Developmental Cell, Vol. 4, 903–915, June, 2003, Copyright ©2003 by Cell Press

Global Regulation of Hox Gene Expression in *C. elegans* by a SAM Domain Protein

Hong Zhang,[1,2,*] Ricardo B.R. Azevedo,[1,3] Robyn Lints,[1] Christina Doyle,[1] Yingqi Teng,[1] Daniel Haber,[2] and Scott W. Emmons[1,*]
[1]Department of Molecular Genetics
Albert Einstein College of Medicine
Bronx, New York 10461
[2]Massachusetts General Hospital Cancer Center and
Harvard Medical School
Charlestown, Massachusetts 02129

Summary

Polycomb group (PcG)-mediated repression of *C. elegans* Hox genes has not been demonstrated, and genes homologous to components of one of the PcG complexes (PRC1) have not been identified in the *C. elegans* genome. We find that a mechanism of general Hox gene repression exists in *C. elegans*, carried out in part by SOP-2, a protein related to, but not orthologous with, any PcG protein. *sop-2* mutations lead to widespread ectopic expression of Hox genes and homeotic transformations. SOP-2 contains a SAM domain, a self-associating protein domain found in other repressors, including a core component of PRC1 and ETS transcription factors. Phylogenetic analysis indicates that this domain is more closely related to those of the ETS family than to those of PcG proteins. The results suggest that global repression of Hox genes has been taken over by a different branch of the SAM domain family during the evolution of nematodes.

Introduction

Gene transcription in eukaryotes can be regulated at multiple steps of the transcription initiation process. Hox genes, encoding conserved homeobox-containing transcription factors, appear to rely heavily on a particular form of chromatin regulation brought about by *Polycomb* group (PcG) genes (Francis and Kingston, 2001; Simon and Tamkun, 2002). Hox genes are expressed in distinct domains along the body axis and act to give cells of diverse tissues their unified regional cell identities. Absence or ectopic expression of Hox genes causes homeotic transformations, involving duplication or loss of body structures (Gellon and McGinnis, 1998). In mutants of PcG genes, Hox genes are expressed in tissues or body regions where their expression normally does not occur, resulting in widespread homeotic transformations. PcG genes encode components of protein complexes that block chromatin remodeling and lead to a repressive chromatin structure. The mutant phenotype of PcG genes indicates that, for Hox genes, if chromatin remodeling is not blocked, positive factors are present in many cells that will initiate transcription outside the normal expression domains.

PcG regulation of Hox genes is an ancient mechanism, conserved in *Drosophila* and vertebrates, but, surprisingly, it has not been described in the nematode *Caenorhabditis elegans*. Two distinct PcG complexes have been identified, ESC-E(Z) and PRC1 (Satijn and Otte, 1999; Francis and Kingston, 2001; Simon and Tamkun, 2002). Only genes encoding components of ESC-E(Z) are found in *C. elegans* (e.g., *mes-2* and *mes-6*) (Holdeman et al., 1998; Korf et al., 1998). *C. elegans* appears to lack any homologs of the core components of PRC1, Polycomb (PC), polyhomeotic (PH), Posterior sex combs (PSC), or the PRC1-associated protein Sex combs on midleg (SCM) (Ruvkun and Hobert, 1998). Moreover, the *C. elegans* ESC-E(Z) PcG genes, which have an important role in regulating germ cell development, have not been reported to repress Hox genes in somatic tissues (Holdeman et al., 1998; Korf et al., 1998; see Ross and Zarkower, 2003 [this issue of *Developmental Cell*]).

These observations have raised the question whether the regulation of Hox genes in *C. elegans*, known for its highly reproducible cell lineages, follows different mechanisms from those in other animals. Prior studies of *C. elegans* Hox gene regulation have tended to emphasize their uniqueness. Unlike in *Drosophila*, expression of the Hox genes *mab-5* and *ceh-13* (orthologs of *Drosophila ftz* and *labial,* respectively [Aboobaker and Blaxter, 2003]) was shown to be controlled during embryogenesis by lineage factors, rather than by region-specific signals (Cowing and Kenyon, 1996; Wittmann et al., 1997). In a postembryonic cell lineage, expression of *mab-5* cycles on and off repeatedly to specify diverse aspects of cell fate, in contrast to the idea of Hox genes as fixed markers of regional cell identity (Salser and Kenyon, 1996). In a variety of cell lineages, expression of Hox genes has been demonstrated to be regulated in a lineage-specific manner by a number of transcription factors and signaling pathways, including LET-60 Ras signaling, Wnt signaling, the NURD complex, and Caudal homolog PAL-1 (Maloof and Kenyon, 1998; Maloof et al., 1999; Ch'ng and Kenyon, 1999; Hunter et al., 1999), yet no factors with a general repressive function in many tissues and affecting all Hox genes have been identified. These observations have lead to the suggestion that lineal mechanisms substitute for regional and global ones in regulation of Hox genes in *C. elegans* (Duboule 1992; Ruvkun and Hobert, 1998).

Here we show that, on the contrary, *C. elegans* Hox genes are subject to a global repressive mechanism similar to that present in other organisms. This repressive mechanism involves a protein related to, but distinct from, PcG proteins. Our results show that a conserved transcriptional regulatory mechanism such as PcG-mediated repression is capable of undergoing major changes during the evolutionary divergence of animal phyla.

*Correspondence: zhangh@helix.mgh.harvard.edu (H.Z.), emmons@aecom.yu.edu (S.W.E.)
[3]Present address: Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204.

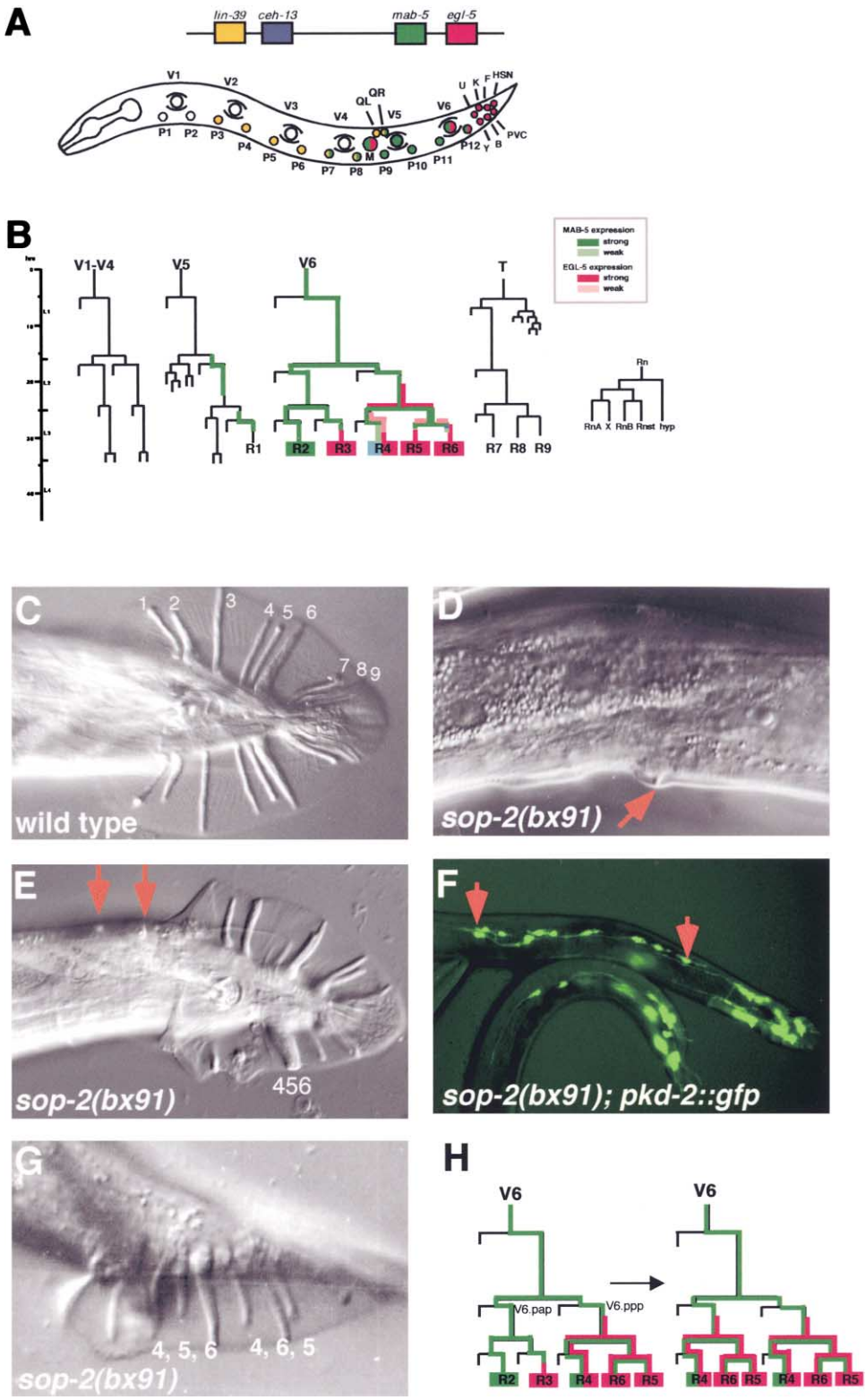


Figure 1. *sop-2* Is Required for Specification of Male Seam Lineage Fates

(A) Schematic diagram of Hox gene expression patterns. Genomic organization of *C. elegans* Hox gene cluster (top) and L1 stage animal showing blast cells that express *lin-39*, *mab-5*, and *egl-5* (bottom). The postembryonic expression pattern of *ceh-13* at this stage has not been fully characterized and is therefore not shown.

