# COG-2, a Sox domain protein necessary for establishing a functional vulval-uterine connection in *Caenorhabditis elegans*

**Wendy Hanna-Rose and Min Han\***

Howard Hughes Medical Institute and Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309, USA

*Author for correspondence (e-mail: mhan@colorado.edu)

## SUMMARY

In screens for mutants defective in vulval morphogenesis, multiple mutants were isolated in which the uterus and the vulva fail to make a proper connection. We describe five alleles that define the gene *cog-2*, for <u>c</u>onnection of <u>g</u>onad defective. To form a functional connection between the vulva and the uterus, the anchor cell must fuse with the multinucleate uterine seam cell, derived from uterine cells that adopt a π lineage. In *cog-2* mutants, the anchor cell does not fuse to the uterine seam cell and, instead, remains at the apex of the vulva, blocking the connection between the vulval and uterine lumens, resulting in an egg-laying defective phenotype. According to lineage analysis and expression assays for two π-cell-specific markers, induction of the π fate occurs normally in *cog-2* mutants. We have cloned *cog-2* and shown that it encodes a Sox family transcription factor that is expressed in the π lineage. Thus, it appears that COG-2 is a transcription factor that regulates a late-stage aspect of uterine seam cell differentiation that specifically affects anchor cell-uterine seam cell fusion.

Key words: *cog-2*, Sox, *Caenorhabditis elegans*, Vulva, Morphogenesis

## INTRODUCTION

The elucidation of organogenesis, in which groups of cells with various identities coordinate to form structurally intact and functional organs, is clearly important to the understanding of the development of multicellular organisms. Since the structure of the *C. elegans* vulva is well-defined at the anatomical level and the process of cell-fate determination in the vulva is well-characterized, vulva development provides an ideal system to study this problem.

*C. elegans* vulval development is an elaborate process requiring the generation of vulval precursor cells (VPCs), the determination of specific vulval cell fates, and finally the execution of vulval fates which involves an intricate morphogenesis process (Horvitz and Sternberg, 1991). VPC generation requires the action of the homeobox gene *lin-39* in the first larval stage, allowing a group of six cells, P3.p-P8.p, to maintain competency to adopt vulval fates (Clark et al., 1993; Salser et al., 1993). Vulval cell fate determination occurs during the third larval stage when three of the VPCs, P5.p-P7.p, are induced by a RTK/Ras/MAPK-mediated signal from the anchor cell (AC) and then divide to form the 22 vulval cells. P6.p adopts a 1° vulval fate and produces eight progeny. P5.p and P7.p adopt a 2° vulval fate and produce seven progeny each. The 1° versus 2° fate decision is determined by each VPC in response to the RTK/Ras/MAPK-mediated pathway activated by the inductive signal as well as *lin-12* mediated, lateral signals between VPCs (reviewed by Sternberg and Han, 1998).

During the execution phase of vulval development, in the third and fourth larval stages, the 22 descendants of the VPCs cooperate in a complex morphogenesis process to form the mature vulva. The most distal vulval cells migrate towards the center of the body and begin to invaginate. Ultimately, a stacked series of ring-like, multinucleate cells forms the tube-like structure connecting the uterus to the outside of the body (Fig. 1; J. White, personal communication). Vulval cells also form attachments to the uterus, the sex muscles and the hypodermis during this period (J. White, personal communication).

The AC plays multiple, essential roles during vulval development (Fig. 1). It invades the basement membrane surrounding the gonad and serves as the source of the inductive signal that promotes vulval fates in the VPCs (Hill and Sternberg, 1992; Fig. 1). In addition, it plays an integral role in establishing a functional connection between the vulva and the uterus. The AC signals six of the twelve granddaughters of the ventral uterine cells, causing them to adopt a fate, π, different from the default fate, ρ, of their six sisters and cousins (Fig. 1; Newman et al., 1995). The six π precursors (three per side) divide to form twelve π cells. Eight of the π cells (four per side) fuse to form the large multinucleate uterine seam (utse) cell that lies over the top of the vulva orifice and underlies the developing uterus (Newman et al., 1996). The remaining four π cells (two per side) adopt the uv1 fate and also make connections with vulval cells (Newman et al., 1996). The AC must also penetrate the dorsal vulval toroid cell (vulF) to form a channel through the top of the vulva (Fig. 1). This

places the AC in a position that blocks the entrance to the vulva from the uterus. The AC then plays a final essential part by fusing with the utse cell (Fig. 1; Newman et al., 1996). Its nucleus and the eight nuclei of the utse cell move to the distal corners of the utse cell, leaving a thin layer of cytoplasm overlaying the vulval orifice, separating the vulval and uterine lumens. This thin laminar process can presumably be broken when the first egg is laid to establish a direct channel between the two organs (Newman et al., 1996).

The complexity of the morphogenesis process as well as previous genetic studies suggest that vulval morphogenesis requires the action of many gene products (Seydoux et al., 1993). We have initiated a genetic dissection of organogenesis by focusing on one of the last steps in vulval morphogenesis, the formation of a functional connection between the vulva and the uterus. We have cloned and characterized a gene, *cog-2*, that plays a key role in the AC-utse cell fusion. In the absence
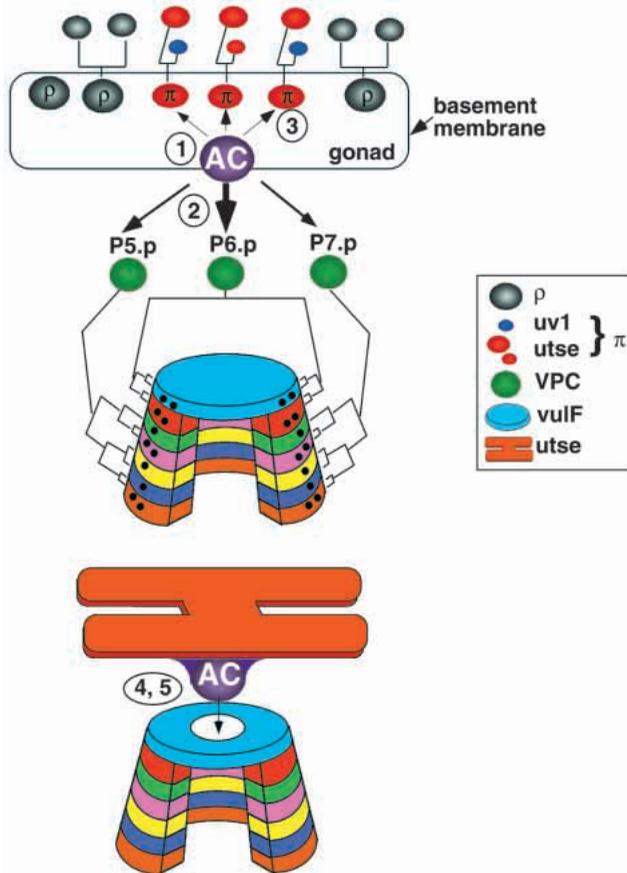


**Fig. 1.** The AC plays multiple roles during vulval differentiation and morphogenesis. (1) The AC invades the basement membrane surrounding the gonad to lie in close proximity to the vulval precursor cells (VPC). (2) In L3, a signal from the AC induces three VPCs, P5.p, P6.p and P7.p, to adopt vulval fates. (3) In L3, a signal from the AC promotes a π fate, as opposed to a default ρ fate, in three VU granddaughters on each side of the gonad. A subset of the π cells fuse to form the uterine seam cell (utse) and the remaining π cells adopt a uv1 fate. (4) In early L4, the AC penetrates the dorsal vulval toroid cell, vulF, forming a channel through the top of the vulva. (5) The AC fuses with the multinucleate utse cell to clear the vulva orifice and unblock the channel between the vulva and the uterus.

of *cog-2* gene activity, the AC performs its initial functions normally so that all the components for establishing the vulval-uterine connection (the vulval cells and the π cells) are coordinated properly in time and space. However, fusion between the AC and utse cell is defective, preventing the formation of a functional vulva.

