are boxed and the homeobox is hatched. Exon junctions in the cDNA are marked. Restriction sites are coded as E, *Eco*RI; H, *Hin*dIII; P, *Pvu*II; and T, *Pst*I. The different probes used in this study, the region sequenced, and the positions of the *mariner*-like element found at the 5′ end are also indicated. (B) Complete genomic sequence of *Dth-2*, and predicted amino acid sequence, deduced by sequence comparison of cDNA and genomic clones illustrated in A. The boundaries of a 42 and a 44 bp intron are indicated by arrowed brackets. The putative CAAT and TATA boxes are boxed and the homeobox sequences are in bold. The polyadenylation signal (AATAGA) and the position of the ES4 oligonucleotide used for primer extension experiments are underlined. The position of the two transcription startpoints (trs); the poly(A)$^+$ site and the 5′ end of the longest cDNA are indicated. *Dth-2* genomic DNA sequence has been deposited in the GenBank/EMBL database under the accesion number X69202.

A



B

```
-1067                                                ACATTATAATGAATATTATATACAGTATACATGTAAAATAAATTAATGCTTTCATAAAAGTATATTTT
-1000 GTAAGTTTCAATGTTAAAATTATCTAGTTGACAGTGTCAACTTGACACTTTTTTCGAGTGAAAAATTGAAAACAATCAAATGAATTAGGAGTTTTACAAT
 -900 TAATGTATTACTAGAACATAGCATTAAATAATAAAATTCAAGTTTTTTAAAGTACAAAGGAGATCGAAAAGTTATAAATATGTATAACATGTATATTAATG
 -800 ATAAAATCAATTAGTTAATGCATAAGTTATTTATTGAGCAATTTATATTCTGATTCAAAATTTACAAATAAATTTACTTGAACAAAATCAAAATTATTTT
 -700 AAAACGTAATTTTAAGTAATCATTACTCTTAAAATTTAGTTTACTTGAAATTGATGAAATTTGAATATTATTGAAATAAATTTTATAATGATAGAAAAAT
 -600 AGTTTTTCCCTTGGATTCTAATATCAAATTAAACAAAACGTAATATTTACATTTCATTATTTAAAATTAACCTAACGAGAAAAATTAAATAGCATTTGAA
 -500 TTCACTAATCGATATTTTCTAGTTATTTAATCGATAGAAAGCATTGACCGCTTACTACTTATCTGGGATTCAACCAATCCATAAACCACTTGAGTCGCAA
 -400 ATTTATTTGAGAATTATTCCGTGGCTCTTCTGCATTTTAATTGGACTACAAGGCAGCAGTAATTATTATGTACTACATTCCCAATTTTGATTCATTATTA
 -300 TGGATAGCAATTTGAAGCATGCTAACCGATGAAATAAGAGTTAGACAACAGTAGTTGTGCATGAAAGACAAAATAATACCTGTGACAAATGTGATGTGTG
 -200 CGTAGTTACGGACTTTCATGGTAAACGTTTTTAAAATATTTCTCACCATGCATAATAATGTATGTAAGGTAGTAGTAAGTACACACACCTAAAGCAACTA
                                                  CAAT                                TATA
 -100 CTGAGTTGAGAAATACAAATAATCAGTGAAGCACGATAAAATCAGCACCAATAAATGCAACCACTTGGTACTTAATAACTCAACTCAGGACTAGATAGTT
      trs trs                                                  cDNA
    1 ATTAATTCAGCTTTGCCTTTAAAGTTGCTACTTAAGTATCTGCGATAGGTATTCACAATCATCTATTTGAATACCACTTAACACACTCTGTAAGAACTTT
                                                  ES4
  101 TATTATCTAAGCTGGTACTAGGAGAAGTACTACTTGAAGACTACTCAATCTACTGAGACACCCATCATTCACAAAGCTGTTATGAGTGGTATTCCTGGAC
                                                                                    M  S  G  I  P  G
  201 TTGCTGGAATGTCCTCAATGTCACCATATTCTGCTTATGCTGCATTAACTCAACATACTGGTGTTAGTTCGGGTTCCTCGCCCTTTGGATCTCAATATTG
      L  A  G  M  S  S  M  S  P  Y  S  A  Y  A  A  L  T  Q  H  T  G  V  S  S  G  S  S  P  F  G  S  Q  Y  C
  301 TTCTAGTGTAAATGACTTCAATCCTTATTCTGATCCAAGAGGAAGTAACACATGGTATGGCATGGCGGCATCAGCCAATGCTAGTAATGATCCGCGTATG
        S  S  V  N  D  F  N  P  Y  S  D  P  R  G  S  N  T  W  Y  G  M  A  A  S  A  N  A  S  N  D  P  R  M
  401 ACAAGTATGGCAAGAATTAACGTTTATAATTTATTGTATTAAATAGTGTCTCGATTAATGGGTTCCGCAGCTGCTGCAGCTAGTAGTAGTATGTCAGCTT
       T                intron 1                  M  S  R  L  M  G  S  A  A  A  A  A  S  S  S  M  S  A
  501 ATTCAGGAATACCCACAAATTTTCATCAAGGAATGCATTCTGCTATGGGCATGGCATCTCTGGGAGCAGCTACGTATGATCATCAAAAAGCTGCAATGCA
      Y  S  G  I  P  T  N  F  H  Q  G  M  H  S  A  M  G  M  A  S  L  G  A  A  T  Y  D  H  Q  K  A  A  M  Q
  601 GTTTAACAATATGTCACAGCGTAGAAAACGAAGAATCTTATTTAGTCAAGCTCAAATTTATGAGCTCGAACGAAGATTTAAACAGCAGAAATATTTATCA
        F  N  N  M  S  Q  R  R  K  R  R  I  L  F  S  Q  A  Q  I  Y  E  L  E  R  R  F  K  Q  Q  K  Y  L  S
  701 GCCCCAGAACGTGAGCATTTGGCTAATTTAATAAATCTAACACCAACACAAGTAATTTAGTCTTTATTTCTGGAACATTTTAATATCAATTTTAGGTAAA
       A  P  E  R  E  H  L  A  N  L  I  N  L  T  P  T  Q                intron 2                V  K
  801 AATATGGTTTCAAAATCATCGTTATAAATGCAAGAGGTCTCAAAAAGACAAAGAAAAAGAGCAACAAAAAGAAAAATCTTATCATTTAAAGAAAAATATA
        I  W  F  Q  N  H  R  Y  K  C  K  R  S  Q  K  D  K  E  K  E  Q  Q  K  E  K  S  Y  H  L  K  K  N  I
  901 GTTGACGATAAAGAACGGTCTCCCAATAAACAAATTTGTAATGCGTCTAGCTCTGATCGAAGTACTCCCGAAGAACCAGTTGCTAAAGCAAAAGAAAGTG
      V  D  D  K  E  R  S  P  N  K  Q  I  C  N  A  S  S  S  D  R  S  T  P  E  E  P  V  A  K  A  K  E  S
 1001 GACTAGATTTTAGTAATCATAAAATTGATAATCTAAATTTAAAAATGGAAGCCGATTTGGAACCAAAGTCTTCTTTATATTCCATAATACCACCATACCT
      G  L  D  F  S  N  H  K  I  D  N  L  N  L  K  M  E  A  D  L  E  P  K  S  S  L  Y  S  I  I  P  P  Y  L
 1101 AACGAACTCTTATGCACAACAAACTCAAAGTGAAGCACAAACGTCTCCTATAATCAATAATGTTCTTGGCTCTAATCTTTTTCCTGAAAGAAAATCGACA
        T  N  S  Y  A  Q  Q  T  Q  S  E  A  Q  T  S  P  I  I  N  N  V  L  G  S  N  L  F  P  E  R  K  S  T
 1201 CCAACAATGGGACCATTAACTAGTTACTCCTTTGGTCAATCAATGGACTCTATTTCAAGCTTTTATTCTCCTGATTTTTCGTTATATAATTGTGCCCATC
       P  T  M  G  P  L  T  S  Y  S  F  G  Q  S  M  D  S  I  S  S  F  Y  S  P  D  F  S  L  Y  N  C  A  H
 1301 CTTATATGGCTGCCAGTTCCTCATATTTTATGAATGCAGCATCAAGACCATGGAATTGAAAGTCACGATGAGAATATTTGAATTTCATATATTGTTAATT
      P  Y  M  A  A  S  S  S  Y  F  M  N  A  A  S  R  P  W  N  *
 1401 TTTTTTATTGTGCAAATAGATTTGATTTCTTTATTTTTCTTTAAATTATTGTTTAAATTCCGCGAATCAGGTAGGTAGCTAAATATATTTAATTGCATAT
                                    polyA
 1501 GAAGATTAATCTAAATGTGGAAAAATGTATGTACATGTGTAATAACGTACAATTAAAATTCAGGCAATTTATTTTATATTTCTAAATTGTGCTGTATGGT
 1601 AGAGAAACATTTTCCTGTCAGACGTTTTACGCACATCAATGTCATACACTGGCGCCGATACGCTAAGTGAGTTTTAAAGATGATACATAGCTCTGTTAAG
 1701 CAGTACTGTTTAACACAAATCAATTAACAATATGGTACACCAACACCTAGTTCACTCCTTAGACTATGAACCTGCTGCATACAAATATTTTTAATATGTC
 1801 AATTATAAAGTACGAGACCATGAAGGCATTAACTTATTCTCTCGCTTATCTTAAACCGTTTCTGGTATGTCTATTGTGAATTATGAAAAATCATCATAAT
 1901 ATTGATAAAGTATGTATAA
```