(B) Postembryonic lineage of V1–V6 and T during male development. The expression pattern of *mab-5* (green) and *egl-5* (red) in the V5 and V6 lineage is shown. Hours of postembryonic development and larval stage are shown to the left. Boxes at the ends of the V6 lineage branches represent the cells of the ray sublineage, which continue to express the proteins as shown. Each Rn cell undergoes the ray sublineage, which generates an A-type neuron (RnA), a B-type neuron (RnB), a structural cell (Rnst), one cell death (X), and one hypodermal cell (hyp).

Table 1. Seam Cell and Ventral Cord Defects in *sop-2* Mutants

	Average Number of Ectopic V1–V5 Rays	Average Number of *tph-1::gfp*-Positive Cells in the Ventral Cord per Animal
Wild-type	0	6
sop-2(bx91)	3.6	2.3 (range 0–10)
sop-2(bx91); mab-5(e1239)	0.35	ND
sop-2(bx91); egl-5(n486)	1.7	8.2 (range 7–11)
sop-2(bx91); mab-5(e1239) egl-5(n486)	0.17	ND
sop-2(bx91); lin-39	ND	0

n > 200 for all genotypes examined.

Results

sop-2 Maintains the Integrity of Domains of Hox Gene Activity

During postembryonic development in *C. elegans*, Hox genes are normally expressed in serial domains along the body axis, defining the region-specific differentiation patterns within various tissues (Figure 1A). In animals lacking function of the gene *sop-2*, these expression domains are greatly expanded. A mutation in *sop-2*, *bx91*, was isolated in a genetic screen involving development of rays in the male tail, described previously (Zhang and Emmons, 2000) (see Experimental Procedures). The *bx91* mutation is recessive and temperature sensitive: at 15°C, animals have no obvious defects, but, at 25°C, they are scrawny and arrest at early larval stages. At 20°C, *sop-2* animals are generally *lon*g (Lon), *unc*oordinated (Unc), and *m*ale *ab*normal (Mab) and show homeotic transformations.

We studied the homeotic transformations in the development of seam and ventral cord cells in detail, since the role of Hox genes in cell fate specification in these cell types is well characterized. Cell division patterns and differentiation within two bilateral rows of epidermal stem cells known as seam cells are specified by the Hox genes *mab-5* and *egl-5* (ortholog of *Drosophila Abdominal-B*). During normal male development, the three most-posterior seam cells, V5, V6, and T, produce nine pairs of sensory rays (Figures 1B–1C). *mab-5* is expressed in the V5 and V6 cell lineages and is required for these lineages to generate rays instead of the longitudinal cuticular ridges known as alae, formed by more-anterior seam cells (Kenyon et al., 1997). *egl-5* is expressed in the V6 lineage and is required for the development and differentiation of V6 rays (rays 2–6) (Ferreira et al., 1999).

The *sop-2(bx91)* mutation affects the development of all V seam cells. In *sop-2(bx91)* males, the anterior seam cells V1–V4 produce rays and fan-like cuticular structures in the anterior body region (Figure 1D; Table 1, line 2). Where rays are formed, there are gaps in the alae (data not shown). V5 also generates one or more ectopic rays in *sop-2* males (Figure 1E). In some cases, V5 appears to adopt the V6 fate, producing five rays instead of one. Generation of ray cells in the anterior body region was further demonstrated by means of a *pkd-2::gfp* reporter gene, which is expressed in one of two ray neurons (Barr and Sternberg, 1999). In *sop-2(bx91)* males, *pkd-2::gfp* is ectopically expressed in the anterior body region (99%, n = 284) (Figure 1F). Thus, *sop-2(bx91)* causes expression of posterior cell fates in the anterior body region.

sop-2 males are also 100% penetrant for a variety of defects in cells derived from the V6 lineage, including ray fusion, missing rays, and ray duplications. In 7% of sides (n = 306), rays 4, 5, and 6 are duplicated at the expense of rays 2 and 3, thus generating the ray pattern 456456, instead of the wild-type V6 ray pattern 23456 (Figure 1G). Other, less frequent, kinds of V6 ray duplications include the ray patterns 2323456, 23456456, and 223456 (4% total, n = 306). These duplications are consistent with anterior to posterior cell fate transformations within the V6 cell lineage. For example, the 456456 pattern suggests that the fate of V6.pap, which normally gives rise to rays 2 and 3, is transformed into that of V6.ppp, which normally gives rise to rays 4–6 (Figure 1H).

Homeotic transformations giving rise to anterior ectopic rays and ray duplications are the result of *mab-5* and *egl-5* gene activity. In a *sop-2; mab-5* double mutant, development of anterior ectopic rays is almost completely eliminated (Table 1, line 3). In a *sop-2; egl-5* double mutant, anterior ectopic rays are formed at a significantly lower frequency than in *sop-2* mutants (Table 1, line 4). The few anterior ectopic rays present in a *sop-2; mab-5; egl-5* triple mutant (Table 1, line 5) might result from the ectopic expression of other Hox genes. It has been shown previously that ectopic expression of *lin-39*, the ortholog of *Drosophila Sex combs reduced*

(C) Wild-type male tail, ventral view.
(D–G) *sop-2(bx91)* male phenotypes.
(D) Central body region bearing an ectopic ray associated with a cuticular fan-like structure.
(E) Male tail with ectopic rays derived from the V5 lineage (arrows). Uppermost side bears ectopic rays (arrows); ray 1 is out of the focal plane. On the bottom side rays 4, 5, and 6 have fused.
(F) Ectopic rays generated from anterior V seam cells. Ectopic rays are visualized with B-type ray neuron marker PKD-2::GFP. Ten ectopic rays are located between the arrows.
(G) Duplication of V6 rays. Lateral view of male tail showing extra rays 4, 5, and 6, duplicated at the expense of rays 2 and 3.
(H) Possible cell fate transformation that could account for the duplication of rays 4, 5, and 6 shown in (G).

(*Scr*), leads to the formation of rays in *mab-5(0)* males (Hunter and Kenyon, 1995; Maloof and Kenyon, 1998). The V6 ray duplications in *sop-2(bx91)* mutant males also require the activities of both *mab-5* and *egl-5*. The V6 ray phenotypes of the *sop-2; mab-5* and *sop-2; egl-5* double mutants are identical to those of the corresponding single mutants *mab-5* and *egl-5* (i.e., absence of V5 and V6 rays in *sop-2; mab-5* mutants and fusion of rays 2–5 in *sop-2; egl-5* mutants) (data not shown). Taken together, these observations indicate that *sop-2(bx91)* causes Hox gene-dependent anterior to posterior cell fate transformation.

Hox Gene Expression Domains Are Expanded by *sop-2(bx91)*

Formation of anterior ectopic rays in *sop-2(bx91)* and their dependence on Hox gene activity suggested that the expression domains of *mab-5* and *egl-5* are expanded into the anterior seam. We therefore examined the expression patterns of these two genes using reporter genes. In wild-type early larvae, expression of *mab-5* is restricted to the cells of posterior body region, including P9–P12, V5, and V6 (Figure 2A) (Kenyon et al., 1997). However, in *sop-2(bx91)* early larvae, *mab-5::gfp* is expressed throughout almost the entire body, including cells in the head, ventral cord, and the syncytial hypodermal cell 7 (hyp7) (Figure 2B). Consistent with its role in generation of ectopic rays from V1–V5, *mab-5::gfp* is ectopically expressed in the descendants of anterior seam cells (data not shown). Similarly, the expression domain of *egl-5* is also expanded in *sop-2(bx91)*. During early wild-type larval development, *egl-5* expression is limited to the tail region (Figure 2C) (Ferreira et al., 1999). Later, during male ray development, *egl-5* is expressed in some descendants of V6 (Figure 1B) (Ferreira et al., 1999). In contrast, in *sop-2(bx91)*, *egl-5::gfp* is ectopically expressed in head and ventral cord neurons (Figure 2D and data not shown). Later, *egl-5::gfp* is also ectopically expressed in the descendants of anterior seam cells, as well as in the additional descendants of V6 generated in *sop-2(bx91)* (Figure 2E and data not shown).

We also examined the expression of a GFP reporter for the *labial* homolog *ceh-13*. Reporters for *ceh-13* are expressed in the lineages of rays 5, 7, and 9, although its function in these cells is not known (Figure 2F) (Stoyanov et al., 2003). In an early *sop-2(bx91)* larva, the expression level of *ceh-13::gfp* in the ventral cord is moderately increased but does not show extensive ectopic expression (data not shown). However, at the late L3 stage, *ceh-13::gfp* is ectopically expressed in *sop-2(bx91)* male rays. In some *sop-2(bx91)* animals, almost all ray cells express *ceh-13::gfp* (3%, n = 109) (Figure 2G). In summary, the Hox genes *mab-5*, *egl-5*, and *ceh-13* are expanded in multiple tissues outside their normal expression domains in *sop-2(bx91)* mutants.

Hox Gene Activity Is Abnormal in the Ventral Cord

In addition to homeotic transformations in the seam, *sop-2(bx91)* animals exhibit defects in the ventral cord. In a wild-type male, *lin-39* specifies the fates of serotonergic motorneurons CP1–CP6, derived from the descendants of P3–8, P(3–8).aapp. In *lin-39* mutant males, no serotonergic CP neurons are generated (Hunter and Kenyon, 1995). Expression of serotonergic fate in CP neurons can be visualized with the serotonin biosynthetic enzyme reporter gene *tph-1::gfp* (Figure 3A) (Sze et al., 2000). In wild-type males, the reporter is expressed in all six CP neurons (n = 23) (Figure 3A). In *sop-2(bx91)* males, it is expressed in an average of 2.3 ventral cord cells (Table 1, line 2), but there is high inter-individual variability, ranging from 0 to 10 (Figure 3B). As in wild-type, development of these neurons requires *lin-39* activity, as no TPH-1::GFP-positive neurons are found in *sop-2; lin-39* males (Table 1, line 6). Thus, it appears that, in *sop-2(bx91)*, the apparent domain of *lin-39* activity has become highly variable and, in some animals, is extended beyond its normal range.

