Title
A homeotic gene cluster patterns the anteroposterior body axis of the nematode C. elegans

Author
Wang, Bruce Boe-Ang

Publication Date
1993

Peer reviewed|Thesis/dissertation

A Homeotic Gene Cluster Patterns the Anteroposterior Body Axis of the Nematode *C. elegans*

by

Bruce Boe-Ang Wang

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Biochemistry

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco

Dedication

To my family

Preface

I thank all Kenyon lab members past and present for scientific assistance and guidance which eventually led to the completion of this thesis. Of those that deserve particular mention are Steve, who graciously spent a great deal of his time identifying and photographing many of the *lacZ* staining cells in the cluster paper, and who suggested many experiments I did and should have done; Lisa, who taught me a great deal of molecular biology; Steve and Craig who helped in the homeobox sequence comparisons; and Michael who helped in assembling all the cluster paper figures.

Although learning about and doing science was an enjoyable experience, it was only a part of my graduate life that I value. The other significant experiences were those fun and personal ones that I gained through friendship. Some thoughts that immediately come to mind are: Vivien's advice and support these past few years, great bike rides and gossip sessions with Julin, Michael's tired look when playing squash and our dinner in a decade, Jeanne's job-hunting advice and her low scores on Jewelbox, Judith's incessant talking about her baby before and after birth, Stevo's distinctive laugh and guitar playing, Lisa's tirades and her disruptiveness (which, incidentally, have caused me to quit graduate school), Craig's skill at making oligos/PCR/cloning reactions that didn't work, all those unproductive games that Craig N. brought over, Nomi's denials about being a gnome, a dwarf like creature that lives underground and hoards treasure, Sonya's stories about all her dates and free dinners, Deborah's frankness and level-headedness, and Supriya's pleasant negative attitude. I'll undoubtedly be able to recall more memories when I get old and reminisce about my youth.

Lastly, I truly thank my advisor, Cynthia, who guided me through easy and difficult times, who was an excellent scientific role model, and who was absolutely instrumental in getting the cluster paper done. Her uncontrollable enthusiasm kept me going and probably made me do experiments that were, in retrospect, stupid. THANKS CHIEF!

Table of Contents

Introduction 1
Results

egl-5
egl-5, a HOM gene candidate for the far posterior body region 3
egl-5 is the homeobox gene *ceh-11*, a probable *Abd-B* homolog 3
An *egl-5-lacZ* fusion is expressed in the far posterior body region 5

lin-39
lin-39, a HOM gene candidate for the central body region 5
lin-39 is the homeobox gene *ceh-15*, an *Scr/Dfd/pb* homolog 6
A *lin-39-lacZ* fusion is expressed in the central body region 8

ceh-23
ceh-23 is a new homeobox gene in the C. elegans HOM cluster 8
ceh-23 is similar to Drosophila *empty spiracles* and *Distal-less* 9
A *ceh-23-lacZ* fusion is expressed in amphid neurons 9

Discussion 10
Experimental Procedures 15
References 21
Figures 26

List of Figures

Figure 1. Nucleotide and predicted amino acid sequences of *egl-5*
Figure 2. Similarities between nematode, fly, and murine HOM-C genes
Figure 3. Embryonic expression pattern of *lin-39-*, *mab-5-*, and *egl-5-lacZ* fusion genes
Figure 4. Expression pattern of an *egl-5-lacZ* fusion
Figure 5. Nucleotide and predicted amino acid sequences of *lin-39*
Figure 6. Expression pattern of a *lin-39-lacZ* fusion
Figure 7. Physical map of the C. elegans HOM cluster
Figure 8. *ceh-23* sequence comparisons and expression of a *ceh-23-lacZ* fusion
Figure 9. Nucleotide and predicted amino acid sequences of *ceh-23*
Figure 10. Structural and functional organization of the C. elegans HOM-C

Introduction

One of the most surprising and remarkable findings in recent years is that both insects and vertebrates, organisms with quite different developmental plans, contain analogous clusters of homeobox (HOM) genes. The eight Drosophila homeotic genes that specify segmental identity are arranged in a Homeotic complex, which consists of the Antennapedia and Bithorax complexes (reviewed in Akam, 1989). The order of these genes along the chromosome correlates with their functional region along the anteroposterior axis of the animal. Hox genes, the vertebrate (murine and human) homologs to the Drosophila genes, are organized in four clusters. The organization of the Hox clusters parallels that of Drosophila (Graham et al., 1989; Duboule and Dollé, 1989) as does the pattern of Hox gene expression along the anteroposterior axis (reviewed in McGinnis and Krumlauf, 1992). Furthermore, evidence for functional homology between the insect and vertebrate HOM genes is provided by mutations in the murine Hox genes and by gene swapping experiments involving insect and murine homologs (McGinnis and Krumlauf, 1992). Since such different organisms use a conserved cluster of homeobox genes in pattern formation along the anteroposterior axis, the question is raised of whether organisms with even different developmental plans also use the same system.

The nematode C. elegans is an organism whose development differs from both insects and vertebrates. For example, it has an invariant cell lineage, an asymmetric early embryonic cleavage pattern, a small number of cells, and it is non-segmented. Although at first glance the nematode does not appear to generate structures along its anteroposterior axis, there are in fact many structures that are located along this axis. For example, an elaborate

egg-laying system is located in the central body region of the hermaphrodite. In the male, a complex mating apparatus is located in the posterior body region and the tail. Other mesodermal and neural structures are located in both sexes along the anteroposterior axis.

Different classes of homeobox genes have been found in C. elegans and it has been estimated that there are about 60 homeobox genes in all (Bürglin et al., 1989). Four Antennapedia-type homeobox genes have been found to lie close to one another in the genome and are arranged in the same order as their closest homologs in the Drosophila and vertebrate clusters (A. Coulson, pers. comm.; Kenyon and Wang, 1991; Bürglin et al., 1991). One of these, the *Antennapedia* (*Antp*) homolog *mab-5*, is known to specify positional identity in the posterior body region (Kenyon 1986; Costa et al., 1988).

The conserved arrangement of the nematode HOM genes and the known function of one of these genes, *mab-5*, suggested that the nematode also uses a homeobox cluster to generate pattern along the anteroposterior body axis as do insects and vertebrates. The work presented here shows that this is indeed the case. The homeobox gene to the left of *mab-5*, a Drosophila *Sex Combs Reduced* (*Scr*)/*Deformed* (*Dfd*)/*proboscipedia* (*pb*) homolog, functions and is expressed in the central body region. The homeobox gene to the right of *mab-5*, a probable Drosophila *Abdominal-B* (*Abd-B*) homolog, functions and is expressed in the far posterior body region and tail. A new homeobox gene in the nematode cluster that appears to be a Drosophila *empty spiracles* (*ems*)/*Distal-less* (*Dll*) homolog is also identified. These findings indicate that the nematode, insect, and vertebrate clusters are structurally and functionally similar, suggesting that they all arose from a common ancient ancestor.

Results

The Abd-B homolog ceh-11 *is the gene* egl-5

Mutations that affected cells in a specific body region and that mapped to the homeobox cluster (HOM-C) were good candidates for HOM-C genes. The gene *egl-5* is required for the correct development of a number of cells located in the far posterior body region and tail (Chisholm, 1991). These cells are not closely related by lineage or by cell type but primarily by location (Figure 4A). Some of the structures that arise from cells in this region include male reproductive structures (such as the spicules, proctodeum, and sex muscles) and specialized neurons (such as the interneuron PVC and the HSN, which innervates the vulval muscles in the hermaphrodite). In *egl-5* mutants, some cells apparently undergo homeotic transformations: they adopt fates characteristic of their anterior homologs. In other cases, cells adopt fates not found elsewhere in the wild type. In addition, *egl-5* maps 0.03 m.u. to the right of *mab-5*. The homeobox *ceh-11* was isolated by several groups (Hawkins and McGhee, 1990; Schaller et al., 1990; "ceh" stands for C. elegans homeobox) and was subsequently found to be located to the right of *mab-5* on the physical map (A. Coulson, pers. comm.; Figure 7), where *egl-5* maps.

A. Chisholm first rescued the egg-laying defect (Egl) of *egl-5* mutants using cosmids that contained the *ceh-11* homeobox (pers. comm.). These initial rescue experiments were confirmed. Because the *ceh-11* homeobox is expected to be required for function, it could be altered in mutants. Thus, to show that *egl-5* and the gene containing the *ceh-11* homeobox were identical, the sequence of the *ceh-11* homeobox was examined in eleven *egl*-5 alleles. Mutations were found in three *egl-5* alleles (Figure 1). *egl-5(u202)* contained a

7 base-pair insertion that creates a stop codon at the beginning of helix 2 of the homeodomain. This mutation should eliminate DNA binding because the resulting protein lacks homeodomain helix 3, the highly conserved DNA recognition helix that mediates DNA binding (reviewed in Gehring et al., 1990). Two other alleles, *n486* and *e2495*, contained the same C to T change, replacing Arg 52 with a Cys residue. Arg 52 is a highly conserved residue within helix 3 of the homeodomain. Because three *egl-5* alleles contain mutations within conserved regions of the *ceh-11* homeobox and because the homeobox is located on a cosmid that rescues *egl-5* mutants, *egl-5* is the gene that contains the *ceh-11* homeobox. These results indicate that *egl-5(u202)* is a likely null allele (Chisholm, 1991).

The *egl-5/ceh-11* homeodomain has comparable levels of homology to several Drosophila Antp-class homeodomains. *egl-5* is most similar to *Antp* (57% identity), *Ubx* (55%), and *abd-A*(53%), which are thought to have arisen from a single Antp-like precursor present in the original insect-vertebrate cluster and diverged from one another following the separation of insects and vertebrates (Figure 2; Akam et al., 1988; Schubert et al., 1993). The *egl-5* homeodomain also resembles *Abd-B* (53%), which is thought to have arisen from another precursor present in the original insect-vertebrate cluster prior to the divergence of insects and vertebrates (Akam et al., 1988; Schubert et al., 1993). To clarify the relationship between *egl-5* and these other HOM-C genes, the sequence of the entire *egl-5* coding region was determined (Figure 1). There are additional regions of homology to *Ubx* and *abd-A* downstream of the homeodomain which suggest that *egl-5* may belong to these classes (Figure 2). However, Antp-class homeodomain proteins such as *Ubx* and *abd-A* share a hexapeptide motif (containing the residues "YPWM") located just N-terminal of the homeodomain. *Abd-B* and vertebrate homologs are not

considered Antp-class homeodomain proteins and lack this conserved motif. *egl-5* also lacks the hexapeptide motif. Thus, it seems most appropriate to consider *egl-5* as most similar to the *Abd-B* class.

egl-5 *is expressed in cells that require its function in the far posterior body region and tail*

HOM-C genes are typically expressed in a position-specific manner in the region that requires their function. To determine where *egl-5* was likely to be expressed, an *egl-5-lacZ* fusion was constructed. In two transgenic lines, the fusion gene was consistently expressed only in a restricted domain in the far posterior body region and tail of embryos (Figure 3), larvae, and adults (Figure 4). Cells that often expressed the fusion in the L1 stage were those that are known to require *egl-5* function at this time (S. Salser, C. Hunter, and C. Kenyon, pers. comm.). These cells include the blast cells B, U, F, Y, and K, and the HSN, which is born and differentiates in the tail region and migrates to the central body region during embryogenesis in an *egl-5* dependent fashion.

The Scr/Dfd/pb *homolog* ceh-15 *is the gene* lin-39

The homeobox *ceh-15* was isolated in a screen for more Antp-class homeoboxes (Kamb et al., 1989) and was subsequently found to physically map to the left of *mab-5* (A. Coulson, pers. comm; Figure 7). The *ceh-15* homeodomain is most similar to the homeodomains of Drosophila *Scr* (77% identity), *Dfd* (77%) and *pb* (73%) (Figure 2). Because these genes specify segmental identity in the domains just anterior to *Antp* in Drosophila, it seemed possible that *ceh-15* could specify identity in the central body region, anterior to the *mab-5* region in C. elegans.