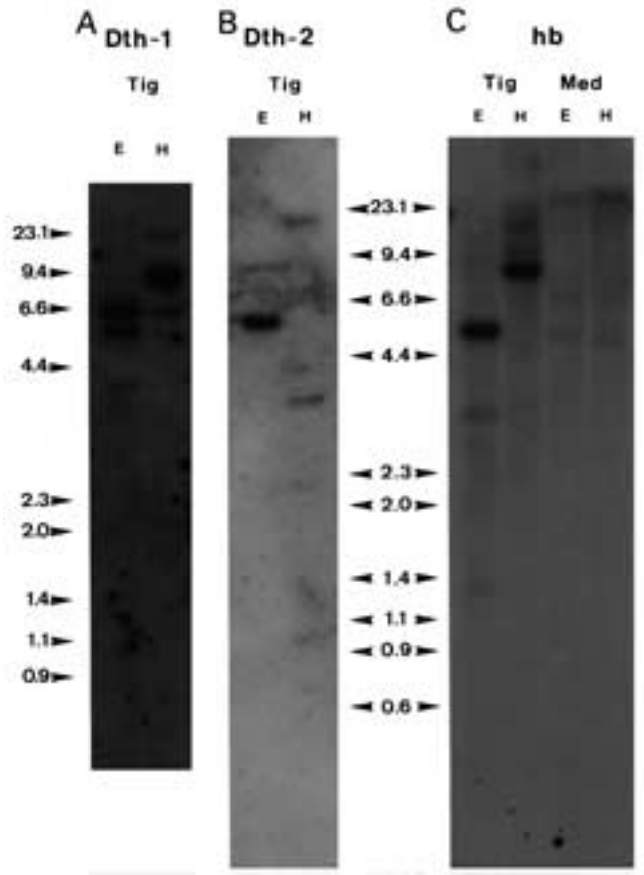


Fig. 3. Genomic Southern blots of *Dugesia (G) tigrina* DNA digested with *Eco*RI(E) or *Hin*dIII(H) and hybridized under high stringency conditions to a (A) *Dth-1* 5′ probe (1-995 nucleotides) and (B) *Dth-2* whole cDNA clone. (C) Genomic Southern blots of *Dugesia (G) tigrina* and *Dugesia (S) mediterranea* digested with the same restriction enzymes as in A and hybridized to a *Dth-1* homeobox-containing fragment of 150 nucleotides (hb probe), under middle stringency conditions.

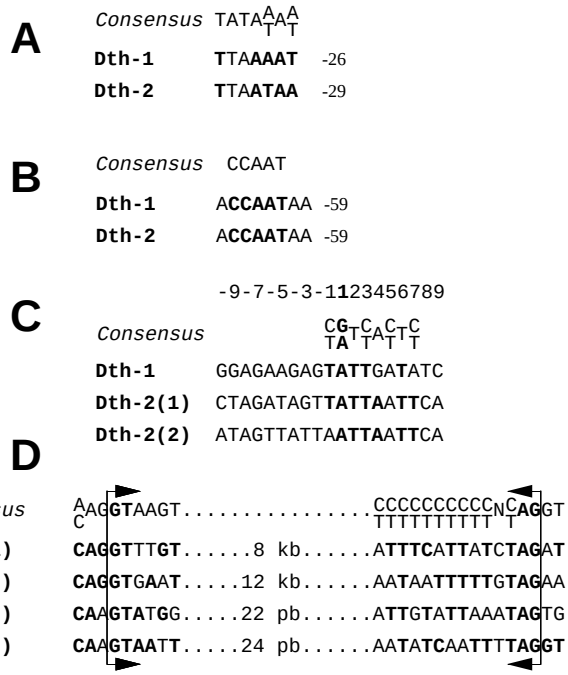


Fig. 4. Comparison of the consensus sequences of the promoter components and the donor and acceptor splice sites of *Dth-1* and *Dth-2* with the consensus sequences (Breathnach and Chambon, 1981; Latchman, 1991; Hultmark et al., 1986; Watson et al., 1992). (A) TATA box; (B) CAAT box; (C) transcription startpoint, and (D) exon/intron boundary sequence.

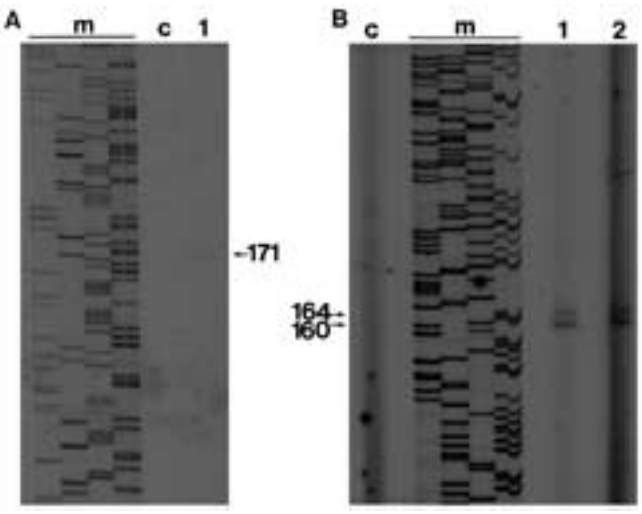


Fig. 5. Mapping the sites of transcription initiation by primer extension. (A) *Dth-1* primer extension using the OD2 oligonucleotide, lane 1: 0.5 µg of poly(A); lane C: control with 25 µg of yeast tRNA. A single band of 171 nucleotides can be observed in the lane 1. (B) *Dth-2* primer extension, lane 1: 25 µg of total RNA and 0.3 pmol of ES4 oligonucleotide; lane 2: 25 µg of total RNA and 0.5 pmol of ES4 oligonucleotide; lane C: control with 25 µg of yeast tRNA and 0.3 pmol of ES4 oligonucleotide. Two bands of 164 and 160 nucleotides not present in the control can be observed. Lane M: sequencing reactions as size markers.