In wild-type, *egl-5* is not expressed in the ventral cord, and *egl-5* loss-of-function has no effect on the number of TPH-1::GFP-positive neurons in the ventral cord (data not shown). However, in *sop-2(bx91); egl-5(n486)* double mutants, the average number of ventral cord neurons expressing *tph-1::gfp* is dramatically increased from 2.3 to 8.2 (range, 7–11) (Table 1, line 4). This indicates that, in *sop-2(bx91)*, *egl-5* has an activity that affects the fates of ventral cord cells. Most likely, ectopic expression of *egl-5* in the ventral cord prevents the expression or function of *lin-39* in specifying the fates of the serotonergic CP neurons.

Ectopic Hox Gene Expression Begins during Late Embryogenesis

We have shown that the expression domains of Hox cluster genes are altered and generally expanded during postembryonic development in *sop-2(bx91)*. This ectopic expression could be due to early initiation of expression in abnormal lineages or to later derepression within a normally silent lineage. To distinguish between these alternatives, we examined the expression patterns of *mab-5::gfp* and *egl-5::gfp* in *sop-2(bx91)* mutant embryos. As in wild-type, these reporters were expressed in a limited number of cell lineages during early embryogenesis (Kenyon et al., 1997). Expression in inappropriate domains first appeared when embryogenesis reached the 3-fold stage, after most cell lineages are completed and cell differentiation is well underway. At this time, *egl-5::gfp* began to be expressed in head neurons, and *mab-5::gfp* was ectopically expressed in hyp7 and some head neurons (data not shown). These observations show that, in *sop-2(bx91)*, Hox gene expression initiates correctly but is later derepressed in differentiating or differentiated cells. We show below, through RNAi experiments, that this is the loss-of-function phenotype of *sop-2*. Hence, *sop-2* gene function is required to maintain the restriction of Hox gene expression to specific cells and tissues.

sop-2(bx91) Affects the Expression of Non-Hox Genes

In addition to homeotic transformations, *sop-2(bx91)* causes some defects that cannot be readily attributed to the ectopic expression of Hox genes, such as long body (Lon), protruding vulva (Pvl) (26%, n = 290), multivulva (Muv) (11%, n = 290), and partial sex transformation (for example, about 1% of *sop-2* hermaphrodites

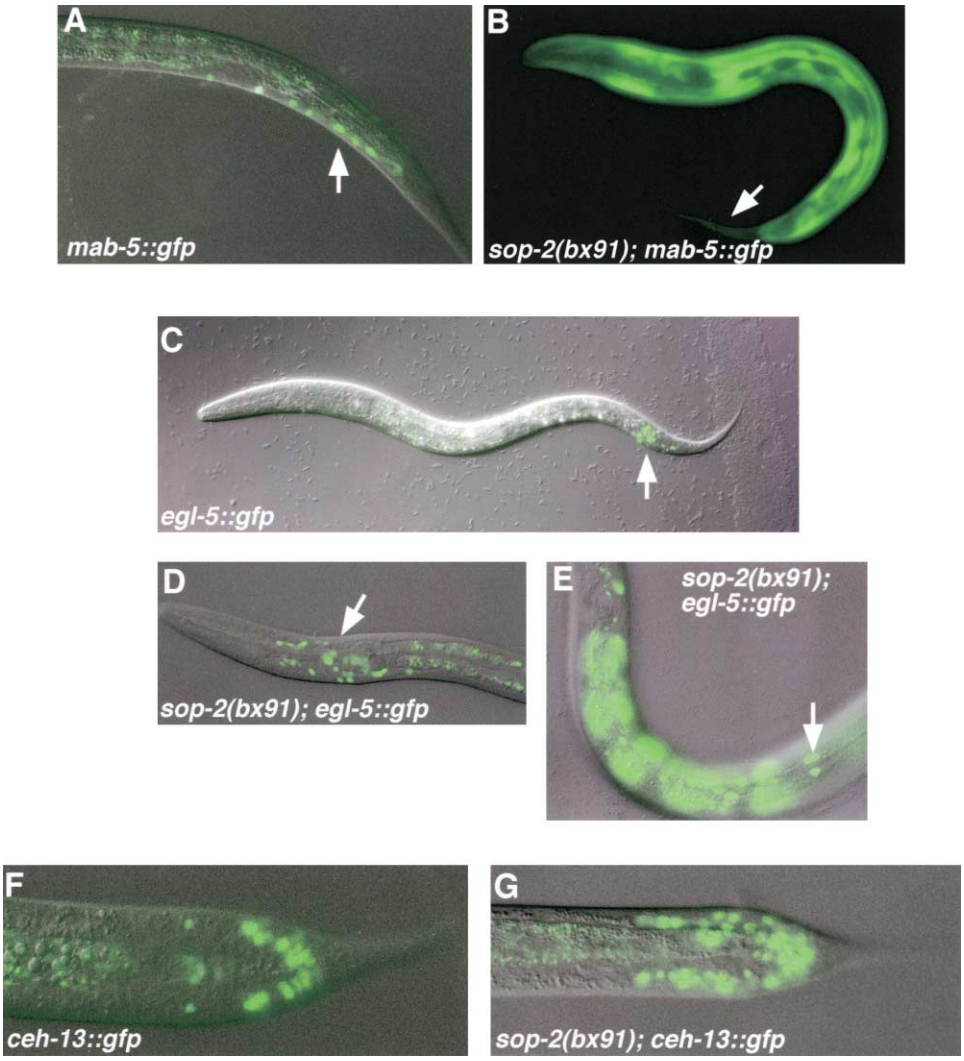


Figure 2. *sop-2* Loss-of-Function Results in Expansion of Hox Gene Expression Domains

(A) Expression of *mab-5* in a wild-type early L2 larva. MAB-5::GFP is confined to a few cells of the posterior (arrow).

(B) Extensive ectopic expression of *mab-5* in a *sop-2* L1 larva. MAB-5::GFP is expressed in the head neurons, hyp7, and ventral cord. Note that MAB-5::GFP is absent from the tail region (arrow).

(C) Expression of *egl-5* in a wild-type early L1 larva. EGL-5::GFP is restricted to the tail region.

(D and E) Ectopic expression of EGL-5::GFP in (D) head neurons and (E) descendants of anterior seam cells in *sop-2(bx91)* mutants.

(F) Expression of *ceh-13* in a wild-type L4 stage male. CEH-13::GFP is in ray cells 5, 7, and 9.

(G) Ectopic expression of *ceh-13* in a *sop-2(bx91)* male of the same stage.

Irregular fluorescence particles in (A)–(G) are gut autofluorescence.

have a blunt, male-like tail, instead of a tail spike). These defects suggest that *sop-2* regulates the expression of some genes in addition to the Hox genes. However, the pleiotropic effects of the *sop-2* mutation are not caused by a nonselective expansion of the expression domains of all genes, since the expression patterns of the following reporter genes at 25°C are not obviously altered in *sop-2(bx91)* mutants: *pal-1::gfp, elt-2::gfp*, and *mec-4::gfp* (data not shown).

sop-2 Regulates Hox Gene Expression Independently of the Wnt Signaling Pathway

Mutations in the negative Wnt regulator *pry-1* (Axin homolog) cause ectopic expression of Hox genes in many cells, including head neurons, anterior seam cells, and ventral cord neurons (Maloof et al., 1999; Gleason et al., 2002; Korswagen et al., 2002). However, several observations suggest that ectopic expression of Hox genes in *sop-2(bx91)* does not require Wnt signaling activity. First, mutations in an essential Wnt signaling component, the β-catenin homolog *bar-1*, have no effect on Hox gene expression in a *sop-2(bx91)* background; in *sop-2; bar-1* double mutants, *mab-5::gfp* and *egl-5::gfp* are still ectopically expressed. Second, *sop-2* and *pry-1* mutants differ in the temporal onset of ectopic Hox gene expression. In *pry-1* mutants, ectopic Hox gene expression is not observed until the late L1 larval stage (Maloof et al., 1999), whereas, in *sop-2(bx91)* mutants, ectopic

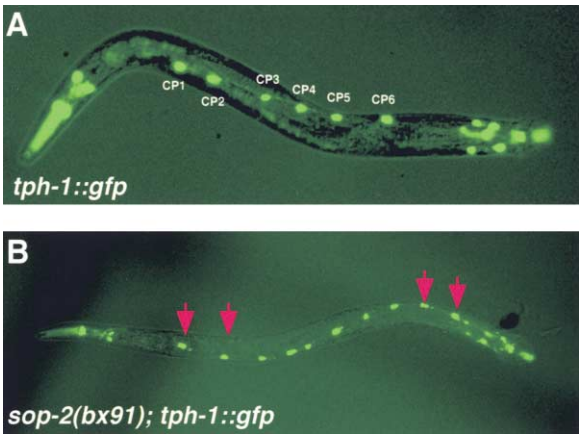


Figure 3. *sop-2* Loss-of-Function Alters Male Ventral Nerve Cord Neuron Fates

(A) Expression of serotonergic fate marker TPH-1::GFP in a wild-type male ventral cord. Six CP motor neurons (CP1–CP6) are serotonergic and express this reporter.

(B) Ectopic expression of TPH-1::GFP in the *sop-2(bx91)* male ventral cord. In this animal, four extra neurons of the ventral cord express TPH-1::GFP (arrows).