## MATERIALS AND METHODS

### Strains and genetic methods

Methods for manipulating the wild-type Bristol N2 and N2-derived strains were as described by Brenner (1974). Strains were maintained and experiments were conducted at 20°C, unless otherwise noted. Mutations used in this study, described by Riddle et al. (1997), are: *unc-119(ed3)* III, *dpy-20(e1282, e1362)* IV, *dpy-8(e130)* X, *lon-2(e678)* X, *mec-2(e75)* X, *mom-1(zu188)* X, and *unc-6(e78)* X. The *lin-11*::gfp line contains *mgIs21*[pRF4; plin-11-ABCDE-GFP] (Hobert et al., 1998). The *cdh-3*::gfp construct is from strain NL1008: *dpy-20(e1362)* IV; *pkEX246* [dpy-20(+); *cdh-3*::gfp] (Pettitt et al., 1996).

### Isolation of *cog-2* mutants

*cog-2* alleles were isolated in screens for temperature-sensitive (TS) and non-TS mutants. The *cog-2* alleles are non-TS. N2 L4 hermaphrodites were treated with 50 mM ethylmethane sulfonate (Brenner, 1974) for 4 hours, cloned and maintained at 15°C. A synchronized population of $F_2$ animals was prepared by picking 30 adult $F_1$ hermaphrodites to each of multiple plates and allowing them to lay eggs for only 3 hours. The $F_2$ eggs were allowed to hatch at 15°C and develop to the L2 stage. $F_2$ progeny were then switched to 25°C for approximately 15 hours, during the stages of vulval induction and morphogenesis, and then returned to 15°C. $F_2$ progeny were screened after 2-3 days for egg-laying defective (Egl) hermaphrodites in which progeny have hatched inside the Egl mother. This population of mutants was subjected to a secondary screen in which the morphology of the vulva was directly observed using Nomarski optics to identify mutants with vulval morphogenesis defects. Five recessive alleles that define *cog-2* were identified, among others. *ku205*, *ku209* and *ku241* all fail to complement *ku207*. *ku205* and *ku194* are rescued by the same cosmid. *ku194*, *ku207* and *ku241* were outcrossed at least 5 times and further characterized. *ku205* and *ku209* hermaphrodites fail to mate and have not been extensively outcrossed. Since *ku205* and *ku209* were not extensively outcrossed, only the phenotypic characterizations of *ku194*, *ku207* and *ku241* are reported. However, characterization of *ku205* and *ku209* indicated that these alleles are similar in severity to *ku194* (data not shown). *cog-2* mutant males have no apparent defects and mate with high efficiency (data not shown).

### Genetic mapping and rescue of *cog-2*

*ku194* was mapped to the X chromosome between the genetic markers *mec-2* and *unc-6*. 27/47 Lon non/Unc and 10/19 Unc non-Lon recombinants from *lon-2 unc-6/cog-2* heterozygotes segregated *cog-2*. 34/37 Lon non-Cog and 1/6 Cog non-Lon recombinants from *lon-2 cog-2/mec-2* heterozygotes segregated *mec-2*. Transgenic strains were generated by germline transformation (Mello et al., 1991). Cosmids were obtained from A. Coulson (Sanger Center). For rescue experiments, 5-10 μg/ml of cosmid or plasmid was coinjected with 100 μg/ml of the transformation marker *sur-5*::gfp, pTG96 (Gu et al., 1998), into *ku194/dpy-8 mom-1 unc-6* heterozygotes. Clones from each line that stably expressed GFP and that no longer segregated Dpy Unc animals were selected to obtain transgenic homozygous *ku194* lines. Positive rescuing activity was scored as restoration of egg-laying ability to transgenic *ku194* hermaphrodites. Other transgenic lines were obtained using unc-119(+) (Maduro and Pilgrim, 1995) as a transformation marker to inject *unc-119(ed3)* hermaphrodites. Extrachromosomal arrays were integrated as described by Mello and Fire (1995).

## *cog-2* cDNA cloning and sequencing

An 1816 base pair *Bam*HI-*Hin*dIII genomic fragment encoding all of the predicted T22B7.1 gene and part of the predicted W01C8.2 gene in the rescuing fragment was used to probe the Stratagene (La Jolla, CA) Nematode UniZAP®XR λ mixed-stage cDNA library. From one million plaques, five positive clones were isolated, excised and sequenced using an ABI automated sequencer. Two cDNAs appeared full-length. Three were incomplete cDNAs likely derived from the same transcript as the full-length clones. The full-length cDNA contained 5 base pairs derived from the SL1 spliced leader sequence, a 3 base pair 5′ UTR, a single open reading frame coding for 471 amino acids and a 672 base pair 3′ UTR. Molecular lesions were identified by sequencing PCR amplified genomic DNA from lysates of mutant worms.

## Plasmid construction

T22B7 was digested with *Kpn*I and religated to produce pKG6. The two largest *Pst*I-*Kpn*I fragments from pKG6 were separately ligated to pBluescript S/K+ to produce pKG6-3′ and pKG6-5′. The 5.8 kilobase pair *Eco*RI fragment from pKG6-3′ was cloned into pBluescript to make pWH4. pWH4 was ligated to the 8.5 kilobase pair *Kpn*I-*Sal*I fragment from pKG6-3′ to make pWH5. pRC1 was created by digesting pWH4 with *Afl*II and religating and then ligating to the *Kpn*I-*Sal*I fragment as above. pRC2 was created by digesting pWH4 with *Sal*I and religating, digesting with *Hin*dIII and *Eco*RI, creating blunt ends with Klenow and then religating, and finally, cloning in the *Kpn*I-*Sal*I fragment, as above.

To make the *cog-2*::gfp translational fusion, pWH14, a fragment that contains sequences including the *Hin*dIII site in the seventh exon through the last amino acid was amplified using primers that introduce a *Not*I and a *Bam*HI site, 5′GCGATGGTATTATTCTAAAGCT-TTCAC3′ and 5′CGGGATCCGCGGCCGCTTCAGCTGTTTGTA-GGAGATGATGTGA3′, and cloned into the *Hin*dIII and *Bam*HI sites of pBluescript. This plasmid was ligated to a 2.3 kilobase pair *Sal*I-*Hin*dIII fragment from pWH5 containing the remaining *cog-2* coding sequence and part of the promoter. The remainder of the promoter was ligated upstream in the form of a 8.5 kilobase pair *Kpn*I-*Sal*I fragment from pWH5. The resulting construct was digested with *Not*I and ligated to a fragment with *Eag*I ends containing the *cog-2* 3′ end amplified from cosmid W01C8 and having *Apa*I and *Pst*I sites introduced 5′ to the UTR. The primers used in this amplification were 5′CGCGGCCGGGGGCCCAATCTGCAGGTAACTTTAACCAAGC GTGTCATTCC3′ and 5′GCGG-CCGCGCGGCCGCACATTTCCT-ACCATAATCCG3′. Finally, to produce pWH14, an *Apa*I-*Pst*I fragment from pPD102.33 containing a GFP cassette (A. Fire, S. Xu, J. Ahnn and G. Seydoux, personal communication) was cloned into the vector cut with *Apa*I and *Pst*I.