Planarians (Platyhelminthes, Turbellaria, Tricladida) are characterized by being triploblastic, acoelomates, unsegmented and bilaterally symmetrical organisms and by the lack of circulatory, respiratory and skeletal structures. The digestive system consists of a pharynx and a blind, three-branched and diverticulated gut lacking an anus (Fig. 7). The gut (intestine) is made up of two different cell types: the columnar or digestive cells and the secretory goblet cells. A solid mass of tissue, the parenchyma, fills the space between the cellular monostratified ciliated epidermis and the gut, and surrounds the internal organs. The parenchyma consists of several non-proliferating differentiated cell types and a particular class of undifferentiated mitotic cells usually called neoblasts that give rise to all the differentiated cell types of the adult (Baguñà, 1981; Ehlers, 1985; Baguñà et al., 1990). Regeneration involves an early stage (0-2 days) of accumulation of undifferentiated cells (neoblasts) beneath the wound epithelium due to a combination of cell proliferation and cell migration (Saló and Baguñà, 1984, 1989). At 2 days of regeneration this accumulation of neoblasts is visible externally, being called the regenerative blastema. It is during this early period that

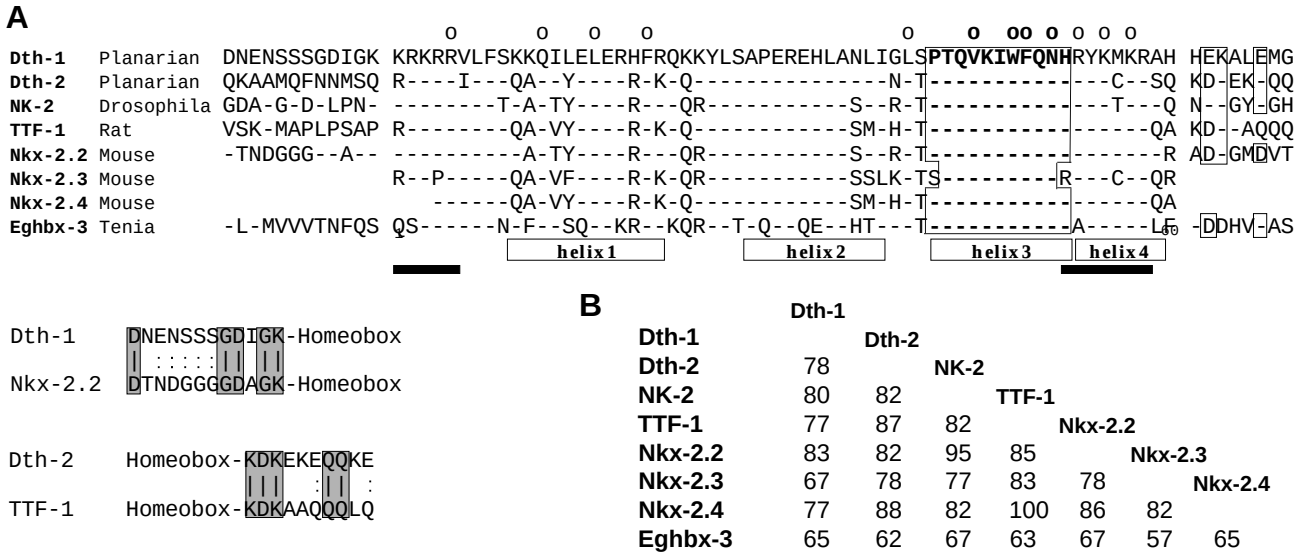


Fig. 6. Alignment and percentage homologies of the amino acid sequences of Dth-1 and Dth-2 with the different members of the NK-2 family: NK-2 of *Drosophila* (Kim and Niremberg, 1989); TTF-1 of rat (Guazzi et al., 1991); Nkx-2, 3 and 4 of mouse (Price et al., 1992); and the cestode *Echinococcus granulosus* Eghbx-3 (Oliver et al., 1992). (A) The amino acid sequences of the homeodomain and flanking regions (single-letter amino acid code). The identical residues are shown by a dash (–) and the helix III and the two basic and one acid consensus residues downstream from the homeobox are boxed. 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 (**O**) are indicated. Underneath the amino acid sequence, homology of upstream sequences from the homeobox between Dth-1 and Nkx-2.2, and downstream from the homeobox between Dth-2 and TTF-1 is represented. (B) Percentage homologies in the homeodomain sequences among all the members of the NK-2 family that include: Dth-1 and Dth-2 (Garcia-Fernàndez et al., 1991), NK-2 (Kim and Niremberg, 1989), TTF-1 (Guazzi et al., 1990), Nkx-2.2, Nkx-2.3, Nkx-2.4 (Price et al., 1992) and Eghbx-3 (Oliver et al., 1992).

determination of new structures along the anteroposterior axis takes place (Saló, 1984; Baguñà et al., 1990). Later on (5-7 days), new structures are differentiated within the blastema and postblastema areas, and the regeneration is complete after 3-4 weeks.

In situ hybridizations to transverse sections, using the antisense strand of the whole cDNA of *Dth-1*, showed abundant silver grains in the intestinal cells, different numbers of intestinal branches being present, depending on the anteroposterior position of the transverse sections (Fig. 7). Sagittal sections show a clear increase in label in the basal region of the intestinal cells (Fig. 9B). No signal was observed in any other tissue or region, even in the pharyngeal epithelium (Figs 7, 9B). A more accurate analysis of grain distribution (Fig. 10A,C) shows a homogeneous expression of *Dth-1* in all intestinal cells and in some clusters of cells surrounding the gut. These clusters of cells could be neoblasts committed or determined to become intestinal cells.

In situ hybridization to transverse sections, using the antisense strand of the complete cDNA of *Dth-2*, showed an almost complementary pattern of labelling to *Dth-1*, with a specific expression in peripheral parenchyma, and higher values in dorsal and lateral regions than in ventral regions (Fig. 8). No signal was detected in other tissues or regions. Sagittal sections showed the increased expression in dorsal and lateral regions more clearly (Fig. 9D). At higher magnification (Fig. 10B,D) non-homogeneous silver grain deposition was observed in the subepidermal parenchyma regions. Since specific organ precursors have not been described in these regions, this pattern of expression could indicate a role of *Dth-2* in determination and/or differentiation of a specific parenchyma cell type.

In situ hybridization analysis performed with both probes in sagittal sections of organisms at early stages of regeneration (Fig. 11) did not show any expression within the blastema, while clear expression was observed in the old differentiated tissues. These results suggest that neither gene has an important role in the process of early pattern formation during regeneration.

To illustrate the specificity of our probes, Figs 8 and 9E show transverse and longitudinal sections hybridized with sense RNA. All the tissues, even the gut and the peripheral parenchyma, display a low level of hybridization. Hence, the control probes do not show any specific binding.