The gene *lin-39* was a strong candidate for the HOM gene that specifies the central body region based on its mutant phenotype and map position. *lin-39* was first identified by H. Ellis and H. Horvitz (Horvitz et al., 1982; Ellis, 1985). In *lin-39(n709)*, cells located in the central body region that normally become VC neurons, motorneurons that innervate vulval muscles in the hermaphrodite, instead adopt a fate characteristic of their homologs in other body regions and undergo programmed cell death. The vulval cell precursors, the P(3-8).p cells, are also located in the central body region and often fail to divide (Ellis, 1985). These vulval cell precursors also fail to divide in stronger alleles isolated by S. Clark and H. Horvitz (pers. comm.) and by J. Austin [pers. comm.; *lin-39(mu26);* Figure 6A]. In addition, the migration of the descendants of the QR neuroblast which traverse the central body region is affected [first noted by C. Bargmann, S. Clark and H. Horvitz (pers. comm.); Figure 6A].

lin-39 defects appear to be restricted to the central body region as structures located in other body regions are normal (Wang et al., in press; C. Kenyon, pers. comm.; Clark et al., in press). Normal posterior structures include the male mating apparatus (rays, hook, diagonal muscles and spicules), the postdeird neuroblast, and the M-derived coelomocyte. Normal anterior structures include the excretory pore and BDU. Serotonergic neurons in the head and tail are not affected (Loer and Kenyon, 1993). *lin-39* is located on chromosome III (Wood et al., 1988) just to the left of *mab-5* (S. Clark and H. Horvitz, pers. comm.), where the *ceh-15* homeobox is located.

To determine whether *lin-39* and the gene containing the *ceh-15* homeobox were identical, cosmids that contained the *ceh-15* homeobox were used to attempt to rescue the vulvaless (Vul) phenotype of *lin-39(mu26)*. One cosmid did rescue the Vul phenotype as did a 10 kb subclone that contained

the homeobox. To show that the *ceh-15* gene was responsible for rescuing activity, a nonsense mutation was made in the predicted hexapeptide sequence just upstream of the homeodomain (Figure 5). The construct containing this mutation failed to rescue. When a 1 kb wild-type DNA fragment overlapping the mutation (Figure 5) was coinjected along with the mutated construct, rescuing activity was restored. These two fragments were together able to rescue the Vul phenotype probably as a result of homologous recombination between the two fragment *in vivo*. This control argues that the nonsense mutation, rather that some other unknown mutation, destroyed rescuing activity. In addition, the *ceh-15* homeobox of the *lin-39(mu26)* allele was sequenced and a G to A transition was found that converts an absolutely conserved Trp residue, found in all Antp-class recognition helices (Scott et al., 1989), to a nonsense mutation (Figure 5). This mutation should eliminate DNA binding since part of the recognition helix is missing. Because only an intact *ceh-15* homeobox gene can rescue *lin-39* and since *lin-39(mu26)* contains a nonsense mutation in the *ceh-15* homeobox, *lin-39* is the gene containing the *ceh-15* homeobox. These results suggest that *lin-39(mu26)* is a null allele.

As discussed above, the *lin-39/ceh-15* homeodomain has homology to the *Scr*, *Dfd*, and *pb* homeodomains. To find out whether *lin-39* had any additional homology outside the homeodomain that would aid in classifying *lin-39* into a particular group, the sequence of the entire *lin-39* coding sequence was determined (Figure 5). The inferred protein sequence has additional homology to each of the above three genes in the regions surrounding the hexapeptide motif, upstream, and downstream of the homeodomain. However, *lin-39* has a similar high level of identity to each of the three Drosophila genes (Figure 2; see Discussion).

lin-39 *is expressed in the central body region*

To determine whether *lin-39*, like *egl-5*, was expressed in a position-specific manner, a *lin-39-lacZ* fusion was constructed using more upstream DNA than was required for rescuing activity. In two transgenic lines, the fusion was expressed in the central body region of embryos (Figure 3), larvae and adults (Figure 6). Among the cells that express the fusion in the L1 stage are the P(3-8) cells and the QR neuroblast (S. Salser, N. T. Robinson, and C. Kenyon, pers. comm.). The descendants of these cells also express the fusion although it is unclear whether this is due to continued transcription or perdurance of β-galactosidase activity. P cells located in other body regions did not express the fusion.

An ems/Dll *homolog is part of the HOM cluster and is expressed in the amphids*

Are there other homeoboxes in the cluster? It is possible that other homeoboxes exist because some structures along the anteroposterior axis, such as the postdeird neuroblast and the T rays, are not affected by known HOM gene mutations. A set of overlapping cosmids and YACs that span the known cluster (200-300 kb) and in addition extend about 200-300 on each side were probed with an oligonucleotide that detects Antp-class homeoboxes (Bürglin et al., 1989; Figure 7). Aside from the known HOM genes, a number of hybridizing bands were sequenced but only one appeared to be an authentic homeobox. This homeobox, designated *ceh-23*, is located to the right of *egl-5* (Figure 7, 10). Other authentic homeoboxes would not have been detected if they were deleted from the cosmid and YAC subclones, were significantly different from the probe, or were not included in the region examined. The

complete nature of the cluster will be known when the C. elegans genome sequencing project determines the nucleotide sequence of this region.

The *ceh-23* homeodomain is most similar to the Drosophila genes *ems* (52% identity) and *Dll* (50%) which are distant relatives of Antp-class homeoboxes (Figure 8). In order to identify additional homologies, the coding sequence of the *ceh-23* gene was determined (Figure 9). No significant additional homologies were found outside the homeodomain although the *ceh-23* gene lacks the hexapeptide sequence found upstream of the homeodomain in *ems* and vertebrate homologs. Interestingly, the Drosophila genes *ems* and *Dll* are required for correct development of the embryonic head. In *ems* mutants, specific head structures are deleted and in particular, the antennal sense organs are missing. The *ems* protein is expressed in the primordia of these sense organs (Jürgens et al., 1984; Dalton et al., 1989). The mouse *ems* homologs, *Emx1* and *Emx2*, are also expressed in defined regions of the brain, including the olfactory bulb (Simeone et al., 1992). *Dll* is required for and is expressed in both the thoracic limbs and head structures (Cohen, 1990). Mouse homologs are also expressed in restricted areas of the brain (Price at al., 1991) and one, *Dlx-2*, is expressed in the olfactory bulb (Porteus et al., 1991; Robinson et al., 1991).

To determine where the *ceh-23* gene was likely to be expressed, a *ceh-23-lacZ* fusion was constructed and two transgenic lines were obtained. In larvae and adults, the fusion was often expressed in the amphid sensilla (C. Bargmann, pers. comm.; Figure 8), a group of chemosensory neurons whose cell bodies located in the head of the animal and whose processes extend into the nose. The fusion was also often expressed in the CAN cells (C. Bargmann, pers. comm.; Figure 8), which may be involved in osmotic regulation.

Discussion

Organisms as diverse as insects and vertebrates use evolutionarily conserved homeobox gene clusters to specify pattern along the anteroposterior axis. This thesis shows that the nematode C. elegans, an organism only distantly related to insects and vertebrates, also uses an analogous cluster to generate pattern along its anteroposterior axis, indicating that the C. elegans cluster is not simply an evolutionary vestige.

Figure 10 compares the C. elegans and Drosophila clusters and shows the cells that require the known HOM genes for correct development. *mab-5*, the *Antp* homolog, patterns and is expressed in the posterior body region (Kenyon, 1986; Costa et al., 1988). To the right of *mab-5* is the gene *egl-5*, a probable *Abd-B* homolog, which patterns the far posterior body region and tail (Chisholm, 1991). As assayed by a *lacZ* fusion, *egl-5* appears to be expressed in this posterior region within cells that require its function. It seems most accurate to classify *egl-5* as an *Abd-B* homolog because like *Abd-B*, *egl-5* lacks the hexapeptide motif that is present in Antp-class homeodomain proteins. Functional studies (for example, where Drosophila or murine *Abd-B* homologs are expressed in *egl-5* mutants or where the relative importance of homeodomain residues is assessed by mutation) would help address this issue of classification.

To the left of *mab-5* is an *Scr/Dfd/pb* homolog, *lin-39*, which patterns and is expressed in the central body region as determined by a *lacZ* fusion. Again, without functional studies, it is unclear to which class *lin-39* most appropriately belongs. It is interesting, however, that *lin-39* has similar levels of identity to each of these three genes in the homeodomain, and the regions surrounding the homeodomain and hexapeptide motif. Perhaps the

ancestral cluster contained one *lin-39*-like gene which, following the separation from worms, subsequently duplicated and diverged in the lines leading to flies and vertebrates. Alternatively, the ancestral cluster may have contained these three genes and the nematode may have created a hybrid gene through recombination and deletion events.

The final Antp-class homeobox gene in the cluster is *ceh-13*. The function of *ceh-13* is not known, although based on its sequence similarity to the Drosophila gene *labial (lab)* and vertebrate homologs, and the results presented here, *ceh-13* is likely to be required for and expressed in the region just anterior to the *lin-39* domain.

Overall, the C. elegans and Drosophila clusters share extensive similarity both structurally and functionally. Although it is unclear exactly which Drosophila gene corresponds to *egl-5* and *lin-39*, it is clear that the C. elegans HOM genes are arranged in the same order as their closest homologs in other organisms. Furthermore, the known C. elegans HOM genes function along the anteroposterior axis in the same order as their closest homologs do in other organisms. What, then, did the ancient cluster look like? Based on sequence analysis of Drosophila, murine, and human homeobox genes, it has been suggested recently that the early metazoan precursor had a three gene cluster: an *Abd-B*-like gene at the 5' end of the cluster, an *Antp/Dfd*-like gene in the middle of the cluster, and a *pb/lab*-like gene and the 3' end of the cluster (Schubert et al., 1993). The C. elegans HOM cluster is consistent with this hypothesis since *egl-5* is *Abd-B*-like, *mab-5* and *lin-39* are *Antp/Dfd*-like, and *ceh-13* is most like *lab*. Further support for an ancient three gene cluster will come when more homeoboxes and genomic organizations are determined from primitive metazoans.

Only one additional homeobox gene, *ceh-23*, was detected in a screen for more Antp-class homeobox genes in the cluster region. Mutants are needed to determine the function of this non-Antp-class gene. Although the function is unknown, the similar sequences and expression patterns of the C. elegans, Drosophila and vertebrate genes are highly suggestive. The Drosophila *ems* and *Dll* genes are not located in HOM clusters. However, a human *ems* homolog, *EMX1*, is located in an analogous position just 5' of the HOX4 locus (E. Boncinelli, pers. comm.), supporting the possibility that an *ems* precursor was present in the ancestral HOM cluster. Perhaps the *ems* precursor generated head structures including olfactory neurons in a common ancestor of nematodes, insects and vertebrates, a function retained throughout evolution.

Are other mechanisms such as HOM gene initiation, maintenance and activation of targets genes also conserved in C. elegans? The anteroposterior structures affected by the C. elegans HOM genes are made primarily postembryonically. Putative null mutations in the HOM genes allow embryogenesis to occur relatively normally. Thus, early developmental events in the worm, for example, the early asymmetric cleavages, do not require HOM gene function, just as the Drosophila HOM genes are not require for early syncytial blastoderm development. Aside from this similarity, however, early work suggests that C. elegans and Drosophila may use different systems in initially setting up the HOM gene pattern (Cowing and Kenyon, 1992; Krumlauf and McGinnis, 1992). Little is known in C. elegans about how the HOM cluster system, once set up, is maintained and activates downstream target genes. Preliminary work suggests that there may be genes in C. elegans that function to stabilize the expression of HOM genes just as the *Polycomb* and *trithorax* genes do in Drosophila (J. Maloof and C.

Kenyon, unpublished). Similar to the results of gene-swapping experiments with Drosophila and murine genes, early work indicates that Drosophila homologs of *mab-5* and *lin-39* can compensate in the corresponding C. elegans mutants, suggesting conserved downstream targets (C. Hunter and C. Kenyon, unpublished). Given the conserved nature of the HOM cluster, it is perhaps not too surprising that other mechanisms may also be conserved.