To produce pWH17, the *cog-2*::gfp transcriptional fusion, pWH14 was digested with *Not*I and *Hin*dIII to release a 7.4 kilobase pair fragment containing the vector, the promoter and the first 19 codons of *cog-2*. This fragment was ligated to a 1.9 kilobase pair *Hin*dIII-*Eag*I fragment from pPD95.69 (A. Fire, S. Xu, J. Ahnn and G. Seydoux, personal communication) to produce a *cog-2*::gfp transcriptional fusion containing the *unc-54* 3′ end. The *cog-2*::gfp translational fusion was injected into *unc-119(ed3)* worms along with an unc-119(+) marker to allow selection of transformed animals and the fluorescence pattern was determined in a non Cog background.

## RESULTS

### *cog-2* mutants have a defective vulval-uterine connection

To identify vulval morphogenesis defective mutants, the F$_2$ generation of mutagenized animals was screened for Egl mutants. All Egl mutants were subjected to a secondary screen in which the vulva structure was examined directly using Nomarski optics to identify mutants with normal vulval lineages but specific structural defects indicative of abnormal morphogenesis. Eight mutants that had defects in establishing a functional vulval-uterine connection were identified in screens covering 13,000 haploid genomes. Five of these alleles define a previously unidentified gene, *cog-2* (connection of gonad defective). The remaining three alleles define two additional *cog* genes that will be described elsewhere.

*cog-2* mutant hermaphrodites display a highly penetrant Egl phenotype, an incompletely penetrant protruding vulva phenotype, and show a diminished brood size (Table 1). Any Egl defect may result indirectly in a small brood size because fertilized eggs hatch inside the hermaphrodite and kill it before the end of its reproductive age. However, a more direct effect of *cog-2* on brood size is suggested by the low frequency of *ku194* and *ku207* hermaphrodites that have no progeny and by the small brood size of the *ku207* homozygous animals that do lay eggs (approximately 134, compared to the wild-type brood size of 300) (Table 1) (Wood, 1988). Even though all alleles are recessive, we were not able to determine the genetic nature of the *cog-2* alleles because *cog-2* is not uncovered by genetic deficiencies in the region (data not shown). However, *ku194* is a likely molecular null (see below), consistent with our analysis of its highly penetrant Egl phenotype.

Additionally, a gonad migration defect is associated with all *cog-2* alleles between two and four times more often than in a wild-type background (Table 2). 13% of gonad arms follow abnormal migration patterns in *ku194* hermaphrodites (Table 2). In 2 of 11 abnormal migrations that were scored, the gonad reflexed prematurely. In the remaining 9 abnormal cases, the arms migrated distally for the correct time and distance and then reflexed and migrated proximally again as expected.

**Table 1. Gross phenotype of *cog-2* mutants**

| Allele | % Egl (*n*) | % Pvul (*n*) | % Sterile (*n*) | Brood size ±s.d.* (*n*) |
|---|---|---|---|---|
| *ku194* | 96 (93) | 44 (183) | 4 (93) | 28±11 (55) |
| *ku207* | 96 (75)‡ | 44 (75) | 1 (75) | 54±13 (48)‡ |
| *ku241* | 100 (94) | 26 (76) | 0 (76) | 61±11 (58)§ |

% Egl refers to percentage of hermaphrodites completely unable to lay eggs.
% Pvul refers to percentage of hermaphrodites with a protruding vulva.
% Sterile refers to the percentage of hermaphrodites that produce no progeny.
Brood size was determined by counting all progeny that emerged from the corpse of an Egl mother.
*Not including sterile or Egl(+) animals.
‡Two hermaphrodites were Egl(+) and had broods of 131 and 137.
§The two-fold difference between *ku194* and *ku241* brood sizes is statistically significant (*P*<0.0001, Mann-Whitney U test).

**Table 2. Nomarski phenotypes of *cog-2* mutants**

| Allele | % Normal mid-L4 morphology* | % AC block (n)‡ | % Grossly abnormal morphogenesis (n)** | % Gonad arms migration defective (n)‡‡ |
|---|---|---|---|---|
| N2 | 97 (32) | 0 (32) | 3 (32) | 3 (108) |
| *ku194* | 0 (53) | 81 (53) | 19 (53) | 13 (82) |
| *ku207* | 1 (82) | 85 (82)§,¶ | 13 (82) | 12 (129) |
| *ku241* | 0 (61) | 95 (61) | 5 (61) | 6 (162) |

*% of L4 hermaphrodites with normal vulval morphology and with a clearly visible utse cell cytoplasmic process at the apex of the vulva under Nomarski optics.

‡% of L4 hermaphrodites with normal vulval morphology except that the AC nucleus was still present at the apex of the vulva in the place of the utse cell cytoplasmic process.

**% of hermaphrodites with grossly abnormal vulva morphology from stages L3 to adult. It is not clear if these animals have the AC-block phenotype as well.

‡‡% of gonad arms that followed an abnormal path of migration as observed under Nomarski optics at the L4 stage.

§4% (*n*=54) of hermaphrodites had normal AC penetration of vulF.

¶6% (*n*=82) had a large, swollen AC.

However, they often followed an abnormally ventralized or dorsalized path. One of these 9 abnormal *ku194* gonad arms failed to move dorsally at all, three moved dorsally too soon (Fig. 2D), two migrated dorsally and then returned abnormally to the ventral side to continue the migration, and three briefly migrated either dorsally or ventrally at the improper time and then returned to their normal migration path.

To investigate the Egl phenotype, we examined by Nomarski optics the AC, VPCs, and uterine cells in *cog-2* mutants at various developmental stages. The vulval structure appears normal at the L3 molt in *cog-2* mutants. All 22 nuclei are positioned properly and the correctly shaped invagination is formed. However, as organogenesis proceeds throughout L4, the vulva structure becomes obviously abnormal (Fig. 2). In wild-type, mid-L4 hermaphrodites, a thin planar process derived from utse cell cytoplasm can be seen in a lateral view lying across the top of the vulval lumen, separating the vulval and uterine lumens (Fig. 2A). The AC that formally occupied the top of the vulva has fused with the utse cell, and its nucleus has migrated distally to clear the orifice (Newman et al., 1996). In *cog-2(ku194)* mutants, the thin planar process at the top of the L4 vulva is not visible, and the AC nucleus abnormally appears in its place (Fig. 2B). *cog-2* mutants with normal vulval morphogenesis are rarely observed (Table 2). The AC-block phenotype is highly penetrant (81-95%) for all *cog-2* alleles (Table 2), and the remainder of the mutants (5-19%) have more grossly affected vulval structures (Fig. 2D,

and Table 2). A majority of the animals with severe defects in vulval morphology also have defects in gonad migration, although this correlation is not absolute (Fig. 2D). Frequently, it also appears that the AC fails to form a channel through the vulF cell in *cog-2* mutants because the anterior and posterior regions of the vulva come into close contact at the apex rather than appearing widely separated as in the wild-type vulva (compare Fig. 2A and B).