DISCUSSION

We present an analysis of the genomic organization of *Dth-1* and *Dth-2* homeobox genes in the planarian *D. tigrina*. We have also established, for the first time in planarians, a working in situ hybridization technique sensitive enough to monitor the expression of its developmental control genes.

Sequence analysis

The analysis of *Dth-1* and *Dth-2* homeobox-containing genes, the first genomic sequences reported in planarians, confirms the high degree of conservation of the different promoter components and exon/intron boundary sequences. The genomic sequences of the two genes, however, show a striking differences in the size of their introns:

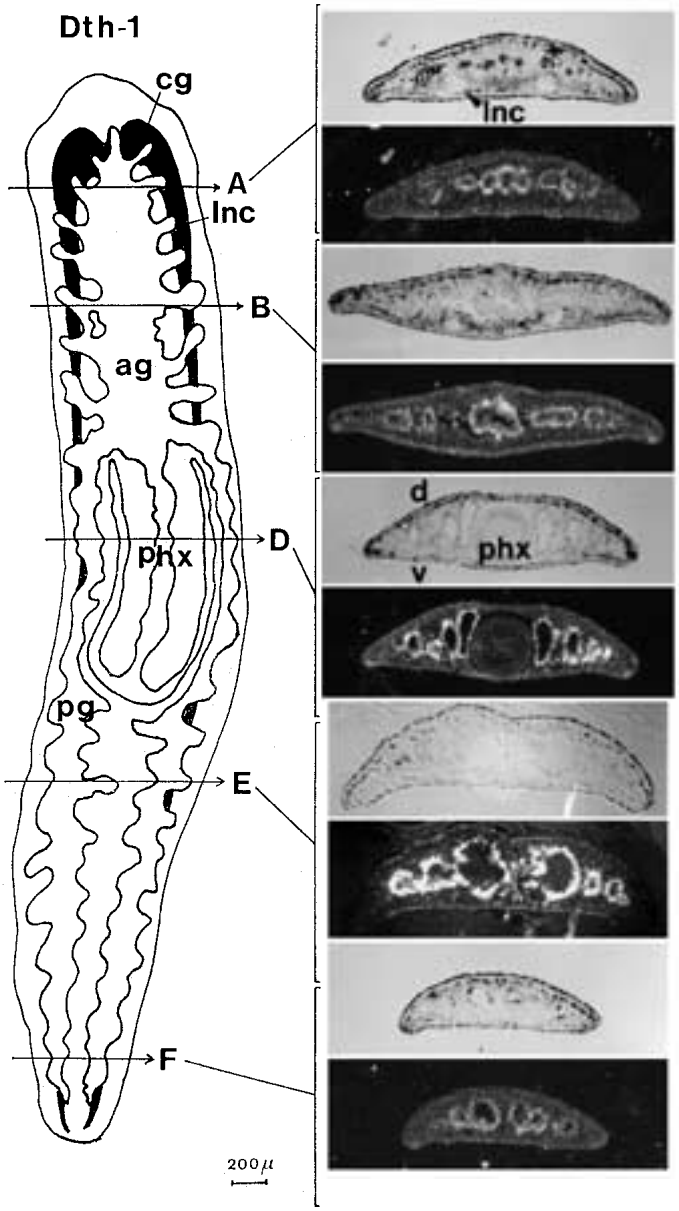


Fig. 7. Pattern of *Dth-1* mRNA accumulation in intact adult planarians. Dark- and bright-field transverse sections hybridized to the *Dth-1* antisense RNA probe. Left schematic representation of a longitudinal section of a planarian showing (in black) the central nervous system (cg, cephalic ganglia; lnc, longitudinal nerve cords), and above it the pharynx and the gut with one anterior (ag) and two posterior (pg) gut branches bearing several diverticulae. (A-F) Transverse sections of the corresponding levels in the scheme on the left. phx, pharnyx; d-v dorsoventral axis. Bar, 200 μm.

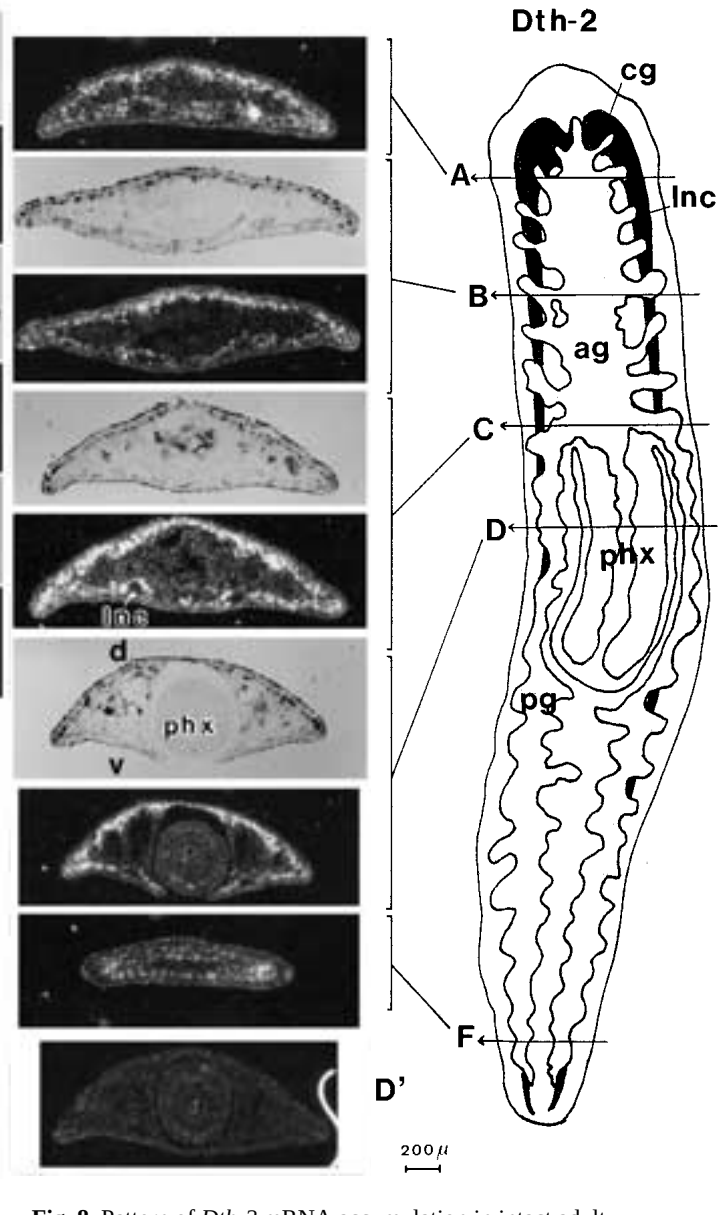


Fig. 8. Pattern of *Dth-2* mRNA accumulation in intact adult planarians. Dark- and bright-field transverse sections hybridized to the *Dth-2* antisense RNA probe. (A-F) Transverse sections of the corresponding levels in the scheme on the right. (D′) same region as D hybridized to the sense RNA probe as a control. Abbreviations as in Fig. 7.

two long introns for *Dth-1* and two short ones for *Dth-2*. No clear difference is observed in the consensus sequences between short and long, in contrast to what was found in *C. elegans* (Fields, 1990). The second intron of *Dth-2* is located in helix III (after amino acid codon 44 of the homeodomain). This intron location, already described in several other homeobox genes of *Drosophila*, honeybee, *C. elegans*, ascidians and mouse, has never been described in lower organisms. This finding may mean that all these genes have maintained this characteristic intron position from a common ancestor and suggests a common ancient origin from an old branch of the homeobox gene evolutionary tree.