At least one aspect of the HOM cluster function, however, appears not to be conserved in C. elegans to a great extent. In Drosophila, the pattern of HOM gene expression is refined by cross-regulatory interactions between the HOM genes. This cross-regulation usually involves the phenotypic suppression of a more anterior gene by a more posterior gene. Likewise, there are hints from loss-of-function Hox mutations that this phenotypic dominance or posterior prevalence of more posterior genes may also play a role in vertebrates (Lufkin et al., 1991). In C. elegans, however, this does not seem to be the case. The phenotypes resulting from null mutations in *egl-5*, for example are not caused by ectopic *mab-5* expression (Chisholm, 1991). Similarly, the posterior body region in *mab-5* mutants becomes like an unspecialized anterior region, not the central body region, suggesting that *lin-39* is not activated in the posterior body region. Interactions between HOM genes in specific cell types have been observed, however, and will be described elsewhere (Salser et al., submitted; Wang et al., in press).

With its non-segmented body plan, asymmetric early embryonic cleavages and invariant cell lineages, C. elegans develops much differently than either Drosophila or vertebrates. Nevertheless, C. elegans uses a conserved cluster to generate anteroposterior pattern suggesting that each of these differences arose during evolution at a time after the HOM patterning system was established. Do simpler organisms contain HOM clusters? Lower

metazoans with even different developmental plans, such as the diploblastic hydra which has no mesoderm (Schummer et al., 1992), and the flatworm which is thought to represent the most primitive bilateral metazoan (Webster and Mansour, 1992), have also been found to contain Antp-class homeobox genes. Whether or not they are clustered is unknown. In the case of the hydra, the wild-type expression patterns of these homeobox genes, *labial* and *Dfd* homologs, are not known. However, they are expressed sequentially in the regenerating hydra head in a manner suggesting that they may be involved in anteroposterior patterning (Schummer et al., 1992).

Thus far, C. elegans is the simplest organism shown to use a HOM cluster. The remarkable conservation of the HOM cluster suggests that the precursor to all bilateral metazoans may have also used an ancient cluster. Homeobox genes have been found in a multitude of organisms ranging from fungi to humans. In the non-animal species (fungi, for example), only divergent non-Antp-class homeoboxes have been found; they are not arranged in clusters. As discussed above, Antp-class homeobox genes have been found in lower animal species but whether they are clustered remain unknown. It has been suggested recently (Slack et al., 1993) that the homeobox cluster may be considered a molecular definition for what constitutes an "animal". If this observation holds true, as more animal and non-animal species are examined, the conservation of the HOM cluster will not seem out of the ordinary. What will seem extraordinary are those animals which generate pattern by non-HOM cluster means.

Experimental Procedures

General Nematode Procedures

Methods for routine culturing and genetic analysis have been described previously (Brenner, 1974; Wood et al., 1988). All analyses were performed at 20°C.

Molecular Cloning of *egl-5* and *lin-39*

egl-5 A. Chisholm first rescued the egg-laying defect (Egl) phenotype of *egl-5* mutants using cosmids containing the homeobox *ceh-11*. These initial observations were repeated and extended by rescuing the Egl phenotype of *egl-5(u202)* using *rol-6(su1006)* as a coinjection marker (100 μg/ml). One cosmid, C08C3 (10 μg/ml), fully rescued the Egl phenotype in 4/9 F1 transformants (progeny of injected animals). The *ceh-11* homeobox from 11 *egl-5* alleles (*n945, n988, u202, e2399, e2502, e2506, e2508, n989, n486, e2495,* and *n1439*) was amplified by PCR and sequenced by dsDNA cycle sequencing (BRL). In the initial screen, DNA changes were detected in three alleles, *u202, e2495,* and *n486*. To eliminate the possibility of PCR artifacts and to confirm the changes, both strands from three independent PCR reactions were sequenced for these alleles.

lin-39 The recessive *lin-39(mu26)* allele was isolated in a screen for mutants in which Q neuroblast migration is altered (J. Austin and C. Kenyon, unpublished) and was characterized by J. Austin. Complementation testing was done using *lin-39(n1760)* following transformation rescue since *lin-39* hermaphrodites are vulvaless (Vul). Because of the similarities in map position of the homeobox *ceh-15* and the gene *lin-39*, cosmids containing *ceh-15* were used to attempt to rescue the Vul phenotype of *mu26* using *rol-*

6(su1006) as a coinjection marker (100 μg/ml). Cosmid F44F12 (30 μg/ml) fully rescued the Vul phenotype (a functional vulva was present and eggs were laid) in 2/8 F1 transformants and a smaller subclone (10 kb; 10 μg/ml) also containing the homeobox fully rescued 18/34 F1 transformants. To show that the *ceh-15* homeobox was part of *lin-39*, a mutation was made at a unique NcoI site which occurs in the predicted hexapeptide sequence of *ceh-15*. The mutation was created in the above 10 kb subclone by digesting with NcoI, filling in with Klenow and religating. Sequencing showed that the resulting mutation was an opal stop codon. This mutant 10 kb fragment (ΔNco4; 10-20 μg/ml) failed to rescue (0/71 F1 transformants). To show that this nonsense mutation was likely to be responsible for the failure to rescue, a 1 kb wild-type fragment (20 μg/ml) that overlaps the mutation was coinjected along with ΔNco4 into *lin-39(mu26)* animals. The Vul phenotype was partially rescued in 1/34 F1 transformants (a nonfunctional vulva was present; no eggs were laid). In addition, a mutation was found in the *ceh-15* homeobox of the *mu26* allele when the homeobox was amplified by PCR and sequenced by dsDNA cycle sequencing (BRL). Sequencing of both strands from three independent PCR reactions confirmed this mutation.

Isolation and Sequencing of *egl-5*, *lin-39* and *ceh-23* cDNAs

To identify *egl-5* cDNA clones, the *egl-5* homeobox was amplified by PCR from a cosmid subclone and was used to probe 800,000 clones of a cDNA library constructed using mRNA from *him-8 (e1489)* embryos (gift of L. Miller and B. Meyer). Two independent 1.0 kb cDNA clones were isolated that had different length poly A tails. Both strands of one cDNA and one strand of corresponding genomic DNA were sequenced using Sequenase 2.0 (USB) and an Applied Biosystems automated sequencer at the Biomolecular Resource

Center (UCSF). This cDNA clone appeared to be a hybrid cDNA based on sequence comparisons with genomic DNA. The C. elegans genome sequencing project provided the sequence data necessary to arrive at this conclusion.

To identify *lin-39* cDNA clones, a 3 kb EcoRI cosmid subclone that contains part of the *lin-39* homeobox was used to probe 1 x 10^6 clones of a mixed stage cDNA library (gift of B. Barstead and R. Waterston) and 800,000 clones of the *him-8 (e1489)* embryonic cDNA library (L. Miller and B. Meyer). Two independent cDNA clones from the first library and a number of cDNA clones from the second library were isolated. Both strands of the longest cDNA (1.3 kb) and one strand of corresponding genomic DNA were sequenced as described above for *egl-5*.

The homeobox oligonucleotide HB-1 (Bürglin et al., 1989) was used to probe a set of overlapping cosmids (ZC104, C45B1, C07B9, C17A3, C47B12, F08F8, T20B12, C30C5, C03D10, C29A11, F44F12, C34A5, K07D8, T04A6, ZK420, T07B11, C55A11, ZK694, C44F11, C03B8, ZC31, C33C3, ZC97, ZC102, ZK686, C08C3, C27D11, M22, C35F6, C39F10, C44C9, ZK581, C04H11) and total yeast DNA containing YACs (Y50B11, Y69F12, Y16G1, Y81H8, Y39H8) that are expected to span the cluster region (see Figure 7). Many hybridizing bands were sequenced. However, other than *ceh-23* (located on cosmid C27D11), no additional new homeoboxes were discovered.

To identify *ceh-23* cDNA clones, part of the *ceh-23* homeobox was amplified by PCR from genomic DNA and was used to probe 2 x 10^6 clones from the mixed-stage cDNA library (B. Barstead and R. Waterston). A single cDNA clone was isolated. At least one strand of the cDNA and corresponding genomic DNA were sequenced using the facilities of the Biomolecular

Resource Center (UCSF). This cDNA clone appeared to be a hybrid cDNA based on sequence comparisons with genomic DNA.

Sequence comparisons were done by visual inspection and by using the FASTA program (Pearson and Lipman, 1988) on the NBRF/PIR protein database (version 33 and 35).

Construction and Analysis of the *egl-5-*, *lin-39-*, and *ceh-23-lacZ* fusions

In an attempt to make *egl-5-* and *lin-39-lacZ* fusions that would accurately reflect the wild-type expression pattern, large regions upstream and downstream of each gene, as well as all introns, were included in the constructs. Fusions were constructed in addition to those described below that had shorter upstream and downstream regions. These fusions had localized staining patterns but also frequently showed additional expression in other body regions. Previous experiments have shown that a *mab-5-lacZ* construct containing only 7 kb of upstream sequence is expressed in a wild-type pattern as assayed by *mab-5* antibodies (Salser and Kenyon, unpublished), suggesting that C. elegans HOM gene control regions, unlike Drosophila, may be small.

A 4.1 kb fragment from the expression vector pPD21.28 (Fire et al., 1990) that contains the *lacZ* gene, SV40 nuclear localization signal and *unc-54* 3' UTR was cloned into a unique ApaI site of a 22 kb NruI subclone of cosmid C08C3 that contains the entire *egl-5* coding sequence (see Figure 1). The resulting *egl-5-lacZ* fusion has about 17 kb upstream of the first predicted AUG and 3 kb downstream of the predicted stop codon. This fusion contains the first 86 residues of the *egl-5* protein.

To construct the *lin-39-lacZ* fusion, pPD21.28 was modified by removing the synthetic intron, SV40 nuclear localization signal, and the *unc-*

54 3' UTR (pPD21.28Δ). The 3.3 kb fragment containing only the *lacZ* gene was then cloned into a unique NcoI site in a 25 kb PstI/EcoRV subclone of cosmid F44F12 that contains *lin-39* (see Figure 5). The resulting *lin-39-lacZ* fusion has about 13 kb upstream of the first predicted AUG and 3.8 kb downstream of the predicted stop codon. This fusion contains the first 150 residues of the *lin-39* protein and has more upstream sequence than is required for rescuing activity.

To construct a *lacZ* fusion likely to represent the actual expression pattern of the gene containing *ceh-23*, a 13 kb fragment of genomic DNA whose 3' terminus was a SnaB1 site within the *ceh-23* homeobox was joined in frame to *lacZ* using the pPD21.28 vector (see Figure 9). It is not certain that this fusion contains all the regulatory regions required for authentic *ceh-23* expression. Based on comparisons between the *ceh-23* cDNA and genomic DNA, this fusion contains no more than about 12 kb of upstream sequence.

The *lacZ* fusion constructs (50-100 μg/ml for *egl-5-lacZ*; 15 μg/ml for *lin-39-lacZ*; 30 μg/ml for *ceh-23-lacZ*) were injected into *him-5(e1490)* or N2 (wild type) hermaphrodites along with the *rol-6 (su1006)* coinjection marker (100 μg/ml). Two lines that expressed the coinjection marker (Rol) were obtained for each fusion. The DNA was not stably inherited in these lines, implying that the injected DNA was maintained extrachromosomally. The majority of Rol animals showed the staining patterns described in Figures 3, 4, 6, and 8.

To assay for β-galactosidase activity, embryos were gently scraped from plates, allowed to settle on polylysine-coated slides, and frozen on dry ice. The embryos were dehydrated in -20°C acetone and rehydrated by a 90%, 60%, 30%, 10% acetone series at room temperature. The embryos were then incubated in staining solution from 3 hrs to overnight at room temperature

References

Akam, M. (1989). Hox and HOM: homologous gene clusters in insects and vertebrates. Cell *57*, 347-349.

Akam, M., Dawson, I., and Tear, G. (1988). Homeotic genes and the control of segment diversity. Development *104* (Suppl.), 123-133.

Brenner, S. (1974). The genetics of *C. elegans*. Genetics *77*, 71-94.

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

Bürglin, T. R., Ruvkun, G., Coulson, A., Hawkins, N. C., McGhee, J. D., Schaller, D., Wittmann, C., Müller, F. and Waterston, R. H. (1991). Nematode homeobox cluster. Nature *351*, 703.

Celniker, S. E., Keelan, 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.

Chisholm, A. (1991). Control of cell fate in the tail region of *C. elegans* by the gene *egl-5*. Development *111*, 921-932.