## The AC and the utse cell do not fuse in *cog-2* mutants

Because the AC nucleus fails to move away from the vulval orifice to the corner of the utse cell in *cog-2* mutants, we investigated whether the AC-utse cell fusion occurs. A *cdh-3*::gfp reporter was used to monitor fusion in wild-type and
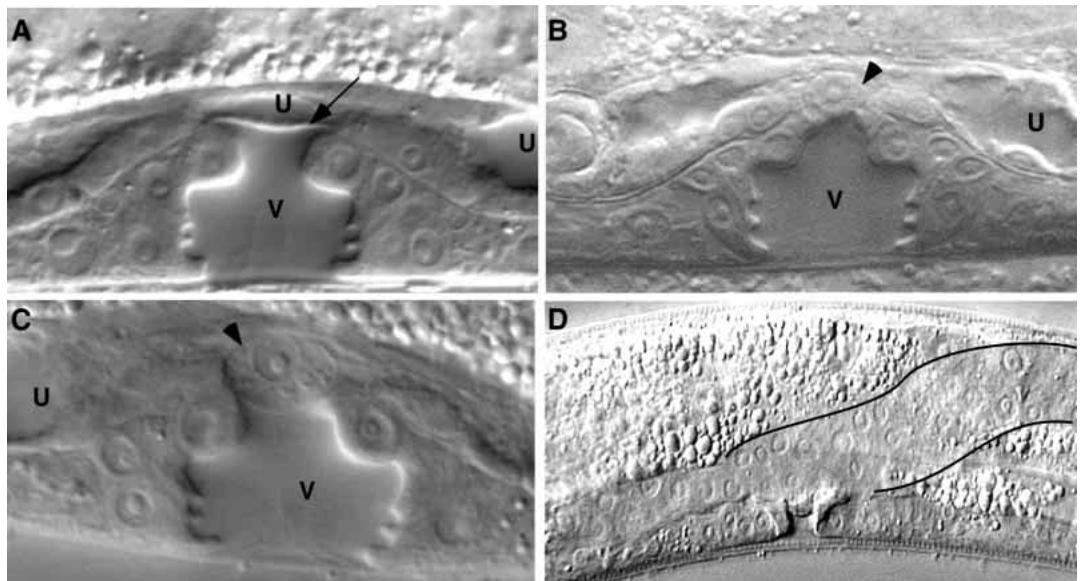


**Fig. 2.** *cog-2* mutants fail to make a proper connection between the vulva and the uterus. Nomarski photomicrograph (using 100× objective) of the lateral side of (A) a wild-type mid-L4 stage vulva (arrow indicates utse cell cytoplasm). (B) A *cog-2(ku194)* mid-L4 stage vulva manifesting an AC-block phenotype. Note that the AC (arrowhead) failed to fully penetrate the vulval apex. (C) A *cog-2(ku207)* mid-L4 stage vulva manifesting an AC-block phenotype and a swollen AC. The AC (arrowhead) successfully penetrated the vulval apex in this mutant. (D) *cog-2(ku194)* L3 stage vulva with an abnormal vulval morphogenesis defect and a gonad migration defect (using 63× objective). The abnormally dorsalized gonad migration path is outlined. U, uterine lumen; V, vulva lumen.

mutant worms (Pettitt et al., 1996). *cdh-3*::gfp carried on an extrachromosomal array is expressed in wild-type hermaphrodites during L3 in the cytoplasm of the AC, resulting in a distinct spot of fluorescence at the apex of the vulva (Fig. 3A,B). As the AC fuses to the utse cell during L3 lethargus, fluorescence spreads from the AC cytoplasm throughout the larger utse cell, resulting in a diffuse fluorescent pattern above the vulva (Fig. 3C,D). In *cog-2(ku207)* mutants, the fluorescence pattern of the CDH-3::GFP fusion protein is observed in the area of the AC cytoplasm in L3 hermaphrodites as is seen in wild-type hermaphrodites. However, in L4 hermaphrodites, after fusion should have occurred, the distinct, AC pattern of expression is still visible (Fig. 3E,F), and AC fluorescence at the apex of the vulva continues into the adult stages of development (Fig. 3G,H). The diffuse staining pattern indicative of AC-utse cell fusion is rarely seen. 30 of 30 wild-type worms displayed the fluorescence pattern indicative of fusion. In contrast, 26 of 33 L4 and young adult *cog-2(ku207)* hermaphrodites displayed a distinct fluorescence pattern indicative of an unfused AC, and the remaining 7 animals displayed an ambiguous expression pattern. These results suggest that the AC fails to fuse with the utse cell in *cog-2* mutants.

## π cell fate appears normal in *cog-2* mutants

In *cog-2(ku207)* mutants, the AC nucleus occasionally adopts an abnormal morphology, appearing large and bloated (Fig. 2C and Table 2). A similar bloated AC phenotype was described previously, under conditions in which the π cells had been ablated (Newman et al., 1996). Because both π cell ablation and a *cog-2* mutation can lead to a similar bloated AC phenotype and because the AC could fail to fuse with the utse cell due to a defect in the π cell lineage from which the utse cell is derived, we investigated the status of the π cell lineage in *cog-2* mutants. π cell precursors that are not induced by a signal from the AC adopt the default ρ lineage. A π lineage is easily distinguishable from a ρ lineage based on the characteristic number and orientation of divisions (Newman et al., 1995). π cell precursors divide once with a dorsal-ventral division plane to produce a larger, dorsal cell and a smaller, ventral cell (Fig. 4). In contrast, two divisions in a mostly anterior-posterior plane are characteristic of a ρ lineage (with the exception of the outer ρ precursor on each side of the animal, each of which undergoes an initial lateral division to restore symmetry to the uterus) (Kimble and Hirsh, 1979; Newman et al., 1995).

In *cog-2* mutants, the π cell precursors adopt a π fate, rather than the default ρ fate, based on their division patterns (Fig. 4). All three π precursors per side of the gonad were lineaged from 21 sides of *ku194* hermaphrodites for a total of 63 cells lineaged. 56 of these cells underwent an asymmetric division in a dorsal-ventral plane (Fig. 4). Furthermore, two worms in which normal lineages were observed for all three π cells on both sides of the gonad were recovered from the slide after lineaging, and these animals manifested a completely Egl phenotype suggesting that the Cog phenotype of *ku194* mutants is the result of something other than a lack of π cell induction. A single π cell was observed to divide in a manner reminiscent of a ρ lineage, and the remaining 6/63 cells were from a single worm in which both the π and ρ precursors failed to divide by the beginning of L4 (Fig. 4). This failure to divide phenotype does not appear to be the cause of the highly penetrant Egl phenotype associated with *ku194* animals. It is possible that the less penetrant grossly abnormal morphogenesis phenotype could be associated with a failure of these lineages to progress. Vulval lineages were normal in all *cog-2* mutants examined (data not shown).

Further support for the observation that normal π cell induction takes place in *cog-2* mutants is that the mutant π cells properly establish expression of two π-specific markers, *lin-11*::gfp and *cog-2*::gfp. LIN-11 is a LIM homeodomain transcription factor that is expressed in the π cells as well as in cells of the vulval secondary lineage (Freyd et al., 1990; Newman and Sternberg, 1996). An integrated array containing a *lin-11*::gfp fusion reporter with GFP fused in frame to the fifth exon of *lin-11* (Hobert et al., 1998) is expressed by wild-type hermaphrodites in the π cells at the beginning of L4 (Fig. 5A,B). *ku194* animals carrying the *lin-11*::gfp integrated array express LIN-11::GFP in a similar pattern, indicating that the π cell fate is intact with regard to expression of this lineage marker (Fig. 5C,D). Identical results were observed with a *lin-11*::lacZ reporter (data not shown). Since *cog-2* is also expressed in the π cells (see below), a *cog-2*::gfp transcriptional fusion was used to mark the π cells in wild-type (Fig. 5E,F) and *cog-2(ku194)* (Fig. 5G,H) hermaphrodites, and once again all six mutant π cells express this reporter (Fig. 5). Expression of these two π-cell-specific markers in *cog-2* mutants indicates that the π cell fate is not severely disrupted in *cog-2* mutants, but does not exclude a later utse cell-specific defect. In fact, a late defect is consistent with observations that while the expression of the *cog-2*::gfp transcriptional reporter remains visible during late L4 in the twelve π cells in a wild-type background, expression becomes fainter in these cells and additional faintly expressing cells are sometimes seen during late L4 and young adult stages in a *cog-2* mutant background (data not shown).