The deduced amino acid sequences of Dth-1 and Dth-2 homeodomains are very similar to the *Drosophila* NK-2 (Kim and Niremberg, 1989), the rodent TTF-1 (Guazzi et al., 1990), the mouse Nkx2.2, Nkx2.3, Nkx2.4 (Price et al., 1992) and the cestode *Echinococcus granulosus* Eghbx-3 (Oliver et al., 1992). The high degree of conservation between the different components of this group, despite the phylogenetic distance between planarians and vertebrates

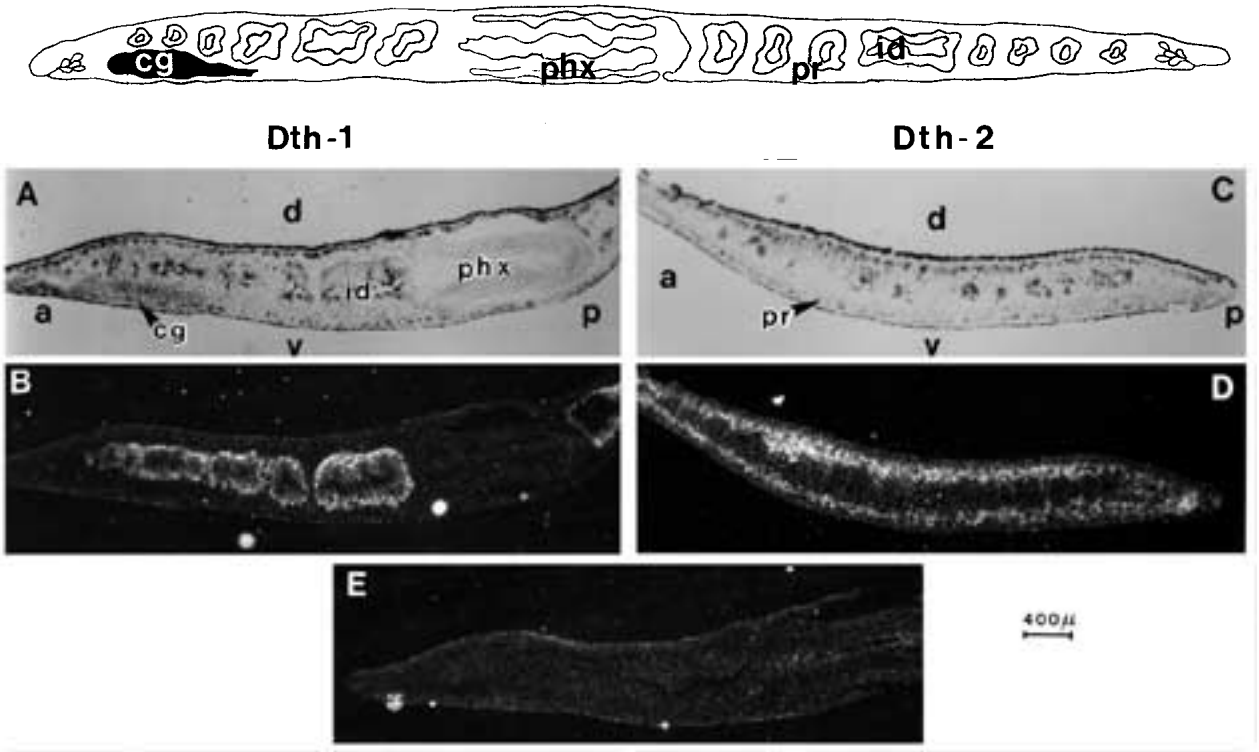


Fig. 9. Localization of the planarian *Dth-1* (A, B) and Dth-2 (C, D) transcripts in sagittal sections of adult planarians. (A) Bright-field and (B) dark-field sagittal sections hybridized to *Dth-1* antisense RNA probe. (C) Bright-field and (D) dark-field sagittal sections hybridized to *Dth-2* antisense RNA probe. (E) Section equivalent to B hybridized to the sense RNA probe as a control. pr, parenchyma; id, intestine diverticula; a-p, anteroposterior axis; d-v, dorsoventral axis. For other abbreviations, see legend of Fig. 7.

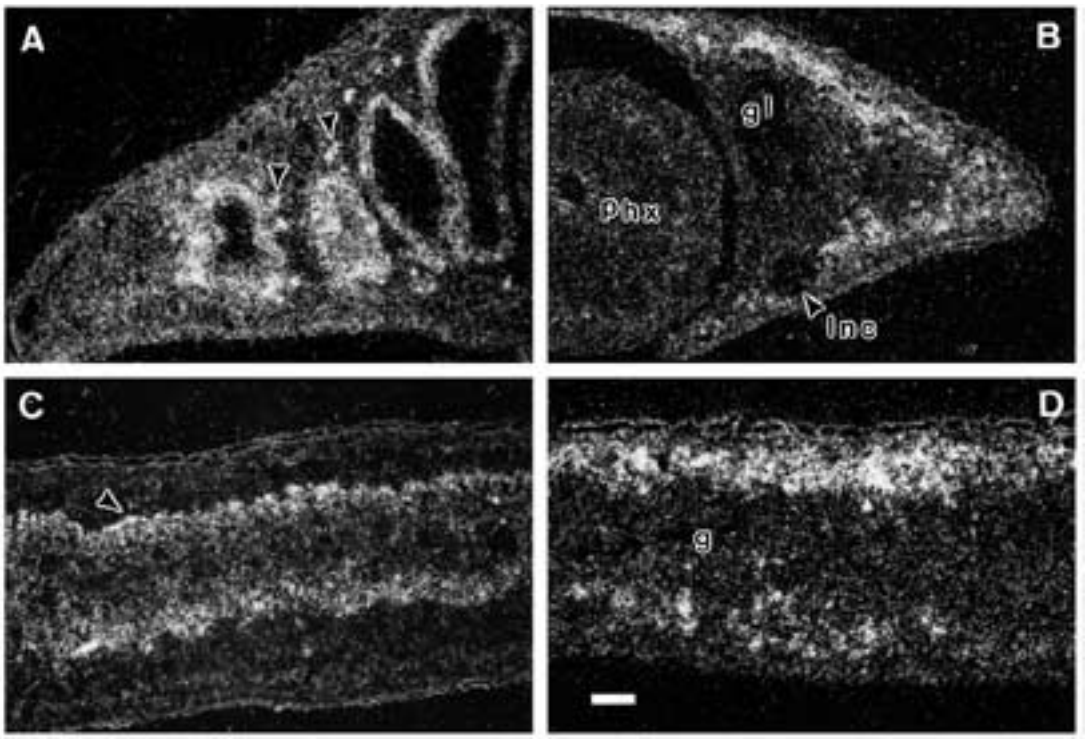


Fig. 10. Dark-field images at high magnification of: (A) transverse sections hybridized to *Dth-1* antisense probe. Clusters of hybridization in areas surrounding the gastrodermis are indicated by arrowheads. (C) Sagittal sections hybridized to *Dth-1* antisense probe. Stronger expression in the basal portion of the intestinal cells is indicated by an arrowhead. (B) Transverse and (D) sagittal sections hybridized to *Dth-2* antisense probe, the lack of hybridization of *Dth-2* probe to the longitudinal nerve cord is indicated by an arrowhead. g, gut; gl, gut lumen. Bar, 50 μm.