Clark, S. G., Chisholm, A., and Horvitz, H. R. (1993). Control of cell fates in the central body region of C. elegans by the homeobox gene *lin-39*. Cell, in press.

Cohen, S. M. (1990). Specification of limb development in the *Drosophila* by positional cues from segmentation genes. Nature 343, 173-177.

Costa, M., Weir, M., Coulson, A., Sulston, J. and Kenyon, C. (1988). Posterior pattern formation in *C. elegans* involves position-specific expression of a gene containing a homeobox. Cell *55*, 747-756.

Cowing, D. and Kenyon, C. (1992). Expression of the homeotic gene *mab-5* during *Caenorhabditis elegans* embryogenesis. Development *116*, 481-490.

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

Dalton, D., Chadwick, R. and McGinnis, W. (1989). Expression and embryonic function of *empty spiracles:* a *Drosophila* homeobox gene with two patterning functions on the anterior-posterior axis of the embryo. Genes Dev. *3*, 1940-1956.

Diederich, R. J., Merrill, V. K., Pultz, M. A. and Kaufman, T. C. (1989). Isolation, structure, and expression of *labial*, a homeotic gene of the *Antennapedia* Complex involved in *Drosophila* head development. Genes Dev. *3*, 399-414.

Duboule, D., and Dollé, P. (1989). The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes. EMBO J. *8*, 1497-1505.

Ellis, H. (1985). Genetic control of programmed cell death in the nematode *Caenorhabditis elegans*. Ph.D. thesis (Massachusetts Institute of Technology).

Fibi, M., Zink, B., Kessel, M., Colberg-Poley, A.M., Labeit, S., Lehrach H., Gruss, P. (1988). Coding sequence and expression of the homeobox gene *Hox 1.3*. Development *102*, 349-359.

Fire, A. Histochemical techniques for locating *E. coli* β-galactosidase activity in transgenic organisms. Gene Analysis Techniques and Applications, in press.

Fire, A., Harrison, S. W. and Dixon, D. (1990). A modular set of lacZ fusion vectors for studying gene expression in *Caenorhabditis elegans*. Gene *93*, 189-198.

Frohman, M. A., Boyle, M. and Martin, G. R. (1990). Isolation of the mouse *Hox-2.9* gene; analysis of embryonic expression suggests that positional information along the anterior-posterior axis is specified by mesoderm. Development *110*, 589-607.

Galliot, B., Dollé, P., Vigneron, M., Featherstone, M. S., Baron, A. and Duboule, D. (1989). The mouse *Hox-1.4* gene: primary structure, evidence for promoter activity and expression during development. Development *107*, 343-359.

Gehring, W. J., Müller, M., Affolter, M., Percival, S. A., Billeter, M., Qian, Y. Q., Otting, G. and Wuthrich, K. (1990). The structure of the homeodomain and its functional implications. Trends Genet. *6*, 323-329.

Graham, A., Papalopulu, N., and Krumlauf, R. (1989). The murine and Drosophila homeobox gene clusters have common features of organization and expression. Cell *57*, 367-378.

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

Horvitz, H. R., Ellis, H.M., and Sternberg, P.W. (1982). Programmed cell death in nematode development. Neurosci. Comment. *1*, 56-65.

Izpisúa-Belmonte, J.-C., Falkenstein, H., Dollé, P., Renucci, A. and Duboule, D. (1991). Murine genes related to the *Drosophila Abd-B* homeotic genes are sequentially expressed during development of the posterior part of the body. EMBO J. *10*, 2279-2289.

Jürgens, G., Wieschaus, E., Nüsslein-Volhard, C., and Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. II. Zygotic loci in the third chromosome. Roux's Arch. Dev. Biol. *193*, 283-295.

Kamb, A., Weir, M., Rudy, B., Varmus, H. and Kenyon, C. (1989). Identification of genes from pattern formation, tyrosine kinase, and potassium channel families by DNA amplification. Proc. Natl. Acad. Sci. USA *86*, 4372-4376.

Karch, F., Bender, W. and Weiffenbach, B. (1990). *abd-A* expression in *Drosophila* embryos. Genes Dev. *4*, 1573-1587.

Kenyon, C. (1986). A gene involved in the development of the posterior body region of *C. elegans*. Cell *46*, 477-487.

Kenyon, C. and Wang, B. (1991). A cluster of *Antennapedia*-class homeobox genes in a nonsegmented animal. Science *253*, 516-517.

Kongsuwan, K., Allen, J. and Adams, J. M. (1989). Expression of *Hox-2.4* homeobox gene directed by proviral insertion in a myeloid leukemia. Nucleic Acids Res. *17*, 1881-1892.

Kornfeld, K., Saint, R. B., Beachy, P. A., Harte, P. J., Peattie, D. A. and Hogness, D. S. (1989). Structure and expression of a family of *Ultrabithorax* mRNAs generated by alternative splicing and polyadenylation in *Drosophila*. Genes Dev. *3*, 243-258.

Laughon, A., Boulet, A. M., Bermingham, J. R., Laymon, R. A. and Scott, M. P. (1986). Structure of transcripts from the homeotic *Antennapedia* gene of *Drosophila melanogaster*: two promoters control the major protein-coding region. Mol. Cell Biol. *6*, 4676-4689.

LeMotte, P. K., Kuroiwa, A., Fessler, L. I. and Gehring, W. J. (1989). The homeotic gene *Sex Combs Reduced* of *Drosophila*: gene structure and embryonic expression. EMBO J. *8*, 219-227.

Loer, C. M. and Kenyon, C. (1993). Serotonin-deficient mutants and male mating behavior in the nematode *C. elegans*. Neuroscience, in press.

Lufkin, T., Dierich, A., LeMeur, M., Mark, M., and Chambon, P. (1991). Disruption of the *Hox-1.6* homeobox gene results in defects in a region corresponding to its rostral domain of expression. Cell *66*, 1105-1119.

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

Meijlink, F., de Laaf, R., Verrijzer, P., Destrée, O., Kroezen, V., Hilkens, J. and Deschamps, J. (1987). A mouse homeobox containing gene on chromosome 11: sequence and tissue-specific expression. Nucleic Acids Res. *15*, 6773-6786.

Pearson, W. R. and Lipman, D. J. (1988). Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA *85*, 2444-2448.

Porteus, M. H., Bulfone, A., Ciaranello, R. D. and Rubenstein, J. L. (1991). Isolation and characterization of a novel cDNA clone encoding a homeodomain that is developmentally regulated in the ventral forebrain. Neuron *7*, 221-229.

Price, M., Lemaistre, M., Pischetola, M., Di, L. R. and Duboule, D. (1991). A mouse gene related to *Distal-less* shows a restricted expression in the developing forebrain. Nature *351*, 748-751.

Regulski, M., McGinnis, N., Chadwick, R. and McGinnis, W. (1987). Developmental and molecular analysis of *Deformed*; a homeotic gene controlling *Drosophila* head development. EMBO J. *6*, 767-778.

Robinson, G. W., Wray, S., and Mahon, K. A. (1992). Spatially restricted expression of a member of a new family of murine Distal-less homeobox genes in the developing forebrain. New Biologist *3*, 1183-1194.

Salser, S. J. and Kenyon, C. (1992). Activation of a *C. elegans Antennapedia* homolog in migrating cells controls their direction of migration. Nature *355*, 255-258.

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

Schubert, F. R., Nieselt-Struwe, K., and Gruss, P. (1993). The Antennapedia-type homeobox genes have evolved from three precursors separated early in metazoan evolution. Proc. Natl. Acad. Sci. USA *90*, 143-147.

Schughart, K., Utset, M. F., Awgulewitsch, A. and Ruddle, F. H. (1988). Structure and expression of *Hox-2.2*, a murine homeobox-containing gene. Proc. Natl. Acad. Sci. USA *85*, 5582-5586.

Schummer, M., Scheurlen, I., Schaller, C., and Galliot, B. (1992). HOM/HOX homeobox genes are present in hydra (Chlorohydra viridissima) and are differentally expressed during regeneration. EMBO J. *11*, 1815-1823.

Scott, M. P., Tamkun, J. W. and Hartzell, G. 3d. (1989). The structure and function of the homeodomain. Biochim. Biophys. Acta *989*, 25-48.

Scott, M. (1992). Vertebrate Homeobox Gene Nomenclature. Cell *71*, 551-553.

Simeone, A., Gulisano, M., Acampora, D., Stornaiuolo, A., Rambaldi, M. and Boncinelli, E. (1992). Two vertebrate homeobox genes related to the *Drosophila empty spiracles* gene are expressed in the embryonic cerebral cortex. EMBO J. *11*, 2541-2550.

Slack, J., Holland, P., and Graham, C. (1993). The zootype and the phylotypic stage. Nature *361*, 490-492.

Sulston, J. E. and Horvitz, H. R. (1977). Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. Dev. Biol. *56*, 110-156.

Tan, D.-P., Ferrante, J., Nazarali, A., Shao, X., Kozak, C. A., Guo, V. and Nirenberg, M. (1992). Murine *Hox-1.11* homeobox gene structure and expression. Proc. Natl. Acad. Sci. USA *89*, 6280-6284.

Vachon, G., Cohen, B., Pfeifle, C., McGuffin, M. E., Botas, J. and Cohen, S. M. (1992). Homeotic genes of the Bithorax complex repress limb development in the abdomen of the *Drosophila* embryo through the target gene *Distal-less*. Cell *71*, 437-450.

Wang, B. B., Müller-Immergluck, M. M., Austin, J., Robinson, N. T., Chisholm, A., and Kenyon, C. (1993). A homeotic gene cluster patterns the anteroposterior body axis of C. elegans. Cell, in press.

Webster, P. J. and Mansour, T. E. (1992). Conserved classes of homeodomains in *Schistosoma mansoni*, an early bilateral metazoan. Mech. Dev. *38*, 25-32.

Wood, W. B. (1988). in *The Nematode Caenorhabditis elegans* (ed. W. B. Wood) (Cold Spring Harbor Laboratory, New York).

Figure 1. Nucleotide and predicted amino acid sequences of *egl-5*

The sequence, based on cDNA and genomic clones, predicts a protein of 223 amino acids. Positions of introns are indicated by arrowheads (▼). The box indicates the most likely translation start site, as inferred from other C. elegans start sites (M. Perry, G. Hertz, and W. Wood, pers. comm.). The homeodomain is underlined. Open circles (o) indicate mutations found in *egl-5* mutants. After position 512, a 7 bp insertion (AAAACAA) was found in *egl-5(u202)*. This insertion results in a stop codon immediately downstream. At position 576, a C to T transition was found in alleles *n486* and *e2495*. This change results in an Arg to Cys substitution. In the *egl-5-lacZ* fusion, the *lacZ* gene was inserted at the ApaI site at position 344. It is not certain that the 5' end of the cDNA is full-length. It is likely, however, that this sequence includes the beginning of the protein because there is an in-frame stop codon 21 base pairs upstream of the first AUG, and there is no putative splice acceptor in between.