## *cog-2* encodes a Sox domain protein

*ku194* was mapped to the X chromosome between *mec-2* and *unc-6*. *cog-2* was cloned using microinjection transformation to rescue the Egl phenotype of *ku194*. Cosmid T22B7 and subclones containing the 3′ end of T22B7 can efficiently restore the ability of transgenic *ku194* hermaphrodites to lay eggs (Fig. 6A). Genefinder (Wilson et al., 1994) predicted two transcripts in this region, T22B7.1 and W01C8.2. However, small deletions in either predicted gene eliminated rescuing activity (Fig. 6A). By screening a cDNA library using a probe that included portions of both predicted transcripts, we identified a single cDNA class. The complete *cog-2* cDNA contains the last three exons of T22B7.1 and all five exons of W01C8.2 (with some slight differences from the predicted intron-exon boundaries in both cases), as well as two additional exons from 1.7 and 3 kilobase pairs upstream that were not predicted by Genefinder. *cog-2* appears to be transpliced to SL1, contains 10 exons and encodes a predicted protein of 471 amino acids (Figs 6B, 7A). The 672 base pair 3′UTR identified in the full-length cDNA is dispensable for functional activity since the rescuing cosmid, T22B7, and the T22B7-derived rescuing plasmids end just 31 base pairs downstream of the predicted stop codon, thereby removing most of the 3′ UTR.

The predicted product of the *cog-2* gene is homologous to a large family of putative, transcription factors called SOX (human) or Sox (murine) domain proteins. All Sox proteins have a single high mobility group (HMG) box with significant

similarity to the HMG box of the mammalian sex determining factor SRY (Laudet et al., 1993). COG-2 is most similar to the group D Sox domain proteins Sox5 and Sox6 (Fig. 7B; Connor et al., 1995; Denny et al., 1992b; Prior and Walter, 1996). COG-2 is greater than 70% identical to Sox5 and Sox6 in the region of the DNA-binding, HMG box, and the COG-2 HMG box is 96% identical to the Sox family HMG consensus (Nambu and Nambu, 1996; Fig. 7B). However, like many other Sox proteins, COG-2 shows no similarity outside of the HMG box to any other Sox protein or any other protein in the data bases.

The location of the lesion associated with each of the five cog-2 alleles was determined. ku209 and ku194 are amber mutations that result in the truncation of over half of the protein including the Sox domain, consistent with the hypothesis that ku194 is a strong, loss-of-function allele and a likely molecular null (Figs 6, 7A). ku207 is a missense mutation resulting in an alanine to threonine substitution within the Sox domain (Figs 6, 7A). This alanine is highly conserved among HMG boxes in general, not just within the SRY related Sox family (Fig. 7B; Laudet et al., 1993). ku205 and ku241 are splice site mutations occurring prior to the region coding for the Sox domain (Figs 6, 7A). Failure to splice at either of these sites would again lead to a severely truncated protein. Unfortunately, splicing in these mutants could not be assayed because no cog-2 transcript could be detected using northern blot analysis (data not shown).

### cog-2 is expressed in the π cells

We investigated the expression pattern of cog-2 since it seems to be acting specifically on fusion between the AC and the utse cell. We examined the expression pattern of cog-2 using a reporter, pWH14, that contains 6.4 kilobase pairs of sequence upstream of the cog-2 coding region, the entire coding region fused to GFP, and the 3′UTR of cog-2 (Fig. 8A). The cog-2::gfp translational fusion encodes a functional protein since it can partially rescue the Egl phenotype of ku194 mutants as an extrachromosomal array (data not shown).

Four stable lines containing an extrachromosomal array of the translational fusion were examined, and one line was integrated and reexamined. All lines shared consistent expression patterns. COG-2::GFP is a nuclear protein consistent with COG-2 homology to transcription factors. cog-2::gfp is expressed in a dynamic pattern in cells of the π lineage (Fig. 8B). No expression is visible in the gonad until the π cell precursors are born. At this point, expression is evident in all π cell precursors as well as in the single ρ precursor on each side that has a π cell precursor as a sister (Fig. 8B,C). After the π cell precursors undergo their single and final division, cog-2::gfp expression is still bright in the π cells, but expression is usually no longer evident in the ρ lineage (Fig. 8B,D). Most animals show no expression of COG-2::GFP expression past the ρ precursor stage in the ρ lineage, although in a few

animals, faint GFP expression was observed in the daughters, but never the granddaughters, of the expressing ρ precursor (Fig. 8B). cog-2::gfp expression remains strong in the π lineage throughout the L4 stage, becoming slightly brighter in the π cells that adopt the uv1 fate (Fig. 8B). In addition, the AC nucleus begins to express cog-2::gfp after fusion with the utse cell (not shown). Expression in the AC is not detectable before the time of fusion. cog-2::gfp expression is not detectable in the vulva at any stage, in any other part of the gonad including the distal tip cells, or in the hypodermis (data not shown). cog-2::gfp is expressed in most, if not all, neurons in the head and the tail of the animal (data not shown).
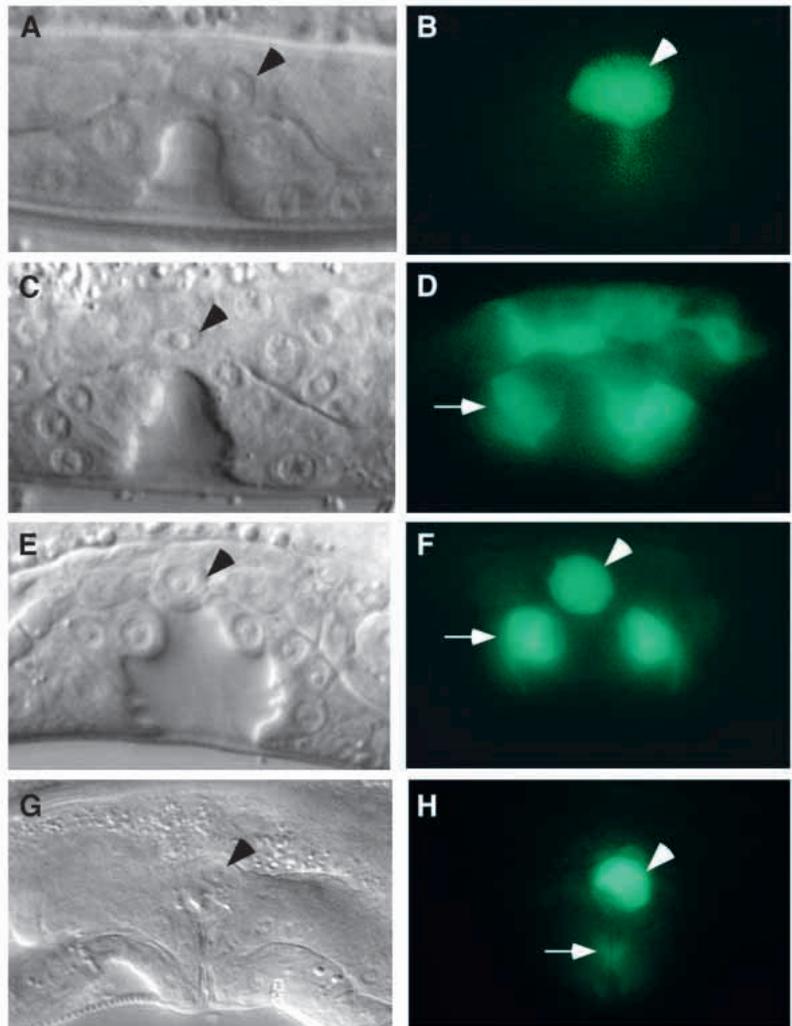