Hox gene expression is apparent during late embryogenesis. Finally, the tissue-specific pattern of ectopic Hox gene expression in *sop-2(bx91)* does not completely overlap with that seen in *pry-1* mutants. For instance, in contrast to *sop-2* mutants, *pry-1* males produce a wild-type set of V6 rays (n = 12), suggesting that Hox gene expression in this lineage is wild-type. Conversely, in *pry-1* mutants, *mab-5* is ectopically expressed in the neuroblast QR and its descendants (Maloof et al., 1999), whereas this is not the case in *sop-2* mutants (n = 16). Taken together, these results indicate that *sop-2* regulates Hox gene expression independently of Wnt signaling.

sop-2 Encodes a SAM Domain Protein

The *sop-2* locus was genetically mapped with a combination of genetic and molecular markers (Figure 4A). Cosmids within the defined genetic interval were tested for complementation rescue of *sop-2(bx91)* larval lethality (see Experimental Procedures). The minimal *sop-2* rescuing region, C50E10.4 (accession AAK73863), encodes a protein of 735 amino acids (Figure 4B). The predicted structure of C50E10.4 was confirmed by sequencing cDNAs from the region. *sop-2(bx91)* contains a missense mutation that converts amino acid position 633 from proline to serine (P633S). Treatment of wild-type animals with C50E10.4 dsRNA induced phenotypes similar to those caused by *sop-2(bx91)* (see Experimental Procedures), confirming that this predicted gene is *sop-2* and indicating that *bx91* is likely to be a loss-of-function mutation. However, unlike *sop-2(bx91)*, the *sop-2(RNAi)*-induced phenotypes are not temperature sensitive in the 15°C–25°C range, suggesting that the *bx91* missense mutation results in a temperature-sensitive protein.

To gain insight into the functional properties of SOP-2, we searched for homologous domains using available databases. A hidden Markov model (HMM) search of the Pfam databases (Bateman et al., 2002) revealed significant similarity ($E < 10^{-9}$) between the C terminus of SOP-2 and the C termini of four predicted *C. elegans* proteins: K04C1.2, ZC376.4, T10D4.6, and F28H6.6 (defining Pfam-B 7.6 domain PB016780) (Figure 5A). The domains of the four proteins are 38%, 32%, 21%, and 26% identical to the SOP-2 domain (621–712), respectively. However, RNAi experiments showed that loss-of-function of these four genes individually did not cause any obvious phenotypic defects or ectopic expression of Hox gene reporters (data not shown).

Outside *C. elegans*, the protein domain in the Pfam-A database most similar to the SOP-2 domain is the sterile α motif/pointed (SAM/PNT) domain. Although the expectation value in this comparison is not statistically significant (HMM profile search; $E = 9$), computational predictions of the secondary and tertiary structure of the SOP-2 domain also suggest that it is a SAM domain. Most fold recognition algorithms (94%) identify a SAM domain as the closest structural match to the SOP-2 domain: the domains of the *Drosophila* PcG protein PH (SPM subfamily; Bornemann et al., 1996) and human ETS transcription factor TEL (SAM/PNT subfamily) were identified as the top matches by 10 and 7 servers, respectively (Figures 5A, 5D, and 5E) (similar results were obtained for the other SOP-2 domain-containing proteins; data not shown). From these results and the phylogenetic analysis described below, we conclude that the nematode-specific SOP-2 domain family constitutes a subfamily within the SAM domain family (Schultz et al., 1997). SAM domains consist of a bundle of four to five α helices and are involved in interactions with proteins (Kim et al., 2001, 2002). A homology model of the SOP-2 SAM domain based on the SAM/PNT domain of TEL is shown in Figure 5C (Kim et al., 2002).

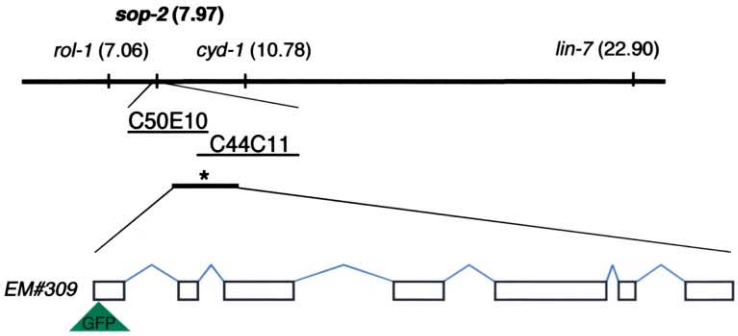
The SAM Domain Mediates the Self-Association of SOP-2

The SAM domain is thought to mediate protein-protein interactions, including homotypic and heterotypic binding between the PcG proteins PH and SCM (Kyba and Brock, 1998). We found that SOP-2 also interacts with itself in the yeast two-hybrid assay (Figure 4C). A deletion construct containing the N-terminal region of SOP-2, but lacking the SAM domain, fails to interact with SOP-2. In contrast, the SAM domain alone can interact with both full-length SOP-2 and with the SAM domain itself (Figure 4C). Self-association of SOP-2, therefore, maps to the SAM domain. Of note, the *sop-2(bx91)* mutation, which is adjacent to the SAM domain, does not affect the ability of the domain to associate with the wild-type SOP-2 SAM domain, but it severely impairs its ability to bind SOP-2(bx91) itself (Figure 4C). This recessive mutation is therefore likely to disrupt the self-association of SOP-2.

The SAM Domain Is Required for Formation of SOP-2 Nuclear Bodies

To define the expression pattern of *sop-2*, we generated a functional *sop-2* reporter gene (*EM#309*) and examined its expression pattern in transgenic animals. This reporter gene contains the entire open reading frame of *sop-2*, with *gfp* inserted between amino acids two and

A



B

```
MSSNLTSNEM SSTSAIVEPP EAVKGNERDK STRRSTSQVV RPEKHEVENA LEDQCSSSSL   60
PKEAQYYAKL DKKLEGKDPR SQFYEAVRLS ADIFAHKFEK AVCSRQTFEP TNSIIKVLNT  120
AEEEMLHEKV VPLPVSSKLQ YYLNRGRYDT IFDRDEQLQR TADPMADEPD HVEKRVTSIL  180
EQASREMEEG EVEPVFYDGS DEDQELPIDL GAMRNLQRTN KFAARSSRMA ARRGGRPGYR  240
GAFRGAARGA PSRRPAPAAE VAPETPVAAP MAPAAPAAPA TPEAAPAAEV MDTSIATEMP  300
QESAVDLSNV SAATEMDTSK EGEASRPTSE KKKKIRTTEM DRLMSMDLGP KDGGRVGELG  360
HMWPESRRRP AAPLPETPAA QPRKSLPRRA AEKKKPEDSD AAEEQEVEME VDNDASTSTP  420
RNARGGRGGG NRRGSRRGQK RTSGGSGKLV EPKKEPVDEP AEKIPKRSEA APEVPATATT  480
KEAPPSTSSS PPDAPATPAT PASSDSRDSP RKIRAMIFSL TGSPPESETP PVLQQEQVIS  540
TAAPTAGRHP NIIQQVPHIN RIPPQPLRRL TAPQAPPASQ PEEPPVQQTV PVVKVELASA  600
PAPIVRDPQS TEPVPPAMPT LVENNHEATL ILRPNKTSDY TRWNAQDLIN WVRLLITNNV  660
DSTIAIMVRE EFDGETLACL VLDDDFRKEV PIPYGHYKKM KIYGTEVLNH YRTEKYQADL  720
RKFHEELAAW KAQQR
```

C

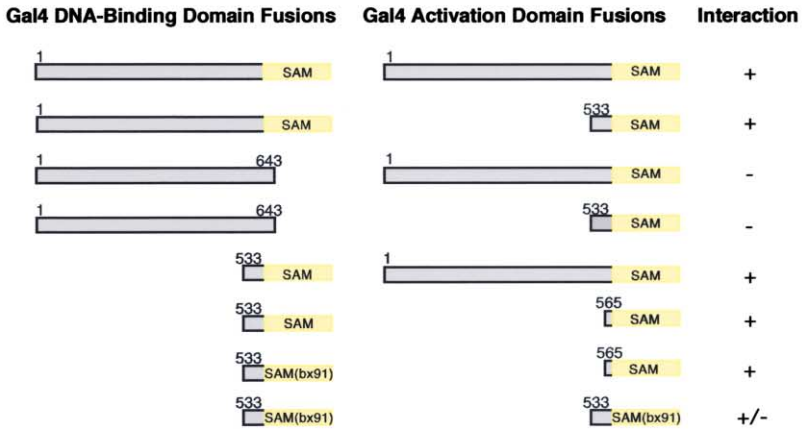


Figure 4. Molecular Structure of *sop-2*

(A) Genetic map of a portion of LGII (top) and the corresponding physical map (middle) showing the position of overlapping cosmids C50E10 and C44C11 that rescue *sop-2(bx91)* mutant phenotypes, as well as the minimal 12 kb rescuing region (*) within these cosmids. *sop-2* gene structure (bottom). The lengths of the introns and cosmids are not to scale. The insertion site of the *GFP* ORF in *sop-2* reporter gene *EM#309* is shown in green.

(B) Predicted protein sequence of SOP-2. The region in red (621–712) is similar to protein domain family PB016780 (see text for details). The residue altered by the *bx91* mutation is boxed.