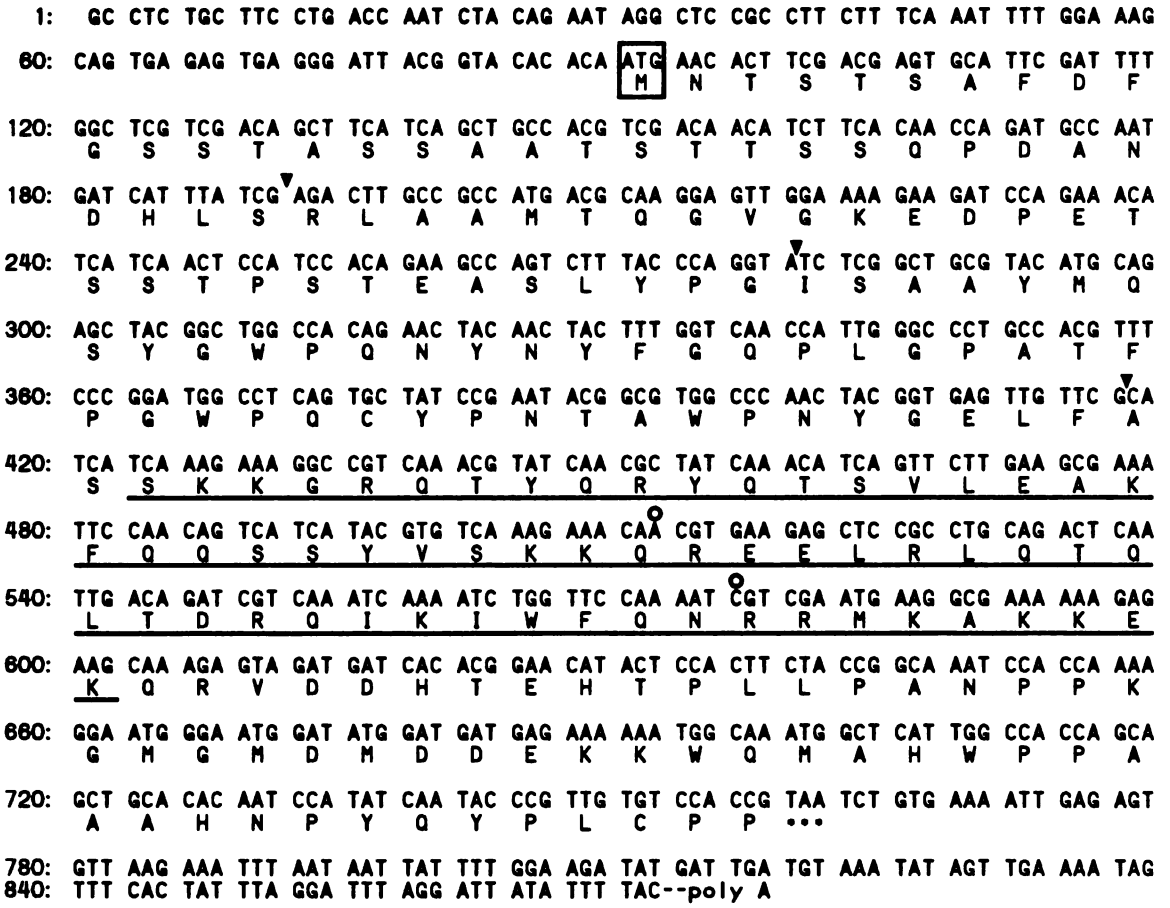

1: GC CTC TGC TTC CTG ACC AAT CTA CAG AAT AGG CTC CGC CTT CTT TCA AAT TTT GGA AAG
60: CAG TGA GAG TGA GGG ATT ACG GTA CAC ACA ATG AAC ACT TCG ACG AGT GCA TTC GAT TTT
M N T S T S A F D F
120: GGC TCG TCG ACA GCT TCA TCA GCT GCC ACG TCG ACA ACA TCT TCA CAA CCA GAT GCC AAT
G S S T A S S A A T S T T S S Q P D A N
180: GAT CAT TTA TCG AGA CTT GCC GCC ATG ACG CAA GGA GTT GGA AAA GAA GAT CCA GAA ACA
D H L S R L A A M T Q G V G K E D P E T
240: TCA TCA ACT CCA TCC ACA GAA GCC AGT CTT TAC CCA GGT ATC TCG GCT GCG TAC ATG CAG
S S T P S T E A S L Y P G I S A A Y M Q
300: AGC TAC GGC TGG CCA CAG AAC TAC AAC TAC TTT GGT CAA CCA TTG GGC CCT GCC ACG TTT
S Y G W P Q N Y N Y F G Q P L G P A T F
360: CCC GGA TGG CCT CAG TGC TAT CCG AAT ACG GCG TGG CCC AAC TAC GGT GAG TTG TTC GCA
P G W P Q C Y P N T A W P N Y G E L F A
420: TCA TCA AAG AAA GGC CGT CAA ACG TAT CAA CGC TAT CAA ACA TCA GTT CTT GAA GCG AAA
S S K K G R Q T Y Q R Y Q T S V L E A K
480: TTC CAA CAG TCA TCA TAC GTG TCA AAG AAA CAA CGT GAA GAG CTC CGC CTG CAG ACT CAA
F Q Q S S Y V S K K Q R E E L R L Q T Q
540: TTG ACA GAT CGT CAA ATC AAA ATC TGG TTC CAA AAT CGT CGA ATG AAG GCG AAA AAA GAG
L T D R Q I K I W F Q N R R M K A K K E
600: AAG CAA AGA GTA GAT GAT CAC ACG GAA CAT ACT CCA CTT CTA CCG GCA AAT CCA CCA AAA
K Q R V D D H T E H T P L L P A N P P K
660: GGA ATG GGA ATG GAT ATG GAT GAT GAG AAA AAA TGG CAA ATG GCT CAT TGG CCA CCA GCA
G M G M D M D D E K K W Q M A H W P P A
720: GCT GCA CAC AAT CCA TAT CAA TAC CCG TTG TGT CCA CCG TAA TCT GTG AAA ATT GAG AGT
A A H N P Y Q Y P L C P P •••
780: GTT AAG AAA TTT AAT AAT TAT TTT GGA AGA TAT GAT TGA TGT AAA TAT AGT TGA AAA TAG
840: TTT CAC TAT TTA GGA TTT AGG ATT ATA TTT TAC--poly A

Figure 2. Similarities between nematode, fly, and mouse HOM genes

The homeodomains, adjacent residues and hexapeptide regions of the C. elegans HOM genes are compared to those Drosophila HOM genes and mouse paralogs that have the highest overall amino acid identity within these regions. In the case of *egl-5/ceh-11*, *Hoxd-11* had the highest level of identity within the homeodomain itself; when adjacent regions were considered, several mouse *Abd-B*-class paralogs had similar levels of identity. Dashes indicate identity with *Antp*. Identical residues between the nematode and fly or mouse genes are boxed. Highlighted residues Gln-6 and Thr-7 in *Antp*, *Ubx*, and *abd-A* homologs appear to be common only to these classes. The precise distance between the hexapeptide regions and the homeodomains is not conserved; some amino acids have been omitted. Homologies outside of the regions shown were not detected. Recently, *ceh-13* has been found to contain a hexapeptide region that is most similar to the *labial* homologs (C. Wittmann and H. Tobler, pers. comm.). *egl-5* contains the residues YP upstream of the homeodomain. Whether or not this is a diverged hexapeptide motif is not known. The sequences (renamed according to Scott, 1992) have been tabulated from the following references: *ceh-13* (Schaller et al., 1990); *mab-5* (Costa et al., 1988); *lab* (Diederich et al., 1989); *pb* (Cribbs et al., 1992); *Dfd* (Regulski et al., 1987); *Scr* (LeMotte et al., 1989); *Antp* (Laughon et al., 1986); *Ubx* (Kornfeld et al., 1989); *abd-A* (Karch et al., 1990); *Abd-B* (Celniker et al., 1989); *Hoxb-1* (Frohman et al., 1990); *Hoxa-4* (Galliot et al., 1989); *Hoxa-2* (Tan et al., 1992); *Hoxa-5* (Fibi et al., 1988); *Hoxb-6* (Schughart et al., 1988); *Hoxb-7* (Meijlink et al., 1987); *Hoxb-8* (Kongsuwan et al., 1989); *Hoxd-11*(Izpisúa-Belmonte et al., 1991).

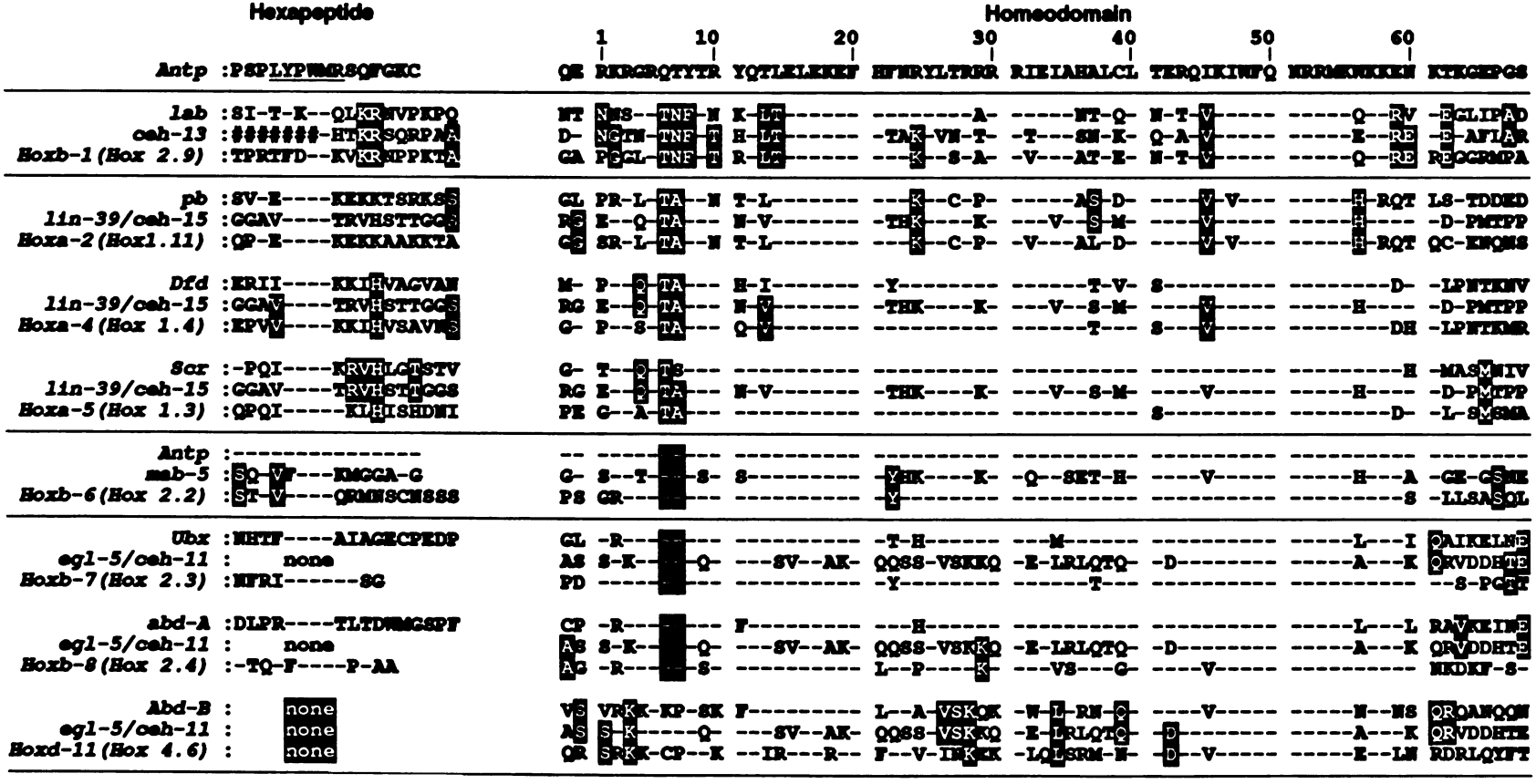
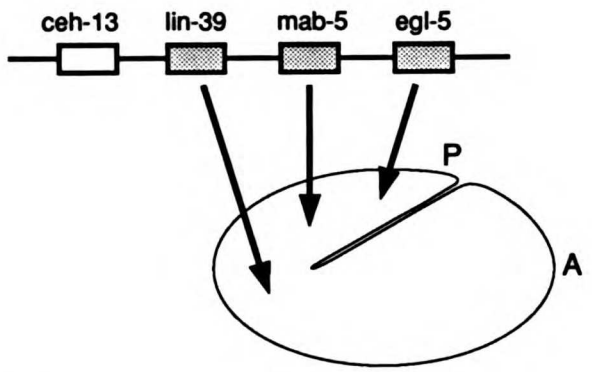

Hexapeptide
Homeodomain
1
10
20
30
40
50
60
Antp
lab
ceh-13
Hoxb-1(Hox 2.9)
pb
lin-39/ceh-15
Hoxa-2(Hox1.11)
Dfd
lin-39/ceh-15
Hoxa-4(Hox 1.4)
Scr
lin-39/ceh-15
Hoxa-5(Hox 1.3)
Antp
mab-5
Hoxb-6(Hox 2.2)
Ubx
egl-5/ceh-11
Hoxb-7(Hox 2.3)
abd-A
egl-5/ceh-11
Hoxb-8(Hox 2.4)
Abd-B
egl-5/ceh-11
Hoxd-11(Hox 4.6)