**Fig. 3**. The AC does not fuse to the utse cell in cog-2 mutants. Nomarski photomicrographs (A,C,E,G) and corresponding fluorescence images (B,D,F,H) of the lateral side of worms carrying an extrachromosomal cdh-3::gfp containing array. (A,B) A wild-type late L3 stage vulva prior to AC-utse cell fusion, showing expression of the fusion protein in the AC cytoplasm. (C,D) A wild-type, early L4 stage vulva after the AC has fused to the utse cell, showing diffusion of the fusion protein throughout the multinecleate utse cell. Expression is also visible in vulva cells at this time point. (E,F) A mid-L4 and (G,H) a young adult stage vulva after the normal AC-utse cell fusion time point in ku207, showing that GFP fusion protein expression is seen only in the AC cytoplasm and not in the utse cell in late stage vulvae of mutants. Again, expression is also visible in vulva cells at these late time points. Arrowhead, anchor cell; arrow, GFP expression in vulval cells.

**Fig. 4**. Lineage analysis indicates that the π precursors adopt a normal fate in *cog-2(ku194)* animals. A normal π lineage, which is characterized by an asymmetric division along a dorsal/ventral axis, an undivided pattern, and a default-like division pattern, which is characterized by a symmetric division along an anterior/posterior axis, are depicted at the top of the columns. The number of π precursor cells that adopted the indicated lineage is shown in the N2 and *ku194* rows as a fraction of the total number of cells observed. The 63 cells were lineaged in groups of three per side of the gonad in a total of 14 worms. Lineages could sometimes be followed only on one side of a hermaphrodite. Two *ku194* hermaphrodites in which a normal π lineage was observed on both sides of the animal were recovered, and they each manifested a completely Egl phenotype. ‡The six cells that failed to divide were from a single worm. *The single cell that divided with an anterior/posterior division axis was in a worm where the remaining five π cells adopted the proper lineage.

| # of putative π precursor cells that adopted: | | |
|---|---|---|
| a π–like division pattern | an undivided division pattern | a default (ρ-like) division pattern |
| **N2** 18/18 | 0/18 | 0/18 |
| **ku194** 56/63 | 6/63 ‡ | 1/63 * |

first 19 codons of the *cog-2* coding region and the *cog-2* 3′UTR and replaces them with a GFP cassette containing a NLS and the *unc-54* 3′ UTR. The *cog-2*::gfp transcriptional fusion is expressed in a similar pattern to the translational fusion in the π and ρ lineages. The expression pattern of the transcriptional fusion differs from the translational fusion in that the body wall muscles express the fusion protein at a higher level than any other cell and fewer neurons expressing *cog-2*::gfp can be detected with the transcriptional fusion, although the same brightly expressing cells that stand out with the translational fusion are visible.

However, four unidentified cells in the head express *cog-2*::gfp at a higher level than others (data not shown). Expression is also faintly visible in the body wall muscles and the intestinal cells (data not shown).

Because *cog-2* is expressed in the π cells, a *cog-2*::gfp transcriptional fusion was produced to use as a marker for the π cell fate. The transcriptional fusion construct deletes all but the