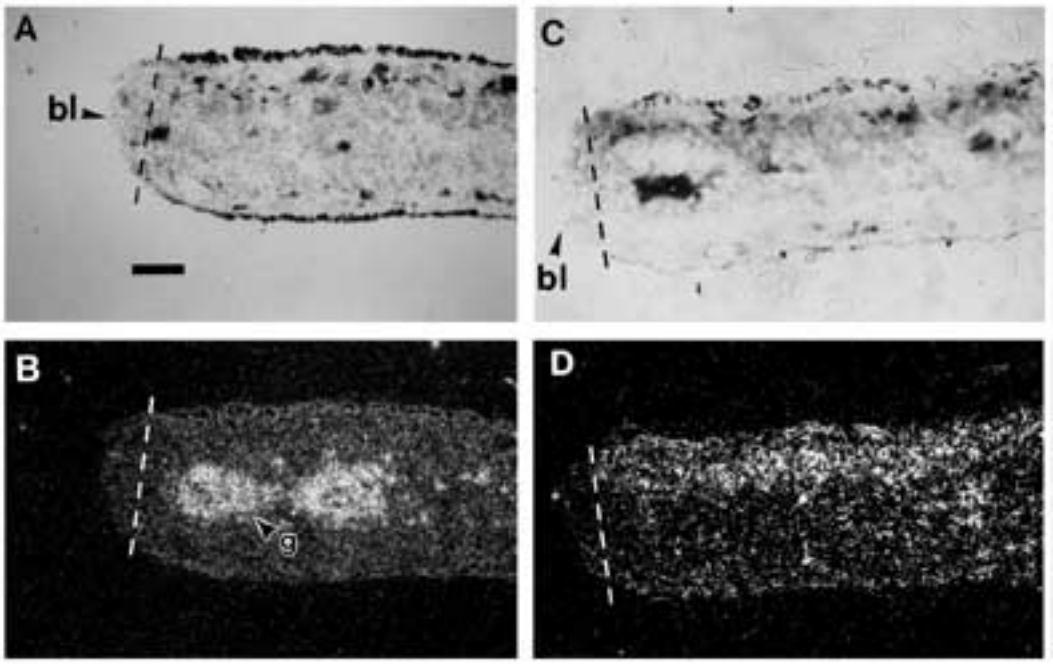


Fig. 11. In situ hybridization of sagittal sections of early (3 days) anterior regenerating organisms. (A, C), Bright-field images stained with methylene blue for morphological examination. bl, blastema. (B) Dark-field image of the section in A hybridized to the *Dth-1* antisense RNA probe. (D) Dark-field image of the section in C hybridized to the *Dth-2* antisense RNA probe. g, gut. Broken lines indicates the limits between blastema and stump regions. Bar, 100 μm.

and the low similarity to the other homeobox classes including Antennapedia, suggests that they may be considered as a new class of ancient homeobox-containing genes, the NK-2 class. In addition, the members of this new class share an identical recognition helix, with the same residue in position 50 and some specific replacements scattered throughout the homeodomain, which may confer a specific consensus recognition sequence other than the Antennapedia, as has been demonstrated for TTF-1, Nkx-2.2, and Nkx-2.3 (Price et al., 1992).

Temporal expression and tissue specificity of *Dth-1* and *Dth-2* transcripts

Large amounts of *Dth-1* mRNA are found throughout the gastrodermis, in intact adult planarians. These results suggest that *Dth-1* expression is associated with differentiation of the endodermal tissues. Several homeobox-containing genes have been described, whose pattern of expression in the embryo, or in some cases in the adult, involves cells or tissues of endodermal origin (Mlodzik et al., 1985; Mlodzik et al., 1988; Saiga et al., 1992; Wang et al., 1990; Rosa, 1989; Wright et al., 1988; Frumkin et al., 1991; Duprey et al., 1988). None shows any clear sequence similarity to the *Dth-1* homeobox. Furthermore, no member of the new NK-2 family shows any specific expression in the endoderm derivatives, and *Dth-1* is thus the only component of the NK-2 family with endoderm-specific expression.

Additional studies also show the expression of *Dth-1* in clusters of undifferentiated parenchyma cells close to the gut in the adult. We postulate that this expression could be associated with determination or commitment of these cells to intestinal cells. Indeed, intestinal cells are short-lived cells (approx. 40-60 days; Baguñà and Romero, 1981) and are continuously being replaced by differentiating neoblasts, as has been extensively documented by light and electron microscopy (Hori, 1986).

Expression of *Dth-2* is restricted to the peripheral parenchyma of the adults with more in the dorsal than the ventral areas. These results suggest that *Dth-2* expression may be associated with differentiation of some parenchymal cell type (or types), widely distributed in the subepidermal region of the whole organism with a higher density in dorsal regions and dorsoventral edges than in ventral regions. Despite several studies at the light and electron microscope level (reviewed by Rieger et al., 1991), parenchyma's cellular components and basic architecture are poorly understood since it is a loosely organized tissue composed of several cell types with numerous interdigitating cell processes that blurr cell limits. In addition, monoclonal antibody staining of parenchyma cells in *D. tigrina* show a pattern that is bewildering in its complexity (Romero et al., 1991), which may reflect an unexpected functional heterogeneity among parenchymal cells or transient different functional stages in single cell types. Although identification of the cells that express *Dth-2* must await the development of polyclonal antibodies, the general, discontinuous, patchy, distribution of *Dth-2* expression suggests that more than one cell type may be labelled. If this is so, *Dth-2* would participate in the determination and differentiation of a variety of mesodermal cell lineages.

Do *Dth-1* and *Dth-2* have any role during regeneration?

In situ hybridization studies at early stages of regeneration (0-3 days; Fig. 11) revealed that *Dth-1* and *Dth-2* are not

expressed in the blastema. Since new structures are determined along the anteroposterior axis during this period (Saló, 1984), we can tentatively conclude that neither gene is involved in early patterning processes. Besides, northern blot analyses also show that at 1-2 days of regeneration *Dth-1* and *Dth-2* have similar levels of expression to those of intact control organisms (García-Fernàndez et al., 1991).

Light and electron microscope studies of gut formation during regeneration (Hori, 1986) indicate that at 3-4 days small clusters of neoblasts gather close to the basal surface of the healed gastrodermis and begin differentiation. Between 5 and 7 days the process is at its maximum, and by 11-15 days the regenerating gastrodermis has already been formed within blastema and postblastema (stump) areas. Northern blot analyses at 5-7 days (García-Fernàndez et al., 1991) show a slight, though significant, increase in the expression of both genes. However, the actual variation in expression should be higher, as poly(A)$^+$ RNA used for northern blots was obtained from the whole regenerant and not from the 1-2 mm area embracing the blastema and postblastema where most processes occur. Indeed, in situ hybridizations of late regenerants (7-13 days) revealed the presence of *Dth-1*-positive gastrodermal cells and *Dth-2*-positive peripheral parenchyma cells in both blastema and postblastema areas (unpublished data).

These results strongly suggest that during regeneration neither gene is involved in early patterning events, although both are involved in late processes of cell determination and differentiation.