Figure 3. Embryonic expression pattern of *lin-39-*, *mab-5-*, and *egl-5-lacZ* fusion genes

The schematic drawing shows a 2-fold embryo (~460 min post first cleavage) with the anterior (A) and posterior (P) ends indicated. Below are 2-fold embryos that carry the extrachromosomal *lin-39-*, or *egl-5-lacZ* constructs. An integrated *mab-5-lacZ* fusion, which has been described previously (Salser and Kenyon, 1992; Cowing and Kenyon, 1992), is shown for comparison. The majority of embryos (85%) exhibited the expression patterns shown or stained with slightly lower intensity. The remaining 15% of embryos showed the same localized staining with slightly extended domains. Localized expression of the fusion genes was also present in earlier and later stage embryos.


ceh-13
lin-39
mab-5
egl-5
P
A

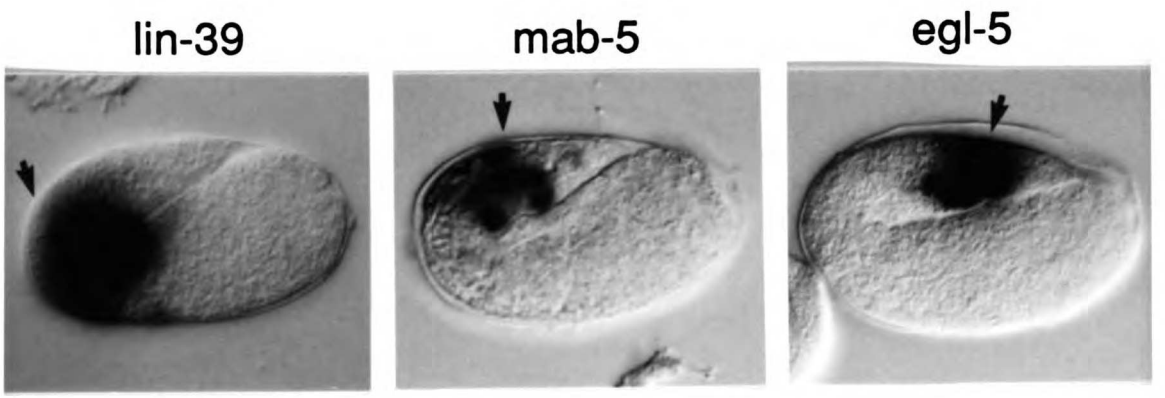

lin-39
mab-5
egl-5

Figure 4. Expression pattern of an *egl-5-lacZ* fusion gene

A. Diagram of an L1 animal (male and hermaphrodite combined) showing the cells that are affected by *egl-5* mutations at this stage. (Chisholm, 1991). The HSN is born in the tail region and migrates to the central body region during embryogenesis.

B. A 0-2.5 hr L1 hermaphrodite larva showing expression of the *egl-5-lacZ* fusion gene in the posterior body region and tail. The majority of L1 larvae (>95%) of both sexes had the expression pattern shown or stained with lower intensity. The number of cells staining varied from animal to animal. Among the cells that often expressed *egl-5-lacZ* at L1 were HSN (arrow), B, F, Y, U and K, (S. Salser, C. Hunter, and C. Kenyon, pers. comm.) which are cells affected by *egl-5* mutations . Older larvae also display posterior specific expression. Cells outside this body region rarely express the fusion gene.

C. Close-up view of an adult male tail showing the intense position specific expression of the *egl-5-lacZ* fusion. In the majority of adults, the fusion is strongly expressed in the posterior body region and tail of both sexes. Bar = 10 μm.

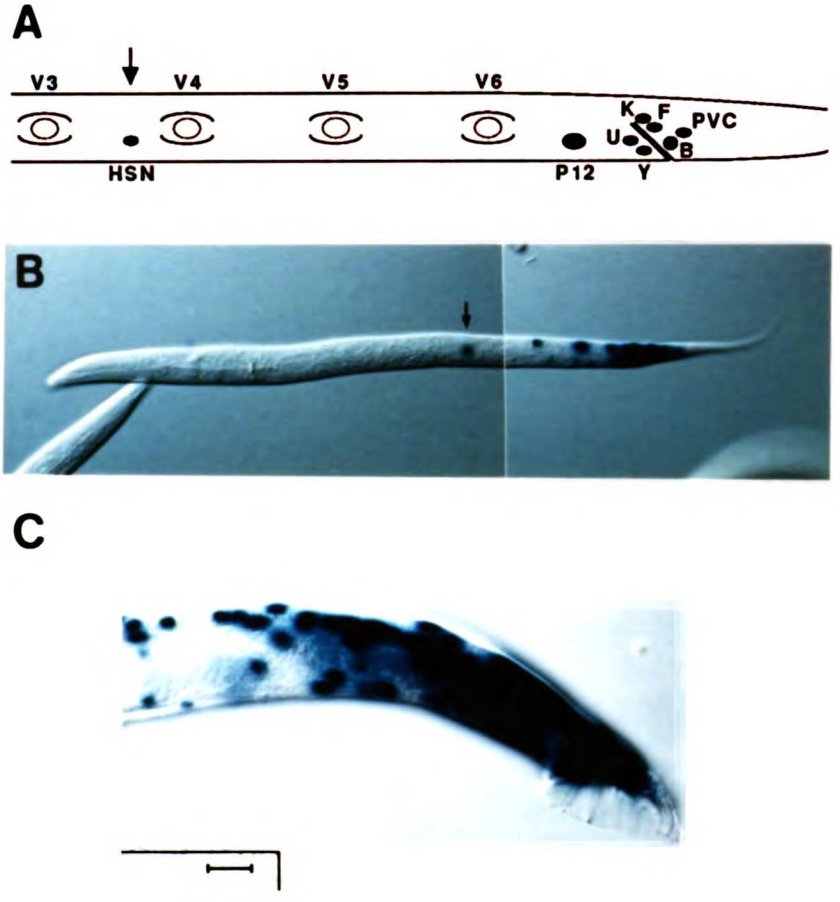

A
V3
V4
V5
V6
K
F
PVC
U
B
HSN
P12
Y
B
C

Figure 5. Nucleotide and predicted amino acid sequences of *lin-39*

The sequence, based on cDNA and genomic clones, predicts a protein of 253 amino acids. Potential translation start sites are boxed. The positions of introns are indicated by closed arrowheads (▼). The homeodomain is underlined. The open arrowhead (▽) indicates the position where a nonsense mutation (TGA) was made in a 10 kb DNA fragment that rescues *lin-39* (subclone ΔNco4). The 1 kb wild-type genomic fragment that was coinjected with ΔNco4 includes the coding sequence from position 406 to 583. The open circle (○) at position 740 indicates the mutation found in the *lin-39(mu26)* allele. This mutation is a G to A transition that changes the Trp residue to a stop codon. In the *lin-39-lacZ* fusion, the *lacZ* gene was inserted at the NcoI site at position 553. This cDNA is likely to be full length or near full length because a message of similar size (approximately 1.3 kb) was observed in all developmental stages using Northern analysis (data not shown), and because the smallest genomic fragment that rescues *lin-39(mu26)* extends only about 360 bp upstream of the cDNA.

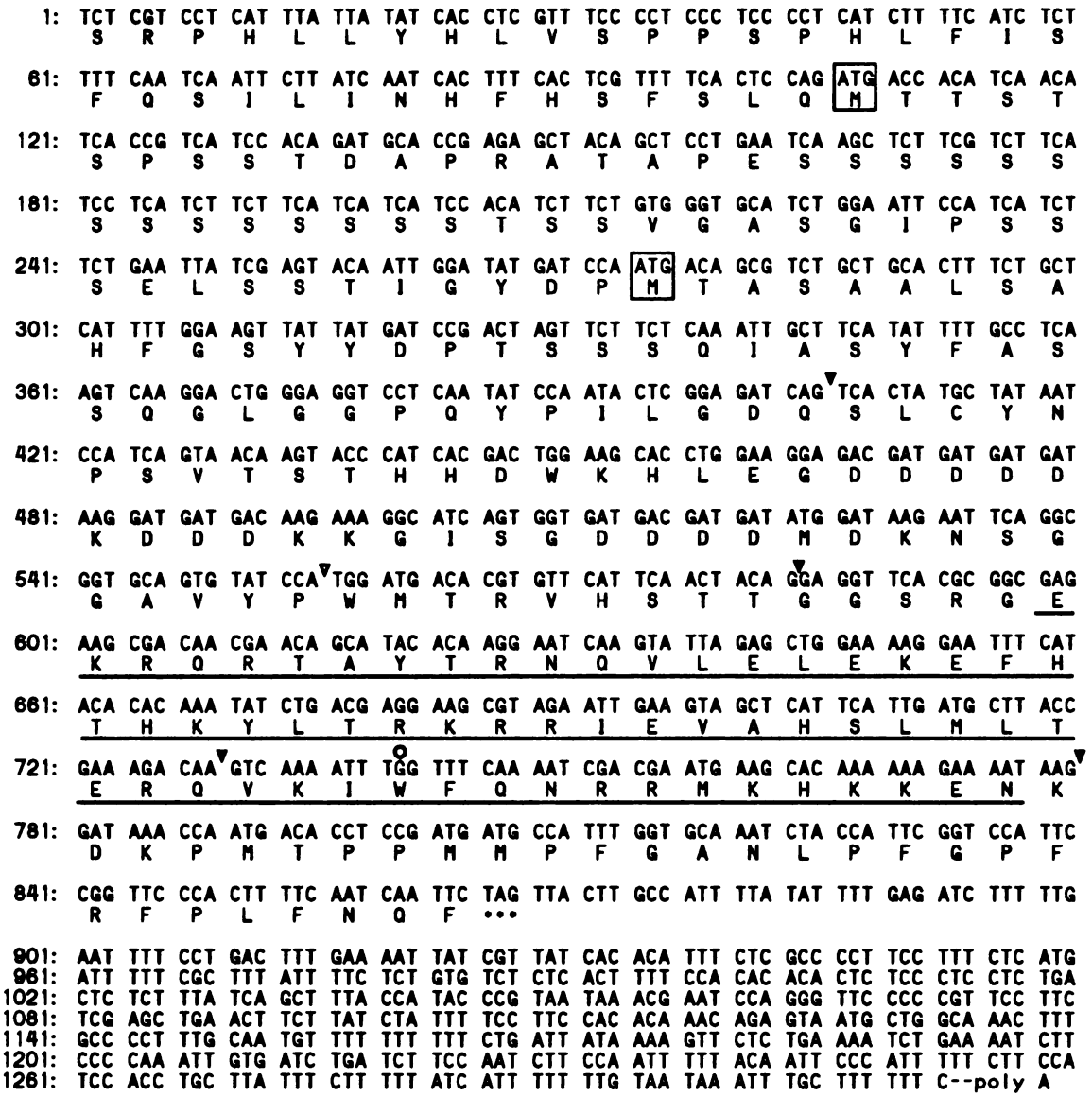
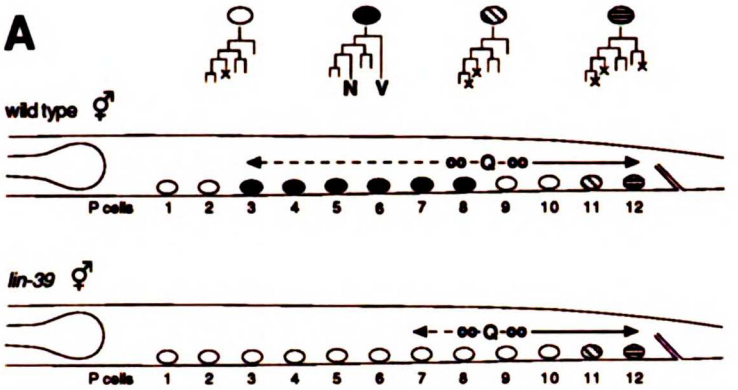