**Fig. 5**. The π cells express lineage-specific markers in *cog-2* mutants. Nomarski photomicrograph of lateral side of early L4 (A) wild-type or (C) *ku194* hermaphrodite carrying integrated *lin-11*::gfp containing array. (B,D) Corresponding fluorescence images showing that the π cells express the *lin-11*::gfp fusion protein in the wild-type and mutant backgrounds. *lin-11*::gfp is also expressed in the secondary lineage in the vulva. Nomarski photomicrograph of lateral side of early L4 (E) wild-type or (G) *ku194* hermaphrodite carrying an integrated *cog-2*::gfp containing array. (F,H) Corresponding fluorescence images showing π cells express the *cog-2*::gfp transcriptional fusion in the wild-type and mutant backgrounds. *cog-2*::gfp is also expressed in body wall muscle nuclei seen as a haze out of the plane of focus. Note that the animals depicted in E-H are slightly older than the animals depicted in A-D accounting for the more dispersed placement of the utse cell nuclei which normally begin to move away from the apex of the vulva shortly after they are born. *, π cells.
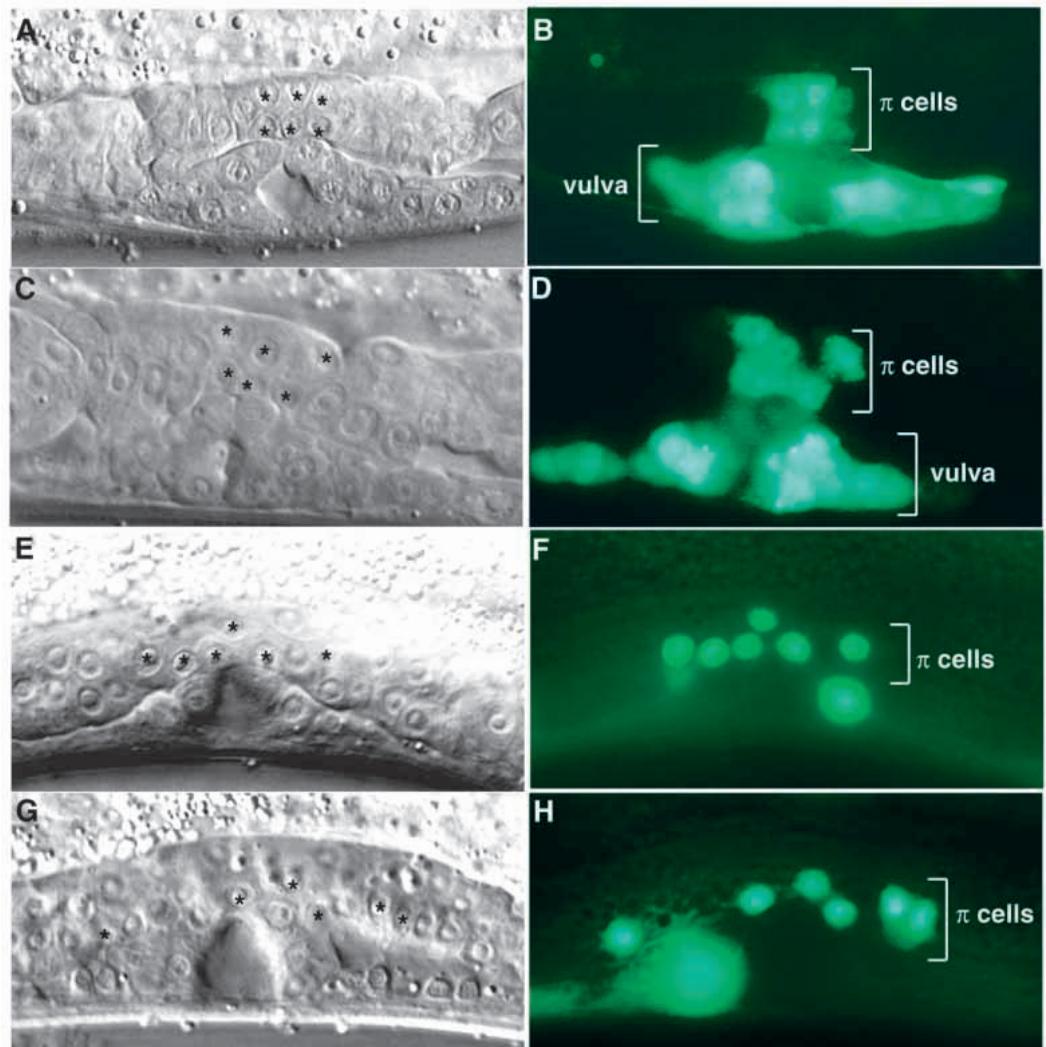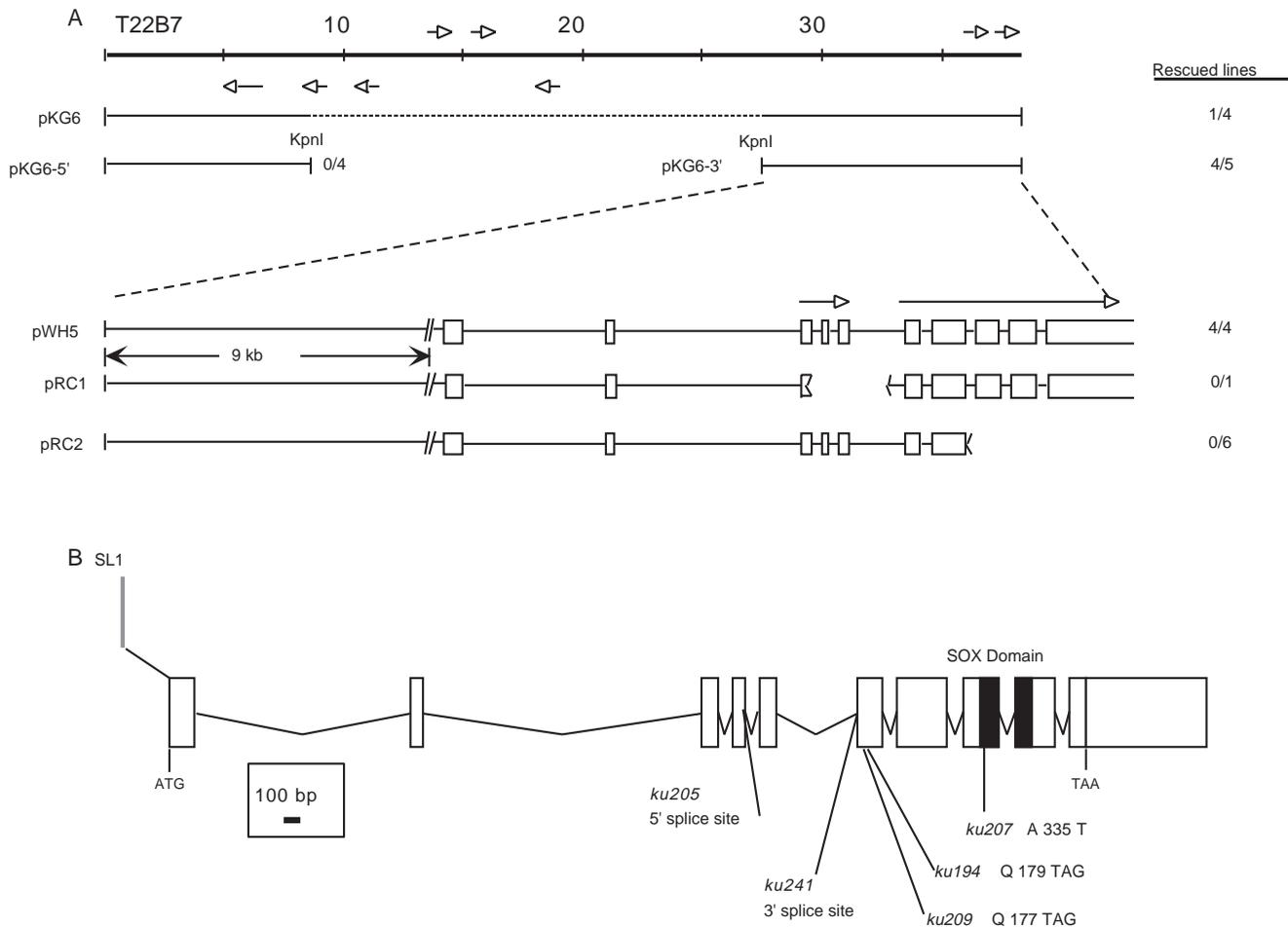
**Fig. 6**. Cloning of *cog-2*.. (A) Cosmid T22B7 and T22B7-derived subclones rescue *ku194*. The coordinates in kilobase pairs are derived from sequence as determined by the *C. elegans* Genome Sequencing Consortium. Arrows represent genes predicted by Genefinder. The name of each subclone is indicated on its left, and to the right, the number of stable lines that had rescuing activity is expressed as a fraction of the total number of lines scored. Deletions predicted to disrupt each of the two predicted genes at the 3′ end of T22B7 were constructed and tested for rescue as indicated. (B) Genomic structure of *cog-2*. Exons are depicted as boxes with intervening introns. The relative positions of the start and stop codons are shown. *cog-2* is transpliced to SL1. The position of the lesion in each of the five alleles is indicated. The black boxes represent the Sox domain.

## DISCUSSION

Mutations in *cog-2* were identified in a screen for vulval morphogenesis defective mutants. *cog-2(ku194)* is likely a null mutation that primarily affects the egg-laying system of *C. elegans* and, to a lesser extent gonad migration. We have shown that the Egl phenotype is a result of a failure of the vulval and uterine lumens to make a functional connection. The AC-utse cell fusion event, a prerequisite for formation of a functional connection, is defective in *cog-2* mutants. Instead, the AC remains at the dorsal opening of the vulva, blocking the passage of eggs from the uterus. We propose that a late-stage defect in utse cell determination, specifically affecting the ability of the utse cell and the AC to fuse, is the likely explanation for the *cog-2* phenotype. This hypothesis is consistent with our observations that a *cog-2* reporter is expressed in the π cells, which are utse cell precursors, but is not expressed in the AC. Furthermore, in *cog-2* mutants the AC and the π cells appear to adopt normal fates and yet fail to fuse. The AC adopts a proper fate as assayed by its ability to perform

its normal roles up until the time of fusion with the utse cell, including induction of the vulva cells and π cells, with the possible exception of penetration of vulF at a high frequency. The π cell fate is induced as assayed by lineage pattern and the wild-type expression pattern of two π cell-specific markers. Therefore, the defect in *cog-2* mutants could be a specific late-stage cellular defect within the utse cell such that it is unable to fuse with the AC. Alternatively, there may be a lack of utse-AC signaling such that fusion is never initiated.