(C) Summary of SOP-2 two-hybrid constructs and results. The structures of the SOP-2 and SOP-2(bx91) fusion proteins used are shown. Each line shows the pair of GAL-4 DNA binding fusion protein and GAL-4 activation domain fusion protein tested and the outcome of this interaction (+, −, or +/−; see Experimental Procedures). Orange boxes represent the SAM domain.

three (Figure 4A). This reporter rescues the *sop-2(bx91)* defects (see Experimental Procedures). SOP-2::GFP localizes to the cell nuclei of essentially all somatic cells and is apparent from the 50-cell stage embryo onward. At first, expression is weak and diffuse within nuclei (Figure 6A), but, by the 200-cell stage, it becomes stronger, and distinct nuclear speckles, which we call "SOP-2 bodies," appear. By the comma stage (~400 cells), the nuclei of most somatic cells contain SOP-2 bodies, but their number and size varies among different cell types (Figures 6B and 6C). Hypodermal and gut nuclei, which undergo endoreduplication, contain large

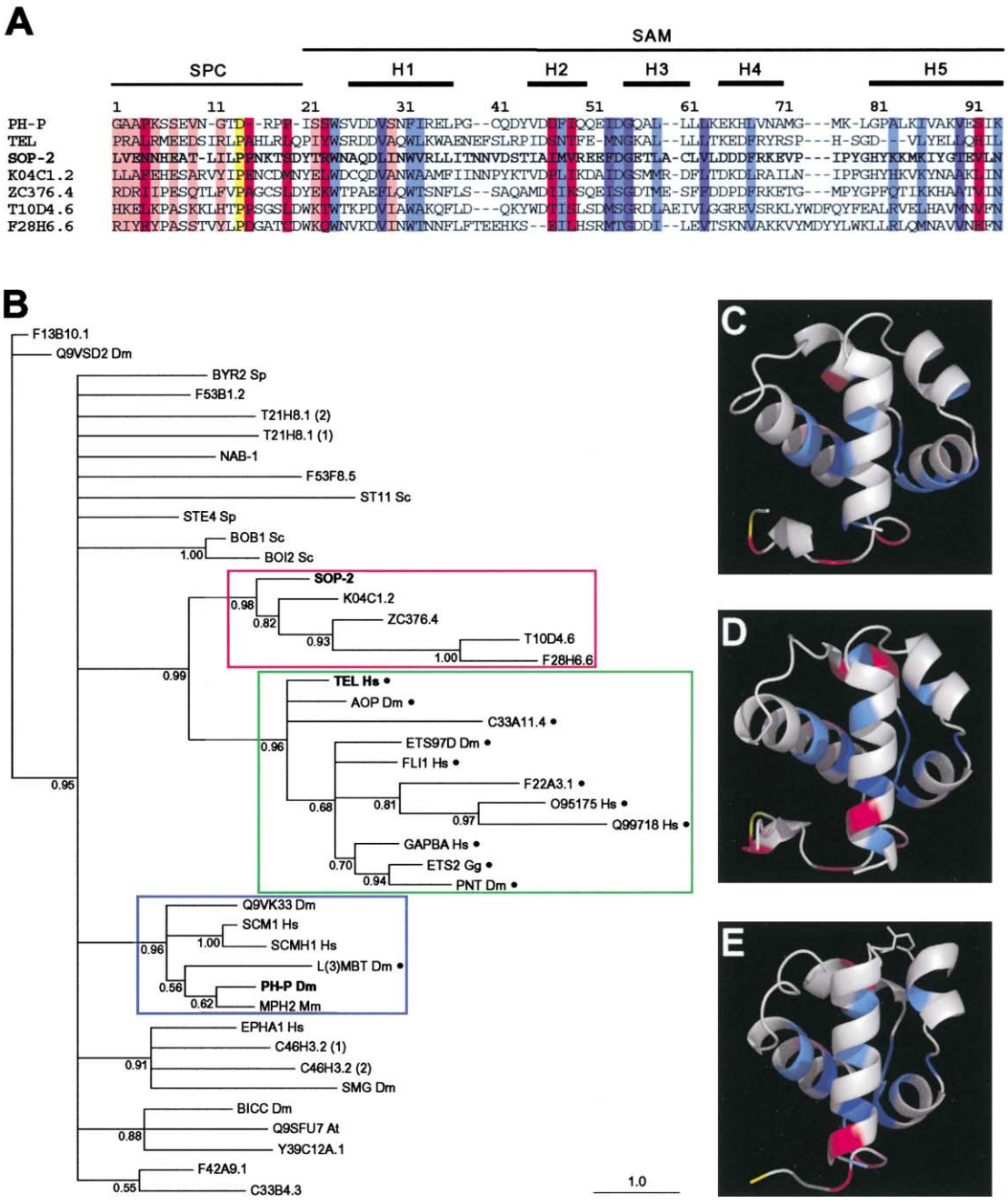


Figure 5. SOP-2 Contains a SAM Domain

(A) Alignment of the SAM domain and preceding SAM/pointed common region (SPC) of SOP-2 (residues 621–712), K04C1.2A (411–503), ZC376.4 (159–249), T10D4.6 (678–768), F28H6.6 (628–716), *Drosophila* PH proximal (1494–1575), and human TEL (38–124). The five α helices are marked H1–H5. Residues marked in red and blue are fast and slow evolving, respectively (relative rates of substitution estimated by ML with the tree in [B]; dark blue, rate less than 0.4; light blue, rate less than 0.5; pink, rate greater than 1.5; red, rate greater than 1.75; overall average rate, 1). The residue marked in yellow is mutated to serine in *sop-2(bx91)*.

(B) Evolutionary relationships of SOP-2, SAM, and SAM/PNT domains. Consensus unrooted tree (50% majority rule) of trees sampled by MCMC. Numbers indicate the posterior probability that the clade is correct under the JTT-F model of amino acid substitution with γ-distributed substitution rates across sites. Polytomous nodes indicate that no resolution of the polytomy has posterior probability greater than 0.5. Branch lengths are ML estimates under the same model and reflect the mean number of substitutions per site (shape parameter, $\alpha = 3.39$, SE = 0.54; total tree length, 57.3). Three monophyletic domain subfamilies of SAM domains with high posterior probabilities are highlighted: SOP-2 domains, red box; SAM/PNT domains, green box; SPM domains, blue box. The SAM domain of SOP-2 is more closely related to the SAM/PNT domains than to the SPM domains of Ph and Scm. The root must be located outside the clade containing both the SOP-2 and SAM/PNT domains because these domains are exclusively present in animals whereas other SAM domains are present in plants and fungi. Domains with a dot after the name are classified as SAM/PNT in Pfam. *At, Arabidopsis thaliana; Dm, Drosophila melanogaster; Gg, Gallus gallus; Hs, Homo sapiens; Mm, Mus musculus; Sc, Saccharomyces cerevisiae; Sp, Schizosaccharomyces pombe*; no code, *C. elegans*. Numbers in parenthesis indicate the position of the domain in a protein.

numbers of SOP-2 nuclear bodies (hypodermis, 39.8 ± 8.8, n = 14; gut, >500, n = 2) compared with other cell types (seam cells, 8.8 ± 3.7, n = 8; neurons, 8.4 ± 2.7, n = 15) (Figures 6D and 6E), raising the possibility that the number of SOP-2 bodies may correlate with DNA content. No expression is detected in the nucleolus.

To determine whether the SAM domain is required for formation of SOP-2 bodies, we introduced a premature stop codon into *EM#309* (Trp643 to stop), deleting the SAM domain. In animals carrying this *sop-2(ΔSAM)::gfp* reporter, SOP-2::GFP is distributed uniformly throughout the nucleus and does not form nuclear bodies (Figure 6F). To test whether formation of SOP-2 bodies is correlated with its function, we tested the effect of the temperature-sensitive *bx91* mutation. We introduced the Pro633 to Ser mutation into *EM#309* and examined its expression pattern at 15°C (wild-type function) and at 25°C (loss-of-function). SOP-2(bx91)::GFP localizes to nuclear bodies at 15°C (Figure 6G), but not at 25°C (Figure 6H). Expression levels of SOP-2(bx91)::GFP were comparable to those of wild-type SOP-2::GFP at all temperatures (data not shown). Of note, *sop-2(bx91)::gfp* transgenic animals weakly phenocopy the *sop-2* mutant at 25°C, despite presence of the endogenous wild-type allele. This might result from a sequestration mechanism, whereby the mutant transgenic SAM domain dimerizes with the endogenous wild-type protein, leading to its impaired nuclear localization. Taken together, our observations suggest that the SAM domain is directly or indirectly responsible for the formation of SOP-2 nuclear bodies. The disruption of self-association, as well as nuclear localization, by the *bx91* mutation suggests that these properties are important to SOP-2 function.

Evolutionary Origin of the SOP-2 SAM Domain

To explore the evolutionary relationship between *sop-2* and PcG genes, we undertook a phylogenetic analysis to trace the origin of their respective SAM domains. A Bayesian phylogenetic analysis of the SAM domain and the SAM/PNT common (SPC) region preceding it (Figure 5A) establishes the SOP-2 domain-containing proteins as a nematode-specific, SAM domain subfamily (posterior probability, p = 0.98; Figure 5B, red box). Furthermore, the SOP-2 domain subfamily is closely related to the SAM/PNT domains from proteins containing ETS domains, such as TEL (posterior probability, p = 0.99; Figure 5B, green box). Our results imply that SOP-2 is not an ortholog of one of the SPM domain-containing PcG proteins because its SAM domain does not group with the SPM subfamily constituted by SCM, PH, and L(3) MBT (posterior probability that SOP-2 domains are more closely related to SPM domains than to SAM/PNT domains, $p < 0.001$; Figure 5B, blue box; Bornemann et al., 1996). Since SOP-2 does not share any motifs or domains with other PcG proteins apart from the SAM domain (e.g., zinc fingers of PH and MBT repeats of SCM), it is unlikely that SOP-2 evolved from an ancestral PcG protein that independently acquired a SAM/PNT-like domain.

A maximum-likelihood (ML) analysis of the evolutionary history of the SAM domain sequences shown in Figure 5B supports the conclusion that there is significant variation in the rates of amino acid substitution among sites ($p < 10^{-44}$; Figure 5A), with most of the slowly evolving residues occurring within the five α helices of the SAM domain and in the loops connecting helices 2–4 (Figures 5A and 5C–5E). Among the slowly evolving sites, three are known to be essential for the structure (Figure 5A, position 24) or the binding interface (positions 58 and 68) of the SAM domain (Kyba and Brock, 1998). Although the SPC region is evolving 63% faster than the SAM domain (Figure 5A), position 14 (Figure 5A, yellow) is evolving 57% slower than the SAM domain. This position, a Pro residue in all the SOP-2 domain-containing proteins and most SAM/PNT domains, is mutated to Ser in *sop-2(bx91)*, confirming its functional significance.