1: TCT CGT CCT CAT TTA TTA TAT CAC CTC GTT TCC CCT CCC TCC CCT CAT CTT TTC ATC TCT
S R P H L L Y H L V S P P S P H L F I S
61: TTT CAA TCA ATT CTT ATC AAT CAC TTT CAC TCG TTT TCA CTC CAG ATG ACC ACA TCA ACA
F Q S I L I N H F H S F S L Q M T T S T
121: TCA CCG TCA TCC ACA GAT GCA CCG AGA GCT ACA GCT CCT GAA TCA AGC TCT TCG TCT TCA
S P S S T D A P R A T A P E S S S S S S
181: TCC TCA TCT TCT TCA TCA TCA TCC ACA TCT TCT GTG GGT GCA TCT GGA ATT CCA TCA TCT
S S S S S S S S T S S V G A S G I P S S
241: TCT GAA TTA TCG AGT ACA ATT GGA TAT GAT CCA ATG ACA GCG TCT GCT GCA CTT TCT GCT
S E L S S T I G Y D P M T A S A A L S A
301: CAT TTT GGA AGT TAT TAT GAT CCG ACT AGT TCT TCT CAA ATT GCT TCA TAT TTT GCC TCA
H F G S Y Y D P T S S S Q I A S Y F A S
361: AGT CAA GGA CTG GGA GGT CCT CAA TAT CCA ATA CTC GGA GAT CAG TCA CTA TGC TAT AAT
S Q G L G G P Q Y P I L G D Q S L C Y N
421: CCA TCA GTA ACA AGT ACC CAT CAC GAC TGG AAG CAC CTG GAA GGA GAC GAT GAT GAT GAT
P S V T S T H H D W K H L E G D D D D D
481: AAG GAT GAT GAC AAG AAA GGC ATC AGT GGT GAT GAC GAT GAT ATG GAT AAG AAT TCA GGC
K D D D K K G I S G D D D D M D K N S G
541: GGT GCA GTG TAT CCA TGG ATG ACA CGT GTT CAT TCA ACT ACA GGA GGT TCA CGC GGC GAG
G A V Y P W M T R V H S T T G G S R G E
601: AAG CGA CAA CGA ACA GCA TAC ACA AGG AAT CAA GTA TTA GAG CTG GAA AAG GAA TTT CAT
K R Q R T A Y T R N Q V L E L E K E F H
661: ACA CAC AAA TAT CTG ACG AGG AAG CGT AGA ATT GAA GTA GCT CAT TCA TTG ATG CTT ACC
T H K Y L T R K R R I E V A H S L M L T
721: GAA AGA CAA GTC AAA ATT TGG TTT CAA AAT CGA CGA ATG AAG CAC AAA AAA GAA AAT AAG
E R Q V K I W F Q N R R M K H K K E N K
781: GAT AAA CCA ATG ACA CCT CCG ATG ATG CCA TTT GGT GCA AAT CTA CCA TTC GGT CCA TTC
D K P M T P P M M P F G A N L P F G P F
841: CGG TTC CCA CTT TTC AAT CAA TTC TAG TTA CTT GCC ATT TTA TAT TTT GAG ATC TTT TTG
R F P L F N Q F •••
901: AAT TTT CCT GAC TTT GAA AAT TAT CGT TAT CAC ACA TTT CTC GCC CCT TCC TTT CTC ATG
961: ATT TTT CGC TTT ATT TTC TCT GTG TCT CTC ACT TTT CCA CAC ACA CTC TCC CTC CTC TGA
1021: CTC TCT TTA TCA GCT TTA CCA TAC CCG TAA TAA ACG AAT CCA GGG TTC CCC CGT TCC TTC
1081: TCG AGC TGA ACT TCT TAT CTA TTT TCC TTC CAC ACA AAC AGA GTA ATG CTG GCA AAC TTT
1141: GCC CCT TTG CAA TGT TTT TTT TTT CTG ATT ATA AAA GTT CTC TGA AAA TCT GAA AAT CTT
1201: CCC CAA ATT GTG ATC TGA TCT TCC AAT CTT CCA ATT TTT ACA ATT CCC ATT TTT CTT CCA
1261: TCC ACC TGC TTA TTT CTT TTT ATC ATT TTT TTG TAA TAA ATT TGC TTT TTT C--poly A

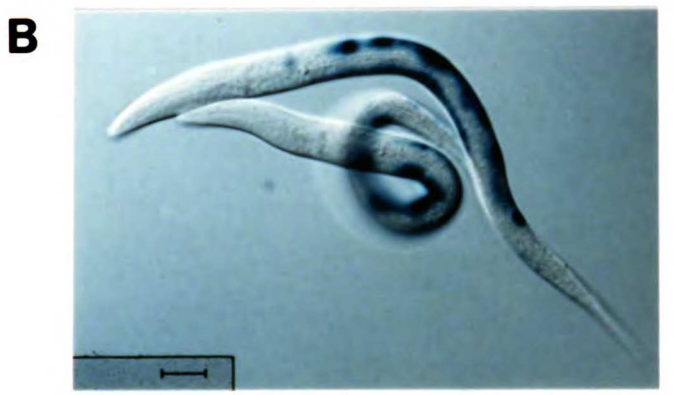
Figure 6. Expression pattern of a *lin-39-lacZ* fusion gene

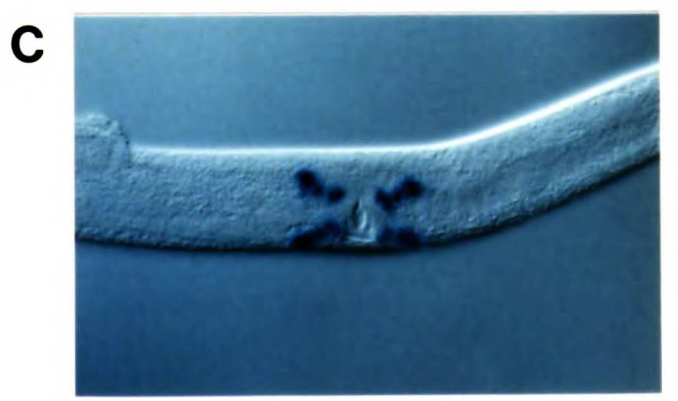
A. Diagram of an L1 hermaphrodite larva showing the fates of cells affected in *lin-39* mutants. The vulva precursor cells are P(3-8).p. The dotted arrows show the extent of migration of QR descendants (which occur on the right side of the animal) while the solid arrows indicate the extent of migration of QL descendants (which occur on the left side). Only QR descendants are affected by *lin-39* mutations. Not all cells affected by *lin-39* mutations are shown. Wild type lineages are from Sulston and Horvitz (1977). This diagram uses data provided by H. Ellis (1985), S. Clark, H. Horvitz, J. Austin, N.T. Robinson, and C. Kenyon (pers. comm.). N = neuronal fate; V = vulval fate; X = cell death.

B. A 0-2.5 hr larva showing localized expression of the *lin-39-lacZ* fusion gene in the mid-body region. The majority of L1 larvae (>95%) showed the above staining pattern or the staining was less intense. The number of cells staining varied from animal to animal. Among the cells that express the *lin-39-lacZ* fusion are P(3-8) and the QR neuroblast (N. T. Robinson and S. Salser, pers. comm.). Staining was also present in descendants of these P and Q cells. It is not clear whether this reflected continued transcription or perdurance of β-galactosidase activity. A few cells outside this body region sometimes express the fusion gene. Older larvae also express the fusion gene in the same position specific manner. Bar = 16 μm.

C, D. Close-up view of *lin-39-lacZ* expression in the vulval muscles (C) and possibly the epithelial D cells (D) surrounding the vulva in two adult hermaphrodites (anterior to the left and up). In the majority of adult hermaphrodites, the fusion gene was expressed only in the central body region.


A
N V
wild type ♂
oo-Q-oo
P cells 1 2 3 4 5 6 7 8 9 10 11 12
lin-39 ♂
oo-Q-oo
P cells 1 2 3 4 5 6 7 8 9 10 11 12


B


C


D

Figure 7. Physical map of the C. elegans HOM cluster

The general locations of *ceh-13*, *lin-39*, *mab-5*, *egl-5*, and *ceh-23* are indicated. The entire known cluster (200-300 kb) and surrounding regions (200-300 kb on either side) were probed with the degenerate oligonucleotide HB-1 (Bürglin et al., 1989). Cosmids (red) and YACs (blue) that were probed are indicated. Other than the above genes no new homeoboxes were detected. Each cosmid is about 35-40 kb. The complete nature of the cluster will soon be elucidated when C. elegans genome sequencing project determines the DNA sequence of this region.

LGIII cluster

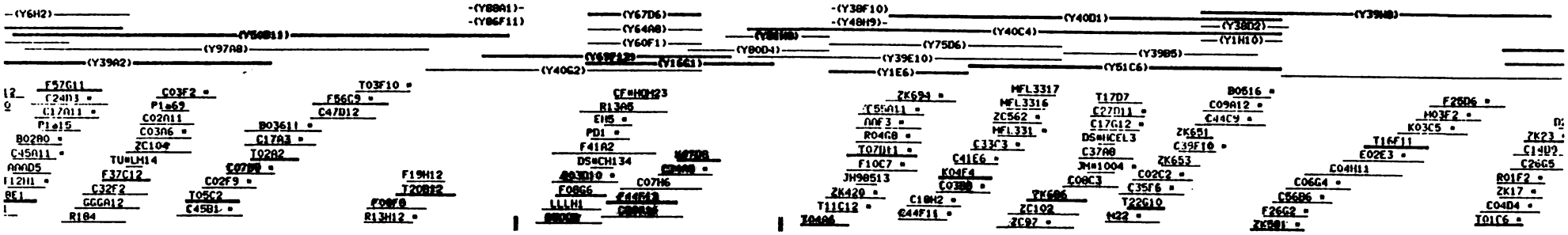


Figure 8. *ceh-23* sequence comparisons and expression pattern of a *ceh-23-lacZ* fusion

A. The *ceh-23* homeodomain is compared to the Drosophila genes *ems* and *Distal-less(Dll)* and the murine homologs *Emx1*, *Emx2*, *Dlx-1*, and *Dlx-2*. Dashes indicate amino acid identity. It is possible that there are additional C. elegans *ems*-like genes. One homeobox, *ceh-2*, has limited homology to *ems* (Bürglin et al., 1989). Sequences have been tabulated from the following sources: *ems* (Dalton et al., 1989); *Emx1* and *Emx2* (Simeone et al., 1992); *Dll* (Vachon et al., 1992); *Dlx-1* (Price et al., 1991); *Dlx-2* (Porteus et al., 1991).

B. A 0-2.5 hr larva showing expression of the *ceh-23-lacZ* fusion, primarily in the amphid sensilla (arrow), which are olfactory neurons in the nerve ring, and also in a bilateral pair of CAN neurons (arrowhead), which run along the excretory canal of the animal and are thought to have an osmoregulatory function (C. Bargmann, pers. comm). Among the amphid neurons that often stain are the olfactory AWC neurons (which mediate chemosensation to volatile odorants) and the putative chemosensory ADL neurons. A similar staining pattern was seen in older larvae and adults (data not shown). Approximately 24% of L1 animals examined had the staining pattern shown. In about 71%, only a subset of these neurons were staining, or staining was less intense. In about 5% of animals expressing the fusion, additional staining was observed in cells along the body.