Implications of *Dth-1* expression for growth/degrowth and embryonic development

Intact adult planarians are widely known for their ability to grow and degrow in body length and volume depending on temperature and feeding conditions (Baguñà and Romero, 1981; reviewed by Baguñà et al., 1990). These processes, fairly well understood at the kinetic level, may alternate back and forth without apparent impairment of the individual. One of the main consequences of growth and degrowth is that the number of iterated structures along the anteroposterior axis, such as the nerve ganglia, commissures and lateral nerves in the central nervous system, as well as the gut diverticula, increase and decrease respectively. How these changes are brought about has never been studied in detail, and competing hypothesis are entertained (see Baguñà et al., 1990, for a discussion on mechanisms). Light microscopy data suggest that during embryonic development, gut diverticula do not arise by bulging from the simple, saccular, embryonic endodermal epithelium but from the clustering of embryonic undifferentiated cells, similar to neoblasts, which later connect to the gut through differential growth and migration (Le Moigne, 1963).

Whether or not a similar mechanism operates during adult growth is an open question. In this context, the endodermal (gut)-specific expression of *Dth-1* may be highly relevant. *Dth-1* probes or antibodies may be used to monitor their expression in the embryonic gut and diverticula and validate the model suggested. They can also be used to analyse how new diverticula arise during growth and, even, how they regress during degrowth. Moreover, although embryonic development in freshwater planarians is highly modified (reviewed by Baguñà and Boyer, 1990), Dth-1 antibodies could also be used to detect the earliest stages and tissues in which this gene is expressed. Finally, *Dth-1* could be used as a probe to isolate the homologous gene in other turbellarian orders, such as Polycladida, Macrostomida, Acoela and Catenulida, which have a cannonical spiralian development. The expression of this gene in these organisms and meaningful molecular embryonic comparisons could be determined.

The authors thanks Drs W. Gehring, E.M. De Robertis and T. Bürglin for helpful dicussions; J. Casanova and M. Corominas for critical reading of the manuscript; L. Ruiz-Avila for technical advice with in situ hybridization; and R. Rycroft for checking the English. We warmly acknowledge the comments of anonymous referees which greatly helped to improve the manuscript. This work was supported by grants from the Dirección General de Investigación Científica y Técnica (Ministerio de Educación y Ciencia, España; DGICYT PB89-0249) and the Comissió Interdepartamental de Recerca i Innovació Tecnològica (Generalitat de Catalunya; CIRIT AR-88; AR-91).

REFERENCES

Adoutte, A. and Philippe, H. (1993). The major lines of metazoan evolution: Summary of traditional evidence and lessons from ribosomal RNA sequence analysis. In *Comparative Molecular Neurobiology* (ed. Y. Pichon), pp. 1-34. Basel, Switzerland: Birhäuser Verlag.

Affolter, M., Schier, A. and Gehring, W. J. (1990). Homeodomain proteins and the regulation of gene expression. *Current Biol.*, **2**, 485-495.

Allen, J. D., Lints, T., Jenkins, N. A., Copeland, N. G., Strasser, A., Harvey, R. P. and Adams, J. M. (1991). Novel murine homeobox gene on chromosome 1 expressed in specific hematopoietic lineages and during embryogenesis. *Genes Dev.*, **5**, 509-520.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1987). *Current Protocols in Molecular Biology*. Greene Publishing Associates and Willey-Interscience.

Baguñà, J. (1981). Planarian neoblasts. *Nature*, **290**, 14-15.

Baguñà, J. and Romero, R. (1981). Quantitative analysis of cell types during growth, degrowth and regeneration in the planarians *Dugesia mediterranea* and *Dugesia tigrina*. *Hydrobiologia*, **84**, 181-194.

Baguñà, J., Saló, E., Collet, J., Auladell, C. and Ribas, M. (1988). Cellular, molecular and genetic approaches to regeneration and pattern formation in planarians. *Fortschr. Zool*., **36**, 65-78.

Baguñà, J., Romero, R., Saló, E., Collet, J., Auladell, C., Ribas, M., Riutort, M., Garcia-Fernàndez, J., Burgaya, F. and Bueno, D. (1990). Growth, degrowth and regenerations as developmental phenomena in adult freshwater planarians. In *Experimental Embryology in Aquatic Plants and Animals* (ed. H. J. Marthy), pp. 129-162. New York: Plenum Press.

Baguñà, J. and Boyer, B. C. (1990). Descriptive and experimental embryology of the Turbellaria: present knowledge, open questions and future trends. In: *Experimental Embryology in Aquatic Plants and Animals* (ed. H. J Marthy), pp. 95-128. New York: Plenum Press.

Benton, W. D. and Davis, R. W. (1977). Screening lambda-gt recombinant clones by hybridization to single plaques *in situ*. *Science*, **196**, 180-182.

Breathnach, R. and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. *Ann. Rev. Biochem.*, **50**, 349-383.

Bürglin, T. R., Finney, M., Coulson, A. and Ruvkun, G. (1989). *Caenorhabditis elegans* has scores of homeobox-containing genes. *Nature*, **341**, 239-243.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. and Rutter, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry*, **18**, 5294-5299.

Cohen, S. M., Brönner, G., Küttner, F., Jürgens, G. and Jäckle, H. (1989). *Distal-less* encodes a homeodomain protein required for limb development in *Drosophila*. *Nature*, **338**, 432-434.

Celniker, S. E., Keellan, D. J. and Lewis, E. B. (1989). The molecular genetics of the bithorax complex of *Drosophila*: characterization of the products of the *Abdominal-B* domain. *Genes Dev.*, **3**, 1424-1436.

Cox, K. H., DeLeon, D. V., Angerer, L. M. and Angerer, R. C. (1984). Detection of mRNAs in sea urchin embryos by *in situ* hybridization using asymmetric RNA probes. *Devl. Biol.*, **101**, 485-502.

Cribbs, D. L., Pultz, M. A., Johnson, D., Mazulla, M. and Kaufman, T. C. (1992). Structural complexity and evolutionary conservation of the *Drosophila* homeotic gene proboscipedia. *EMBO J.*, **11**, 1437-1449.

De Robertis, E. M., Morita, E. A. and Cho, K. W. Y. (1991). Gradient fields and homeobox genes. *Development*, **112**, 669-678.

Deveraux, J. P., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.*, **12**, 387-395.

Duprey, P., Chowdhury, K., Dressler, G. R., Balling, R., Simon, D., Guenet, J. L. and Gruss, P. (1988). A mouse gene homologous to the *Drosophila* gene *caudal* is expressed in epithelial cells from the embryonic intestine. *Genes Dev.*, **2**, 1647-1654.

Ehlers, U. (1985). Das Phylogenetische System der Plathelminthes. Stuttgart: Gustav Fischer Verlag.

Field, K. G., Olsen, G. J., Lane, D. J., Giovannoni, S. J., Ghiselin, M. T., Raff, E. C., Pace, N. R. and Raff, R. A. (1988). Molecular phylogeny of the animal kingdom. *Science*, **239**, 748-753.

Fields, C. (1990). Information content of *Caenorhabditis elegans* splice site sequences varies with intron length. *Nucl. Acids Res.* **18**, 1509-1512.

Frumkin, A., Rangini, Z., Ben-Yehuda, A., Gruenbaum, Y. and Fainsod, A. (1991). A chicken *caudal* homologue, *Chox-cad*, is expressed in the epiblast with posterior localization and in the early endodermal lineage. *Development*, **112**, 207-219.

Garcia-Fernàndez, J., Baguñà, J. and Saló, E. (1991). Planarian homeobox genes: Cloning, sequence analysis, and expression. *Proc. Natl. Acad. Sci. USA*, **88**, 7338-7342.