Discussion

The Role of *sop-2* in Gene Regulation

We have shown that a repression mechanism mediated by *sop-2* maintains Hox gene expression patterns in several diverse *C. elegans* tissues and body regions. Thus, in spite of the reproducible cell lineages and lineage-specific mechanisms controlling the expression of Hox genes in *C. elegans*, Hox gene expression domains are also maintained, in part, by a combination of positively acting (e.g., trxG proteins LIN-49 and LIN-59 [Chamberlin and Thomas, 2000]) and negatively acting regulatory pathways, just as they are in other animals. SOP-2 contains a SAM domain, a protein domain associated with the formation of repressive chromatin complexes by other repressors, including the two PcG proteins PH and SCM. Therefore, the mechanism of repression in *C. elegans* may be similar to that in other animals.

As seen in fly or mouse PcG mutants, in *sop-2* mutants, Hox genes are not expressed in every cell, and each Hox gene is ectopically expressed at a different level. For instance, *mab-5* is not expressed in the tail region (Figure 2B), and the ectopic expression domains of *mab-5* and *egl-5* appear to be much broader than those of *ceh-13* and *lin-39*. The simplest explanation for these gene-specific patterns is that, for each gene, transcription factors that will activate expression in the absence of *sop-2* gene function are present or active in only a subset of cells. Alternatively, *sop-2* repression may be redundant with other repressive mechanisms in some tissues. Crossregulation between Hox genes may also contribute to the Hox gene expression patterns observed in *sop-2* mutants, as our data suggest for *lin-39* and *egl-5*.

sop-2 also appears to play a role in regulating the

(C–E) EH surfaces of the SAM domains.
(C) Homology model of the SOP-2 domain (632–706) based on the (D) SAM/PNT domain of TEL (Kim et al., 2001).
(E) SAM domain of PH (Kim et al., 2002). Residues are colored as in the alignment in (A).

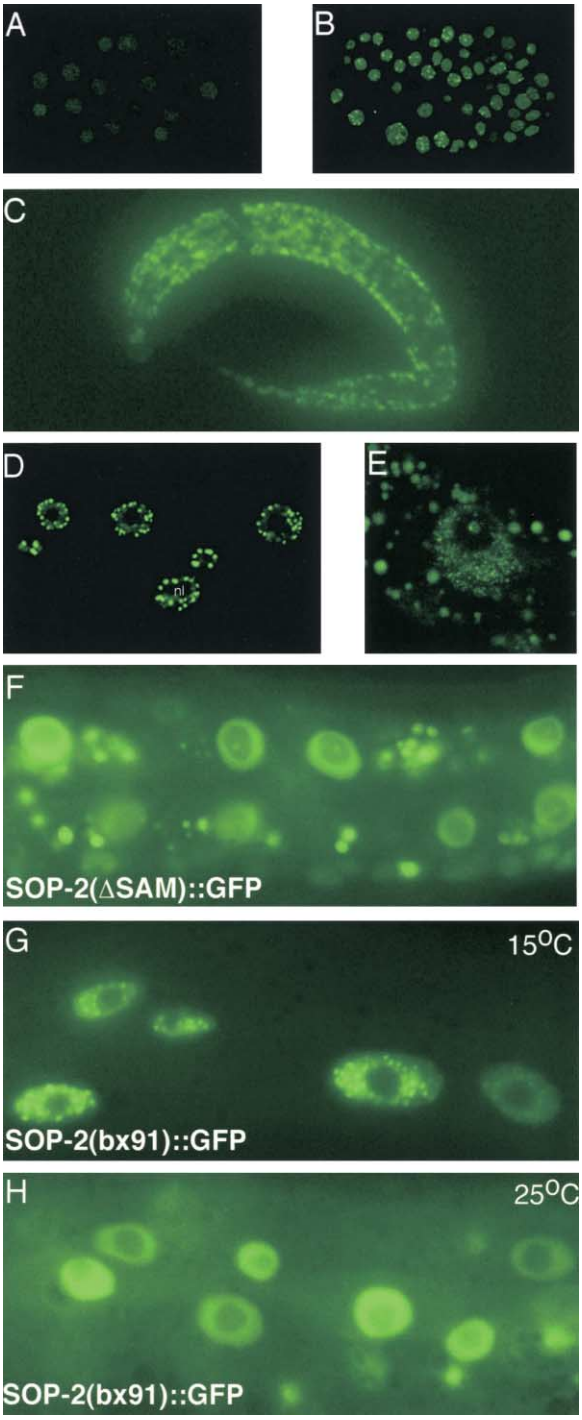


Figure 6. SOP-2 Forms Nuclear Bodies

(A–E) Expression of SOP-2::GFP at different developmental stages and in different cell types.

(A) Fifty-cell stage embryo showing weak nuclear expression.

(B) Three hundred-cell stage embryo showing strong nuclear expression and nuclear bodies.

(C) Early adult hermaphrodite.

(D) Hypodermal (large) and seam (small) nuclei and (E) gut cell nucleus of adult hermaphrodite. Note that there is no expression in the nucleolus (nl).

(F) Expression of SOP-2(ΔSAM)::GFP in early adult reared at 20°C, showing strong, homogenous nuclear expression.

(G) Expression of SOP-2(bx91)::GFP in early adult reared at 15°C, showing nuclear bodies.

(H) Expression of SOP-2(bx91)::GFP in early adult reared at 25°C, showing strong, homogeneous nuclear expression.

(A, B, and D) Single confocal sections.

(E) Three confocal sections superimposed (0.5 μm apart).

expression of nonhomeotic genes, since *sop-2* mutants have pleiotropic effects not known to be associated with Hox gene misexpression, including abnormalities in body size, sex determination, and vulva development. PcG mutants in other organisms also cause defects that may be caused by inappropriate expression of nonhomeotic genes (Jacobs and van Lohuizen, 2002). For instance, *M33* (*Pc* homolog) mutant mice have slow gonad growth that leads to male to female sex reversal (Katoh-Fukui et al., 1998), and loss of function of *mel-18, mph-1/rae28, bmi-1* (*Pc*, *ph*, and *Psc* homologs, respectively), and *M33* result in cell proliferation defects (Raaphorst et al., 2001). Thus, regulation of these distinct pathways may be inherent properties of some PcG genes.

The Mechanism of SOP-2 Repression

We do not know the direct gene targets of *sop-2* repressive activity; *sop-2* could repress Hox genes directly or indirectly by a variety of mechanisms. An attractive hypothesis we favor is that SOP-2 is directly targeted to Hox gene promoters and functions in the same way that the PRC1 PcG complex functions in *Drosophila* and mammals, namely, by the formation of a stable and spreading repressive protein complex on DNA. The SAM domain of SOP-2 is likely to be structurally similar to that of PH and TEL, both of which form head to tail, left-handed helical polymers (Figure 5A) (Kim et al., 2001, 2002). The SAM domain of PH plays an important role in formation of the PRC1 complex, mediating the interaction of PH with itself and with the SAM domain of SCM (Kyba and Brock, 1998). While the precise interactions that lead to the spread of transcriptional repressor complexes along the chromatin remain to be defined for PcG proteins, the conserved structure of the SAM domains of PH, TEL, and SOP-2 points to similar mechanisms.

This hypothesis for SOP-2 function raises several questions, including the nature of additional proteins that may interact in a complex with SOP-2 and how a SOP-2 repressive complex is recruited to DNA. If SOP-2 has partners in a repressive complex, these remain unknown at present. We could not gain evidence by RNAi experiments that the other proteins in the *C. elegans* genome with SOP-2 domains have this role, and homologs of the other components of PRC1 in flies and mammals cannot be identified in the *C. elegans* genome. In *Drosophila* and mammals, PRC1 is recruited to DNA through the activities of PcG proteins of the ESC-E(Z) complex. Recently, *C. elegans* genes of this group (*mes* genes) have been shown to play a role in Hox gene repression (Ross and Zarkower, 2003). Hence, one possibility is that SOP-2 activity is potentiated by *mes* genes. However, whereas, in *Drosophila*, PcG mutations of the ESC-E(Z) and of PRC1 groups have similar phenotypes and the activities of these complexes are tightly coupled, in *C. elegans*, the *mes* mutant and *sop-2* mutant phenotypes are dissimilar. *mes* mutants have weakly penetrant effects on male ray development,

whereas, in *sop-2(bx91)*, these effects are fully penetrant. Moreover, the additional phenotypes seen in a *sop-2* background, including the highly penetrant larval lethality, are not seen in *mes* mutants. Therefore, *sop-2* has functions not dependent on *mes* gene activity. If SOP-2 acts as we hypothesize, by forming a repressive complex on DNA, this complex must be recruited in part by novel mechanisms.

A further question of interest is how SOP-2-mediated repression is relieved in cells that normally express Hox genes, since *sop-2* appears to be widely expressed in all cells. In some cells, it appears that the Wnt pathway can overcome *sop-2* repression. Both *sop-2* and a Wnt pathway repressor, *pry-1*, repress Hox gene activity in the anterior seam. Hox gene expression in a *pry-1* mutant, but not in a *sop-2* mutant, requires the Wnt pathway cofactor β-catenin. This epistasis relationship indicates that *sop-2* acts downstream of, or in parallel to, the Wnt pathway. Wnt pathway activity can activate Hox genes in cells containing a wild-type *sop-2* gene, whereas loss of *sop-2* function alone is sufficient to activate Hox genes in the absence of Wnt pathway activity. This shows that, in a *pry-1* mutant, the Wnt pathway activates Hox genes either by blocking *sop-2* function or by activating a parallel pathway insensitive to *sop-2* activity.