A

```
                1         10          20          30          40          50          60
                |         |           |           |           |           |           |
      ems  : PKRIRTAFSP SQLLKLEHAF ESNQYVVGAE RKALAQNLNL SETQVKVWFQ NRRTKHKRMQ Q
  ceh-23   : HRKA--IYGT T-TQQ--DM- KGQM------ -EN---R-G- -PS--RI--- ---S--R-K- -
     Emx1  : ---------- ----R--R-- -K-H------ --Q--GS-S- ---------- -----Y--QK L
     Emx2  : ---------- ----R----- -K-H------ --Q--HS-S- T--------- -----F--QK L

     Dll   : MRKPRTIYSS LQLQQLNRRF QRTQYLALPE RAELAASLGL TQTQVKIWFQ NRRSKYKKMM K
  ceh-23   : H--A----GT T-T---EDM- KGQM-VVGA- -EN--QR--- SPS--R---- -----HRRKQ Q
    Dlx-1  : I--------- ----A----- -Q-------- ---------- ---------- -K---F--L- -
    Dlx-2  : V--------- F--AA-Q--- -K-------- ---------- ---------- -----F---W -
```

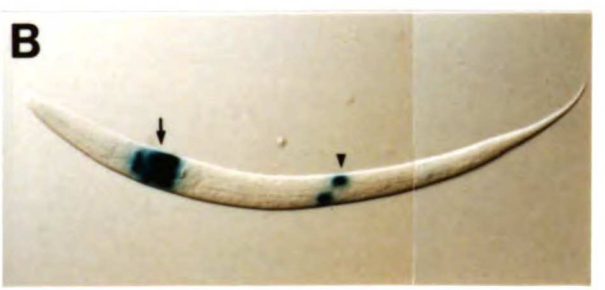


Figure 9. Nucleotide and predicted amino acid sequences of *ceh-23*

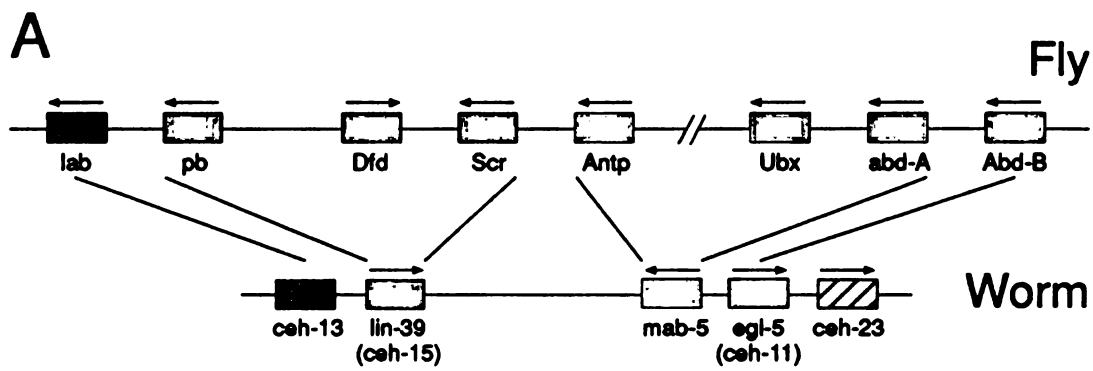
The sequence, based on cDNA and genomic clones, predicts a protein of at least 303 amino acids. Positions of introns are indicated by arrowheads (▼). The homeodomain is underlined. In the *ceh-23-lacZ* fusion, the *lacZ* gene was joined at the SnaBI site at position 697. It is not certain that the 5' end of the cDNA is full-length because there is a putative splice acceptor site just upstream in the genomic sequence.

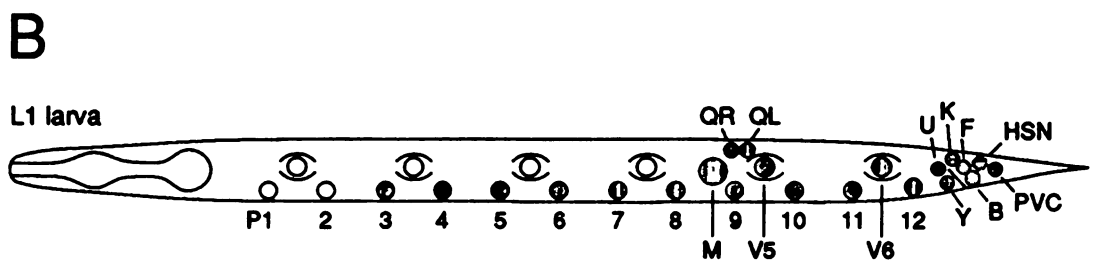
```
   1: ACC CAT CTT CCA TTC CAA ACT TTG CCA GTT TCA ACG CCG CTG CCT GTG TCT TCA TCA TCT
       T   H   L   P   F   Q   T   L   P   V   S   S   T   P   L   P   V   S   S   S   S

  61: CTA ACA GAC GTT CTT CAA ACA ATT GCT GCT TTA CAA GCC TGT CCC ACA TCC TGT ATA CCT
       L   T   D   V   L   Q   T   I   A   A   L   Q   A   C   P   T   S   C   I   P

 121: TCA ACA TCA ACT GGA ATG CTT TCA CCG AAT CTT CCC TTC TCG GCT ACA ATT CCA CGT GTC
       S   T   S   T   G   M   L   S   P   N   L   P   F   S   A   T   I   P   R   V

 181: AAC TTG TTC CCA CCA TCA CAA CCA GCC AAT TCT TTA ATA CTG CCA ACA ATT CCA GCA CAG
       N   L   F   P   P   S   Q   P   A   N   S   L   I   L   P   T   I   P   A   Q

 241: CCT TTT ATT CCA AAT CCA TCT CTC CTG CAA GCC AAT CCA TCT GCT GTC GAA GCT CTT GCA
       P   F   I   P   N   P   S   L   L   Q   A   N   P   S   A   V   E   A   L   A

 301: AAT GCA CTT TTT GCA ACA ACA AGC CGT CGG GCT TCT TGC CCA GAA CCT CCT GCA TCC TCT
       N   A   L   F   A   T   T   S   R   R   A   S   C   P   E   P   P   A   S   S

 361: CAA GCA ACC GTC ACG CTT CAA GTC CCA TCT ACG GGT TCT CCA GAA CGA CGA CGT TAT TCA
       Q   A   T   V   T   L   Q   V   P   S   T   G   S   P   E   R   R   R   Y   S

 421: GAA ACA AAT ATG GAG GTT CTG CTA CGT GAA CAA CTT GCA CAA TTG ATG CCT CCA ACC TCT
       E   T   N   M   E   V   L   L   R   E   Q   L   A   Q   L   M   P   P   T   S

 481: CAA CTA CCC GGA ATG CCT GGA TGT TAT TAT CAA CAT GTT CCA GCT GCT GGA ACT TCA GGA
       Q   L   P   G   M   P   G   C   Y   Y   Q   H   V   P   A   A   G   T   S   G

 541: ATT CAA GGA TCA TTA GAT GCG GCG TTA ATG GGA GCT GTT CCC TTG GCT ATG AAT AGT ATG
       I   Q   G   S   L   D   A   A   L   M   G   A   V   P   L   A   M   N   S   M

 601: GCA CAT AGT CGG AGA GCA GCA AAT CAT AGA AAA GCA AGA ACA ATC TAC GGT ACA ACA CAA
       A   H   S   R   R   A   A   N   H   R   K   A   R   T   I   Y   G   T   T   Q

 661: ACT CAA CAA CTG GAA GAT ATG TTT AAA GGA CAA ATG TAC GTA GTC GGA GCA GAA CGA GAA
       T   Q   Q   L   E   D   M   F   K   G   Q   M   Y   V   V   G   A   E   R   E

 721: AAT CTT GCT CAA CGC CTT GGC CTA TCA CCT TCT CAA GTT AGA ATT TGG TTT CAA AAC CGG
       N   L   A   Q   R   L   G   L   S   P   S   Q   V   R   I   W   F   Q   N   R

 781: AGA AGC AAA CAT CGT AGG AAA CAA CAA GAG GAG CAA CAG AGC ACG ACG TTA GAA GAA AAA
       R   S   K   H   R   R   K   Q   Q   E   E   Q   Q   S   T   T   L   E   E   K

 841: TCT GAA GAA ATT GGA AAA GAT GAA GAG GAA GAT GAT GAA GAA GAT GAA GAT GAT GTA AAA
       S   E   E   I   G   K   D   E   E   E   D   D   E   E   D   E   D   D   V   K

 901: GTT CTG AAT TAA TTT AAT TTT TTT TTT CGT GTT TAC TTG TTT ATT TCC TGC CAT ATT CTG
       V   L   N  ...

 961: TTT TAG AAA ATA ATT AGA AAA AAT GTG TAT ATT CTT GAT TAT AAA GCA TTA AAC AAA TAA
1021: AGT TTT ATG TGA ACT T--poly A
```

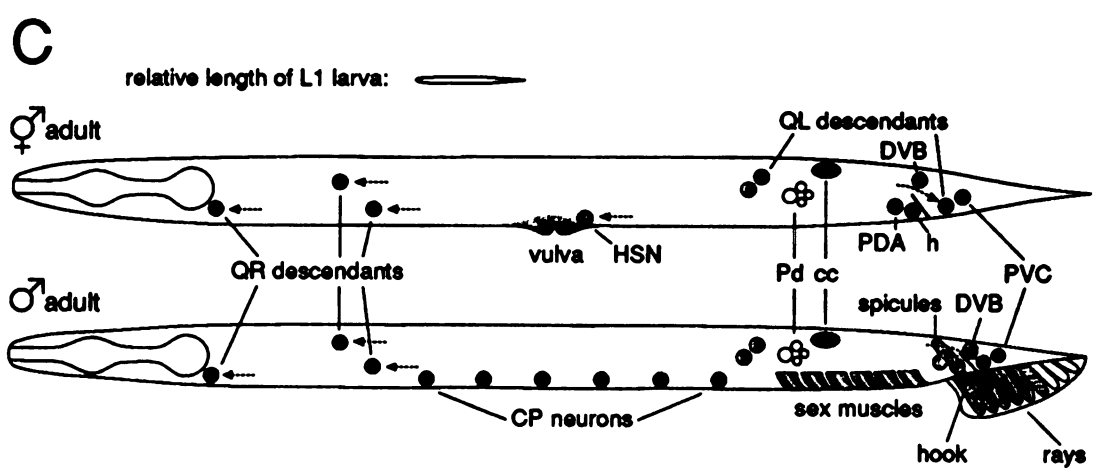
Figure 10. Structural and functional organization of the C. elegans HOM-C

A. Possible evolutionary relationships, inferred only from the level of amino acid identity in and around the homeodomain, and from the lack of a hexapeptide region in *egl-5*, are indicated by dashed lines. The tightly linked *ceh-13* and *lin-39* genes are separated by an estimated 200-300 kb from the tightly linked *mab-5*, *egl-5* and *ceh-23* genes. (See Figure 7). The *ceh-23* (*ems*/*Dll*-like) homeobox is included because its position in the C. elegans and human clusters appears to be conserved (see Discussion). As in Drosophila, a number of non-HOM of genes have also been found to map in the C. elegans cluster. The directions of transcription were inferred from DNA sequence and restriction analysis. The direction of transcription of *mab-5* was determined by S. Salser. Unlike in Drosophila and vertebrates, the direction of transcription in C. elegans is not conserved.

B. Diagram of a newly hatched L1 (male and hermaphrodite combined), showing cells that require the function of *lin-39* (green), *mab-5* (red), and *egl-5* (blue) to develop correctly. Cells that require the function of more than one HOM gene are color-coded for those genes. The HSN is drawn in its birthplace in the tail. These cells give rise to the structures shown in the adult (C), again color-coded to represent the HOM-C gene(s) required for their development. cc = coelomocyte; PVC, PDA, DVB, and the Q descendants are neurons; CP = serotonergic neurons; h = hypodermal cell. Structures only rarely affected by a HOM-C gene ($\leq$ 10%) are not color-coded for this gene. Considerable cell migration takes place during development (arrows), so that structures generated by different HOM-C genes become intermingled. The migration of the ray precursor cells (not shown) begins from the position of V(5,6) and moves into the tail. Not all structures and cell types specified by the HOM-C genes are shown but most of these additional structures arise within the primary domains of the HOM-C genes whose function they require. One exception, the male gonad (not shown), develops in the central/posterior body region in an *egl-5*-dependent fashion. A few of the anteroposterior structures, such as the three most posterior T-derived rays and the postdeirid (Pd), are not affected by *lin-39*, *mab-5*, or *egl-5* mutations. Compiled from data in Wang et al., in press; Loer and Kenyon, in press; Ellis, 1985; Kenyon, 1986; and Chisholm, 1991. This figure was made with significant assistance from M. Müller-Immergluck.


A
Fly
lab
pb
Dfd
Scr
Antp
Ubx
abd-A
Abd-B
Worm
ceh-13
lin-39
(ceh-15)
mab-5
egl-5
(ceh-11)
ceh-23


B
L1 larva
QR QL
U K F HSN
PVC
Y B
P1 2 3 4 5 6 7 8 9 10 11 12
M V5 V6


C
relative length of L1 larva:
♂♀adult
QL descendants
DVB
PDA h
vulva HSN
QR descendants
Pd cc
PVC
♂adult
spicules DVB
CP neurons
sex muscles
hook
rays