Gehring, W. J. (1985). The homeobox: a key to the understanding of development? *Cell*, **40**, 35.

Gehring, W. J. (1987). Homeoboxes in the study of development. *Science*, **236**, 1245-1252.

Guazzi, S., Price, M., De Felice, M., Damante, G., Mattei, M. G. and Di Lauro, R. (1990). Thyroid nuclear factor 1 (TTF-1) contains a homeodomain and displays a novel DNA binding specificity. *EMBO J.*, **9**, 3631-3639.

Hawkins, N. C. and McGhee, J. D. (1990). Homeobox containing genes in the nematode *Caenorhabditis elegans. Nucl. Acids Res*. **18**, 6101-6106.

Henikoff, S. (1987). Unidirectional digestion with Exonuclease III in DNA sequence analysis. *Methods Enzymol.*, **155**, 156-165.

Hori, I. (1986). Ultrastructural observations on gastrodermal regeneration in the planarian *Dugesia japonica* (Turbellaria). *Hydrobiologia* **132**, 217-222.

Hultmark, D., Klemenz, R. and Gehring, W. J. (1986). Translational and transcriptional control elements in the untranslated leader of the heat-shock gene *hsp-22*. *Cell*, **44**, 429-438.

Ingham, P., Howard, K. R. and Ish-Horowicz, D. (1985). Transcription pattern of the *Drosophila* segmentation gene *hairy. Nature*, **318**, 439-445.

Kim, Y. and Nirenberg, M. (1989). *Drosophila* NK-homeobox genes. *Proc. Natl. Acad. Sci. USA*, **86**, 7716-7720.

Langdale, J. A., Metzler, M. C. and Nelson, T. (1987). The *argentia* mutation delays normal development of photosynthetic cell-types in *Zea mays*, *Dev. Biol.* **122,** 243-255.

Latchman, D. S. (1991). Eukaryotic transcription factors. London: Academic Press Limited.

Lazzaro, D., Price, M., De Felice, M. and Di Lauro, R. (1991). The transcription factor TTF-1 is expressed at the onset of thyroid and lung morphogenesis and in restricted regions of the foetal brain. *Development*, **113**, 1093-1104.

Le Moigne, A. (1963). Etude du développement embryonnaire de *Polycelis nigra* (Turbellarié, Triclade). *Bull. Soc. Zool. Fr*. **88**, 403-422.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual*, 1st edition. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

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

Mlodzik, M., Fjose, A. and Gehring, W. J. (1985). Isolation of *caudal*, a *Drosophila* homeo box-containing gene with maternal expression, whose transcripts form a concentration gradient at the pre-blastoderm stage. *EMBO J.*, **11**, 2961-2969.

Mlodzik, M., Fjose, A. and Gehring, W. J. (1988). Molecular structure and spatial expression of a homeobox gene from the *labial* region of the Antennapedia complex. *EMBO J.*, **7**, 2569-2578.

Oliver, G., Vispo, M., Mailhos, A., Martinez, C., Sosa-Pineda, B., Fielitz, W. and Ehrlich, R. (1992). Homeobox-containing genes in the parasitic flatworm *Echinococcus granulosus. Gene*, **121**, 337-342.

Price, M., Lazzaro, D., Pohl, T., Mattei, M. G., Rüther, U., Olivo, J. C., Duboule, D. and Di Lauro, R. (1992). Regional expression of the homeobox gene *Nkx-2. 2* in the developing mammalian forebrain. *Neuron*, **8**, 241-255.

Rieger, R., Tyler, S., Smith III, J. P. S. and Rieger, G. E. (1991). Plathyelminthes: Turbellaria. In *Microscopic Anatomy of Invertebrates*. **3** (ed. F. W Harrison and B. J. Bogitsh), pp. 7-140. New York: Wiley-Liss, Inc.

Riutort, M., Field, K. G., Turbeville, J. M., Raff, R. A. and Baguñà, J. (1992). Enzyme electrophoresis, 18S rRNA sequences, and levels of phylogenetic resolution among several species of freshwater planarians (Platyhelminthes, Tricladida, Paludicola). *Can. J. Zool.*, **70**, 1425-1439.

Romero, R., Fibla, J., Bueno, D., Sumoy, L., Soriano, M. A. and Baguñà, J. (1991). Monoclonal antibodies as markers of specific cell types and regional antigens in the freshwater planarian *Dugesia(G.) tigrina*. *Hydrobiologia*, **227**: 73-79.

Rosa, F. M. (1989). *Mix. 1*, a homeobox mRNA inducible by mesoderm inducers, is expressed mostly in the presumtive endodermal cells of Xenopus embryos. *Cell*, **57**, 965-974.

Saiga, H., Mizokami, A., Makabe, K. W., Satoh, N. and Mita, T. (1991). Molecular cloning and expression of a novel homeobox gene AHox1 of the ascidian, *Halocynthia roretzi*. *Development* **111**, 821-828.

Saló, E. (1984). Formació del blastema i re-especificació del patró durant la regeneració de les planàries *Dugesia (S) mediterranea i Dugesia (G) tigrina*: anàlisi morfologica, cellular i bioquímica, Ph. D. Thesis, University of Barcelona.

Saló, E. and Baguñà, J. (1984). Regeneration and pattern formation in planarians. I. The pattern of mitosis in anterior and posterior regeneration in *Dugesia(G)tigrina*, and a new proposal for blastema formation. *J. Embryol. exp. Morph.* **83,** 63-80.

Saló, E. and Baguñà, J. (1989). Regeneration and pattern formation in planarians. II. Local origin and role of cell movements in blastema formation. *Development* **107**, 69-76.

Schaller, D., Wittmann, C., Spicher, A., Müller, F. and Tobler, H. (1990). Cloning and analysis of three new homeobox genes from the nematode *Caenorhabditis elegans. Nucl. Acids Res*. **18**, 2033-2036.

Shashikant, C. S., Utset, M. F., Violette, S. M., Wise, T. L., Einat, P., Einat, M., Pendleton, J. W., Schughart, K. and Ruddle, F. H. (1991). Homeobox genes in mouse development. In: *Critical Reviews in Eukaryotic Gene Expression*. **1**, 207-245.

Vollbrecht, E., Veit, B., Sinha, N. and Hake, S. (1991). The developmental gene *Knotted-1* is a member of a maize homeobox gene family. *Nature*, **350**, 241-243.

Walldorf, U., Fleig, R. and Gehring, W. J. (1989). Comparison of homeobox-containing genes of the honeybee and *Drosophila. Proc. Natl. Acad. Sci. USA*, **86**, 9971-9975.

Wang, G. V. L., Dolecki, G. J., Carlos, R. and Humphreys, T. (1990). Characterization and expression of two sea urchin homeobox gene sequences. *Dev. Gen.* **11**, 77-87.

Watson, J. D., Gilman, M., Witkoswki, J. and Zoller, M. (1992). *Recombinant DNA*, 2nd edition. New York: Scientific American Books.

Wilhelmine, A. M. L. and Blattner, F. R. (1983). Lambda Charon vectors (Ch32, 33, 34 and 35) adapted for DNA cloning in recombination-defective hosts. *Gene* **26**, 171-179.

Wright, C. V. E., Schnegelsberg, P. and De Robertis, E. M. (1988). XLHbox 8: a novel *Xenopus* homeo protein restricted to a narrow band of endoderm, *Development* **104**, 787-794.

(Accepted 16 February 1993)