The apparent localization of SOP-2 into distinct nuclear bodies is a striking property of this protein and is consistent with our hypothesis that SOP-2 acts by forming a repressive complex on DNA. Two lines of evidence, the temperature sensitivity of the missense mutation *sop-2(bx91)* for both *sop-2* activity and localization to nuclear bodies and the correlation between the number of nuclear bodies and tissue ploidy, argue for a functional significance of these bodies. Other SAM domain-containing proteins also form nuclear bodies. A Ph-EGFP reporter protein in *Drosophila* forms nuclear bodies from the early gastrula stage onward, which increase in number in the nuclei of epidermal cells as development proceeds (Netter et al., 2001). The transcriptional regulator TEL also forms nuclear bodies, which are dependent on its SAM domain (Chakrabarti et al., 2000). The nature and function of these nuclear structures are not yet understood.

Evolution of Global Repression

Despite its C-terminal SAM domain, *sop-2* does not appear to be a *C. elegans* homolog of the PcG genes *ph* or *Scm*. Rather, the SAM domain of *sop-2* is more closely related to the SAM/PNT domains of ETS transcription factors such as TEL. This pattern could be explained by two evolutionary scenarios: either a functional PRC1 complex was absent from the lineage leading to *C. elegans* or it was initially present, but its members subsequently degenerated or disappeared. We favor the second scenario, since conservation of the core components of the PRC1 complex (PC, PSC, PH, and dRING1) between *Drosophila* and mammals indicates that this complex must have been present in groups ancestral to nematodes, as well (Aguinaldo et al., 1997; Peterson and Eernisse 2001). This inference is further supported by two observations on the evolution of the *Pc* and *ph* genes. First, although *C. elegans* lacks any proteins containing an SPM domain (data not shown), the parasitic nematode *Trichinella spiralis* contains an EST (accession number BG521169), encoding a putative SPM domain similar to that of the mouse PH homolog Edr1 (BLAST; $E < 0.0001$). Second, a homolog of PC has recently been discovered in the cnidarian *Podocoryne carnea*, implying that PC must have been present in the ancestor of all bilaterian animals (Lichtneckert et al., 2002). However, none of the *C. elegans* chromodomain-containing proteins contain the C-terminal C box required to recruit the other members of the PRC1 complex, which is present in the PC proteins of *P. carnea*, *Drosophila*, and vertebrates. Both observations are consistent with the loss or degeneration of a preexisting PRC1 complex in the nematode lineage leading to *C. elegans*.

Not only has there been a loss of the PRC1 complex, but there also appears to have been a degeneration of the Hox cluster itself: several genes (*Hox2-4*, *Antp, Ubx*, and *abd-A*) have been lost, and the physical cluster has disintegrated, such that it now contains thousands of intervening genes (Ruvkun and Hobert, 1998; Aboobaker and Blaxter, 2003). This correlation led Duboule (1992) to propose a causal link between the degeneration of the Hox cluster and the substitution of global mechanisms of Hox gene regulation by lineage-specific ones. Our observations and those of the accompanying paper by Ross and Zarkower (2003) suggest an alternative, namely, the preservation of a globally acting repressive mechanism, but one in which the loss of Hox gene organization has been accompanied by the evolution of novel protein(s) assuming the role of the PRC1 protein complex.

Experimental Procedures

Strains

Most strains carry the *him-5(e1490)* mutation, which produces a high frequency of male self-progeny. The following strains were used in this work: LGII: *sop-2(bx91), swIs1(ceh-13::gfp, rol-6(su1006)), muIs16(mab-5::gfp, dpy-20), rol-1(e91), lin-7(e1413)*, and *cyd-1(he112)*; LGIII, *mab-5(e1239), lin-39(n1760), egl-5(n945), and pha-1(e2123ts)*; LGIV, *muIs6(lin-39::lacZ, rol-6(su1006))*; LGV, *Is(tph-1::gfp, rol-6(su1006))* and *bxIs14(pkd-2::gfp, pha-1(+))* (L. Jia and S.W.E., unpublished data); LGX: *bar-1(ga80)* and *bxIs13(egl-5::gfp, lin-15(+))*.

Isolation, Mapping, and Cloning of *bx91*

sop-2(bx91) was isolated in a screen for suppressors of *pal-1(e2091)* (*sop*) V6 ray loss as described in Zhang and Emmons (2000). Eighty-one percent of V6 lineages produce rays in *sop-2(bx91); pal-1(e2091)* ($n = 384$), compared with 5% in *pal-1(e2091)* ($n = 232$). Presence of the *pal-1(e2091)* mutation in the background did not cause any difference in the phenotype from that of the *sop-2(bx91)* single mutant.

Mapping with polymorphic markers *jsp301*, *jsp303*, and *jsp304* (identified by Dr. Nonet and located in F35C5, W09H1, and F54F11, respectively) placed *sop-2* at approximately 270 kb to the left of cosmid W05H5. Cosmids from this region and PCR products spanning the gap were injected into *sop-2(bx91)* together with transformation marker pRF4 [*rol-6(su1006)*], and rescue of larval lethality at 25°C was assessed. Coinjection of C50E10 and C44C11 fully rescued the *sop-2(bx91)* larval arrest phenotype. Rescuing activity was further delimited to a 12 kb region (C50E10 nt position 36901 to C44C11 nt position 5941) that contained a single ORF, C50E10.4.

RNAi Experiments

T7- and T3-flanked PCR templates (C50E10 nt 40846–41962) were used for RNA synthesis. Single-stranded RNA was transcribed with MEGAscript T3 and T7 kits (Ambion), annealed, and then injected into wild-type, *muIs16*, and *bxIs13* animals. F1 progeny generated

in the first 4–24 hr were scored for larval lethality and ectopic expression of Hox gene reporters at 20°C; larval arrest (12%, n = 312 animals), ectopic expression of *egl-5::gfp* (28%, n = 75) and *mab-5::gfp* (49%), Pvl (13%, n = 102).

Bioinformatic Analyses

Profile hidden Markov model (HMM) searches were done with HMMER 2.2g (http://hmmer.wustl.edu/). We used 18 fold-recognition servers (3DPS, 3DS3, BasB, BasC, FFA3, FUG2, GETH, INBG, MGTH, ORFb, ORFs, PCO3, PCO4, PMO3, PMO4, SFAM, SHGU, and ST99) (Bujnicki et al., 2001). The structure of the SOP-2 SAM domain was modeled on the basis of the SAM/PNT domain of TEL (Protein Data Bank code 1LKYA) with the PMO4 server. Less than 5% of the model shows an error score >95% in the structure verification algorithm ERRAT, a level considered satisfactory for good crystallographic models (Colovos and Yeates, 1993). Structures were drawn with PyMOL 0.84 (http://pymol.sourceforge.net/).

The SOP-2 domain-containing sequences (5) and representative SAM and SAM/PNT sequences (26 and 12 sequences, respectively, contained in the seed alignments used to define each Pfam-A family) were chosen for phylogenetic analysis. The amino acid sequences containing the SAM domain plus the SAM/PNT common region (SPC) were aligned with ClustalW. Phylogenetic inference was done by a Bayesian approach under the JTT empirical model of amino acid substitution (Yang et al., 1998), with γ-distributed substitution rates among amino acid sites (four rate categories). The posterior probabilities of trees were approximated with the Markov chain Monte Carlo (MCMC) algorithm implemented in the MrBayes 3.0B program (Huelsenbeck and Ronquist, 2001). The states of the Markov chain were saved every 100 generations. The consensus tree and posterior probabilities of branches were calculated on the basis of two independent sets of 10,000 trees sampled by the MCMC procedure after the chains had reached stationarity.

A consensus phylogeny was used in maximum-likelihood (ML) analyses with PAML 3.13 (Yang, 1997). We used the JTT-F amino acid substitution model, assuming either a single rate for all sites or γ-distributed rates among sites (ten rate categories) (Yang et al., 1998). Branch lengths and rates of amino acid substitution at each residue were estimated under the γ rates model, and the single-rates and γ rates models were compared by a likelihood ratio test.

Yeast Two Hybrid

Yeast two-hybrid experiments were performed with the ProQuest Two-Hybrid System (Gibco) as recommended by the manufacturer. Full-length *sop-2* cDNA or cDNA corresponding to the C-terminal SAM domain were subcloned in-frame into pPC86 and pPC97. pPC97-SAM (P633S), which contains the SOP-2(bx91) mutation, was generated by PCR-based mutagenesis. Interactions were determined by scoring blue color on X-gal assay and confirmed by growing on 3AT, Ura$^-$, and 5FOA media. The control constructs A–E were used to monitor the interaction signal. The interaction between SOP-2 is as strong as the interaction between rat c-Fos and mouse c-Jun control (control D), which shows blue color within 1 hr in X-gal assay. The self-association of SOP-2(bx91) is as weak as the interaction between human Rb and E2F-1 control (control B), which shows very faint blue to white color after 24 hr in X-gal assay.

sop-2::gfp Reporter and Transgenic Line Construction

sop-2::gfp reporter *EM#309* consists of the *sop-2* coding region with GFP inserted between the second and third codon as well as all noncoding 5′ and 3′ flanking sequence (1.5 kb and 1.3 kb, respectively) up to the next gene. *EM#309* was made by inserting PCR-generated fragments of 1.5 kb (C50E10 nt positions 37182–38827) and 5.7 kb (C50E10 nt position 38868 to C44C11 nt position 1917) into the SacII/NotI and NheI/StuI sites of pPD114.108 (a gift from A. Fire, Carnegie Institution of Washington, Baltimore, MD), respectively. Mutation P633S or W643Stop was introduced into *EM#309* by PCR-based mutagenesis.

bxEx99 and *bxEx103* were generated independently by coinjection of *EM#309* (10 ng/μl) and pBX-1*(pha-1+)* (100 ng/μl) DNA into *pha-1(e2123ts); him-5(e1490)*. *bxEx99* and *bxEx103* arrays produced essentially identical SOP-2::GFP expression patterns.

Acknowledgments

We thank J. Ross and D. Zarkower for communicating their results prior to publication and for helpful discussions, D. Portman and R. Palmer for helpful comments on the manuscript, and S. van den Heuvel and A. Hart for providing laboratory space to finish some experiments. This work was supported by the NIH (2T32CA09216 to H.Z., CA58596 to D.A.H., and GM39353 to S.W.E.), NARSAD (Young Investigators Award to R.L.), and EMBO (Postdoctoral Fellowship to R.B.R.A.). S.W.E. is the Siegfried Ullmann Professor of Molecular Genetics.

Received: January 23, 2003
Revised: March 18, 2003
Accepted: March 20, 2003
Published: June 2, 2003

References

Aboobaker, A., and Blaxter, M. (2003). Hox gene loss during dynamic evolution of the nematode cluster. Curr. Biol. *13*, 37–40.

Aguinaldo, A.M.A., Turbeville, J.M., Linford, L.S., Rivera, M.C., Garey, J.R., Raff, R.A., and Lake, J.A. (1997). Evidence for a clade of nematodes, arthropods and other moulting animals. Nature *387*, 489–493.

Barr, M.M., and Sternberg, P.W. (1999). A polycystic kidney-disease gene homologue required for male mating behaviour in *C. elegans*. Nature *401*, 386–389.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., and Sonnhammer, E.L.L. (2002). The Pfam protein families database. Nucleic Acids Res. *30*, 276–280.

Bornemann, D., Miller, E., and Simon, J. (1996). The *Drosophila* Polycomb group gene Sex comb on midleg (Scm) encodes a zinc finger protein with similarity to polyhomeotic protein. Development *122*, 1621–1630.

Bujnicki, J.M., Elofsson, A., Fischer, D., and Rychlewski, L. (2001). Structure prediction meta server. Bioinformatics *17*, 750–751.

Chakrabarti, S.R., Sood, R., Nandi, S., and Nucifora, G. (2000). Posttranslational modification of TEL and TEL/AML1 by SUMO-1 and cell-cycle-dependent assembly into nuclear bodies. Proc. Natl. Acad. Sci. USA *97*, 13281–13285.

Chamberlin, H.M., and Thomas, J.H. (2000). The bromodomain protein LIN-49 and trithorax-related protein LIN-59 affect development and gene expression in *Caenorhabditis elegans*. Development *127*, 713–723.

Ch'ng, Q., and Kenyon, C. (1999). *egl-27* generates anteroposterior patterns of cell fusion in *C. elegans* by regulating Hox gene expression and Hox protein function. Development *126*, 3303–3312.

Colovos, C., and Yeates, T.O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. *2*, 1511–1519.

Cowing, D., and Kenyon, C. (1996). Correct Hox gene expression established independently of position in *Caenorhabditis elegans*. Nature *382*, 353–356.

Duboule, D. (1992). The vertebrate limb: a model system to study the Hox Hom gene network during development and evolution. Bioessays *14*, 375–384.

Ferreira, H.B., Zhang, Y.H., Zhao, C.N., and Emmons, S.W. (1999). Patterning of *Caenorhabditis elegans* posterior structures by the *Abdominal-B* homolog, *egl-5*. Dev. Biol. *207*, 215–228.

Francis, N.J., and Kingston, R.E. (2001). Mechanisms of transcriptional memory. Nat. Rev. Mol. Cell Biol. *2*, 409–421.

Gellon, G., and McGinnis, W. (1998). Shaping animal body plans in development and evolution by modulation of Hox expression patterns. Bioessays *20*, 116–125.

Gleason, J.E., Korswagen, H.C., and Eisenmann, D.M. (2002). Activation of Wnt signaling bypasses the requirement for RTK/Ras signaling during *C. elegans* vulval induction. Genes Dev. *16*, 1281–1290.

Holdeman, R., Nehrt, S., and Strome, S. (1998). MES-2, a maternal

protein essential for viability of the germline in *Caenorhabditis elegans*, is homologous to a *Drosophila* Polycomb group protein. Development *125*, 2457–2467.

Huelsenbeck, J.P., and Ronquist, F. (2001). MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics *17*, 754–755.

Hunter, C.P., Harris, J.M., Maloof, J.N., and Kenyon, C. (1999). Hox gene expression in a single *Caenorhabditis elegans* cell is regulated by a caudal homolog and intercellular signals that inhibit Wnt signaling. Development *126*, 805–814.

Hunter, C.P., and Kenyon, C. (1995). Specification of anteroposterior cell fates in *Caenorhabditis elegans* by *Drosophila* Hox proteins. Nature *377*, 229–232.

Jacobs, J.J.L., and van Lohuizen, M. (2002). Polycomb repression: from cellular memory to cellular proliferation and cancer. Biochim. Biophys. Acta *1602*, 151–161.

Katoh-Fukui, Y., Tsuchiya, R., Shiroishi, T., Nakahara, Y., Hashimoto, N., Noguchi, K., and Higashinakagawa, T. (1998). Male-to-female sex reversal in M33 mutant mice. Nature *393*, 688–692.

Kenyon, C.J., Austin, J., Costa, M., Cowing, D.W., Harris, J.M., Honigberg, L., Hunter, C.P., Maloof, J.N., Muller-Immergluck, M.M., Salser, S.J., et al. (1997). The dance of the Hox genes: patterning the anteroposterior body axis of *Caenorhabditis elegans*. Cold Spring Harb. Symp. Quant. Biol. *62*, 293–305.

Kim, C.A., Gingery, M., Pilpa, R.M., and Bowie, J.U. (2002). The SAM domain of polyhomeotic forms a helical polymer. Nat. Struct. Biol. *9*, 453–457.

Kim, C.A., Phillips, M.L., Kim, W., Gingery, M., Tran, H.H., Robinson, M.A., Faham, S., and Bowie, J.U. (2001). Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. EMBO J. *20*, 4173–4182.

Korf, I., Fan, Y.A., and Strome, S. (1998). The Polycomb group in *Caenorhabditis elegans* and maternal control of germline development. Development *125*, 2469–2478.

Korswagen, H.C., Coudreuse, D.Y.M., Betist, M.C., van de Water, S., Zivkovic, D., and Clevers, H.C. (2002). The Axin-like protein PRY-1 is a negative regulator of a canonical Wnt pathway in *C. elegans*. Genes Dev. *16*, 1291–1302.

Kyba, M., and Brock, H.W. (1998). The SAM domain of polyhomeotic, RAE28, and scm mediates specific interactions through conserved residues. Dev. Genet. *22*, 74–84.

Lichtneckert, R., Muller, P., Schmid, V., and Reber-Muller, S. (2002). Evolutionary conservation of the chromatin modulator Polycomb in the jellyfish *Podocoryne carnea*. Differentiation *70*, 422–428.

Maloof, J.N., and Kenyon, C. (1998). The Hox gene *lin-39* is required during *C. elegans* vulval induction to select the outcome of Ras signaling. Development *125*, 181–190.

Maloof, J.N., Whangbo, J., Harris, J.M., Jongeward, G.D., and Kenyon, C. (1999). A Wnt signaling pathway controls Hox gene expression and neuroblast migration in *C. elegans*. Development *126*, 37–49.

Netter, S., Faucheux, M., and Theodore, L. (2001). Developmental dynamics of a polyhomeotic-EGFP fusion in vivo. DNA Cell Biol. *20*, 483–492.

Peterson, K.J., and Eernisse, D.J. (2001). Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA gene sequences. Evol. Dev. *3*, 170–205.

Raaphorst, F.M., Otte, A.P., and Meijer, C. (2001). Polycomb-group genes as regulators of mammalian lymphopoiesis. Trends Immunol. *22*, 682–690.

Ross, J.M., and Zarkower, D. (2003). Polycomb group regulation of Hox gene expression in *C. elegans*. Dev. Cell *4*, this issue, 891–901.

Ruvkun, G., and Hobert, O. (1998). The taxonomy of developmental control in *Caenorhabditis elegans*. Science *282*, 2033–2041.

Salser, S.J., and Kenyon, C. (1996). A *C. elegans* Hox gene switches on, off, on and off again to regulate proliferation, differentiation and morphogenesis. Development *122*, 1651–1661.

Satijn, D.P.E., and Otte, A.P. (1999). Polycomb group protein complexes: do different complexes regulate distinct target genes? Biochim. Biophys. Acta *1447*, 1–16.

Schultz, J., Ponting, C.P., Hofmann, K., and Bork, P. (1997). SAM as a protein interaction domain involved in developmental regulation. Protein Sci. *6*, 249–253.

Simon, J.A., and Tamkun, J.W. (2002). Programming off and on states in chromatin: mechanisms of Polycomb and trithorax group complexes. Curr. Opin. Genet. Dev. *12*, 210–218.

Stoyanov, C., Fleischmann, M., Suzuki, Y., Tapparel, N., Gautron, F., Streit, A., Wood, W., and Müller, F. (2003). Expression of the *C. elegans labial* orthologue *ceh-13* during male tail morphogenesis. Dev. Biol., in press.

Sze, J.Y., Victor, M., Loer, C., Shi, Y., and Ruvkun, G. (2000). Food and metabolic signaling defects in a *Caenorhabditis elegans* serotonin-synthesis mutant. Nature *403*, 560–564.

Wittmann, C., Bossinger, O., Goldstein, B., Fleischmann, M., Kohler, R., Brunschwig, K., Tobler, H., and Muller, F. (1997). The expression of the *C. elegans labial*-like Hox gene *ceh-13* during early embryogenesis relies on cell fate and on anteroposterior cell polarity. Development *124*, 4193–4200.

Yang, Z.H. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. *13*, 555–556.

Yang, Z.H., Nielsen, R., and Hasegawa, M. (1998). Models of amino acid substitution and applications to mitochondrial protein evolution. Mol. Biol. Evol. *15*, 1600–1611.

Zhang, H., and Emmons, S.W. (2000). A *C. elegans* mediator protein confers regulatory selectivity on lineage-specific expression of a transcription factor gene. Genes Dev. *14*, 2161–2172.