Cell-to-cell fusion is an important component of multiple developmental events such as fertilization (Wilson and Snell, 1998), formation of the placental lining (Hernandez-Verdun, 1974), myogenesis (Wakelam, 1985), and formation of ostoeclasts (Roodman, 1991), and is a mechanism employed frequently during *C. elegans* development. More than 20% of the nuclei in an adult hermaphrodite are found within syncytial cells that can have upwards of 100 nuclei (Podbilewicz and White, 1994; Sulston and Horvitz, 1977; Sulston et al., 1983). Although the purpose of many of these fusion events in *C. elegans* remains mysterious, the AC-utse cell fusion appears to be required for

**A**

```
  1       gtttgagc ATG AGC CGT AGA CGA AAA GCG AAT CCG ACA AAA CTG AGT GAA AAC GCG AAG
                    M   S   R   R   R   K   A   N   P   T   K   L   S   E   N   A   K

 60       AAG CTT GCC AAG GAA GTT GAA AAT AGT CAA GAA GAA AAC GAC TGC GAC ATG CTG GCC AAA
 18        K   L   A   K   E   V   E   N   S   Q   E   E   N   D   C   D   M   L   A   K

120       CCG CAA ACC CCC ACG ATT ATC ATA CCT GGT CAT ATG GAC GAC ACG CCG AAT CCA GCC GGG
 38        P   Q   T   P   T   I   I   I   P   G   H   M   D   D   T   P   N   P   A   G

180       TCA CCT CAC GAC GCC TCA ATC AAA TCG AGC TCG AGC ACC ATC TCC GAT CAT ACG ACG ACG
 58        S   P   H   D   A   S   I   K   S   S   S   S   T   I   S   D   H   T   S   T

240       TCA GCA ACA ACC GGA ATT TCT GAT TTT CCC GAT ATT CTG GCA CAA ACT GAA CAC GGC TGT
 78        S   A   T   T   G   I   S   D   F   P   D   I   L   A   Q   T   E   H   G   C

300       TCT GTA CTA ATA GAC GGA AAT CAC TTA ACA GAG ATT ATC AAT TCA GTG GAT ACG CAA GAC
 98        S   V   L   I   D   G   N   H   L   R   E   I   I   N   S   V   D   T   Q   D

360       GGA AAA CAA GAT CTA TTA TCA GAC GTT ATC CGA CAG CTG ACA TCA ATA AAA GAG AGG TTA
118        G   K   Q   D   L   L   S   D   V   I   R   Q   L   T   S   I   K   E   R   L

420       ACA AAC GAC GAG AGT CCT GTC AAA GAC GAT TTA AAA GAA GAT CCT GAC GAC ATG AGC CCG
138        T   N   D   E   S   P   V   K   D   D   L   K   E   D   P   D   D   M   S   P

480       ATG CTC CAT GCC GGC AAT TTC GAC TCG GAG ATG CTC CTG AGG CAG CAC GAA CTG ATG CAG
158        M   L   H   A   G   N   F   D   S   E   M   L   L   R   Q   H   E   L   M   Q

540       CAT CAG CAG CAA CAA ATG ATA ATT GCT AAC ATG TTG AAA GCC ACC CAA TCT CTC CCA CTC
178        H   Q   Q   Q   Q   M   I   A   N   M   L   K   A   T   Q   S   L   P   L

600       CTT TTT AAT GGC GGT TTG AAT TAT GAA GCT ATA CTG AAC AAT CCC GTC TTA AAT GCT ACT
198        L   F   N   G   G   L   N   Y   E   A   I   L   N   N   P   V   L   N   A   T

660       ATT GCT GGA CAT TTG CCA AAT CCT CTT GCA TCG AAC ATT TCC TTG CTT CAA AAG AGT ATT
218        I   A   G   H   L   P   N   P   L   A   S   N   I   S   L   L   Q   K   S   I

720       AGC GCA AAA CTT GCA GCC GCT GGA AAC ATG CAA ACA GTC GAA AAA GTG GAG ACA CCT TTA
238        S   A   K   L   A   A   A   G   N   M   Q   T   V   E   K   V   E   T   P   L

780       AAT CTT TCA AAA GAC ACC CCA TCT CCA ACT GCT ATT CCA CAA TCT CCA CTA TCT GGA TTT
258        N   L   S   K   D   T   P   S   P   T   A   I   P   Q   S   P   L   S   G   F

840       CGG CTA CCA TAC TCA CTT GGG ACA AAT TAT GGA AGC GAT GGT CAA TTA TTT AAC AAT TGC
278        R   L   P   Y   S   L   G   T   N   Y   G   S   D   G   Q   L   F   N   N   C

900       TCA CCA AAT TCC AGT GGT AAG AGC ACG CCT GGT AAC ACA TCA GTA ACT AGT GAA GTA GCA
298        S   P   N   S   S   G   K   S   T   P   G   N   T   S   V   T   S   E   V   A

960       ACT CCA CGA CCA CAG GCC AAA TCA CCA AAC CAT ATC AAA AGA CCA ATG AAC GCA TTC ATG
318        T   P   R   P   Q   A   K   S   P   N   H   I   K   R   P   M   N   A   F   M

1020      GTA TGG GCT CGT GAC GAG AGA CGA AAA ATT TTG AAA GCG TAT CCG GAT ATG CAC AAC AGC
338        V   W   A   R   D   E   R   R   K   I   L   K   A   Y   P   D   M   H   N   S

1080      AAC ATC TCG AAG ATC CTC GGA TCT CGT TGG AAA GGT ATG AGC AAT TCC GAG AAA CAG CCA
358        N   I   S   K   I   L   G   S   R   W   K   G   M   S   N   S   E   K   Q   P

1140      TAC TAT GAA GAG CAA TCA CGG CTG AGC AAA TTG CAC ATG GAG CAG CAC CCG GAC TAT CGG
378        Y   Y   E   E   Q   S   R   L   S   K   L   H   M   E   Q   H   P   D   Y   R

1200      TAC AGA CCA CGT CCG AAG AGA ACA TGT GTG ATT GAT GGA AAG AAG GTT CGA GTC AAC GAA
398        Y   R   P   R   P   K   R   T   C   V   I   D   G   K   K   V   R   V   N   E

1260      TAT AAG ACA ATC ATG AAG ACT AAA AAA GAT TTG ATG TGG GGA GAT GAG CCA GGA TTT TCT
418        Y   K   T   I   M   K   T   K   K   D   L   M   W   G   D   E   P   G   F   S

1320      CAA CCA TCA GAT CTA CAA ATG GAC TTG GCA TCA CAT GTA AAT CTT CTC AAC GAC CTC ACA
438        Q   P   S   D   L   Q   M   D   L   A   S   H   V   N   L   L   N   D   L   T

1380      CAG CAC CAC CAC CAG TCA CAT CTC CTA CAA ACA GCT GAA TAActttaaccaagcgtgtcattccat
458        Q   H   H   H   Q   S   H   L   L   Q   T   A   E   Z
1446      ctgatcctccaaacattctctgtattttttaaaaccccccacctttggaattttgtcgtgttgttctcatgccagctttactttggt
1531      actcctcacccaccacatttacttacctggtccatcatgtattattacagtatctcaatttttattttttttagcagatcggttt
1616      cttgattttcaaaatcttggacaatcaaaatctgcttgccggtggacaatgcatttttttttccattcttcctttc
1701      ctcatacaatagcataatctgaagcgttatttcctttttctatgtgtttccttattgtacctctttcaaaagcttcccacactcca
1786      agaatctcagcgtttccgcattgtttcctacatattctgaagactgcatttcagtttgatttactgtgttcttcatatcaattcc
1871      tttcctattcataataggccttttttttccatttttttttcctcctttttactgcattgtgttcccattttccacctggcaccagatct
1956      tttcctttatcttgttgcaacctttaggctagtcatcaatcaaacatccactcccgtttccatcaaattttttactctatgcatgca
2041      tctgattatggttcacactcaattttctgataaatgtcatattcaagaacaagag
```

**Fig. 7.** *cog-2* encodes a member of the Sox domain family of proteins. (A) The sequence of the complete cDNA including the 3′ and 5′ UTRs and the coding region is shown on the top line with the predicted protein sequence below it. The corresponding nucleotide and amino acid coordinates are indicated to the left. The Sox domain is boxed. The putative polyadenylation signal is underlined. Ƴ, splice junctions. Residues and splice junctions affected by the five *cog-2* alleles are circled. (B) The COG-2 Sox domain is aligned with Sox domains of the murine Sox5 and Sox6 proteins as well as SRY. A Sox domain consensus is shown on the bottom line (Nambu and Nambu, 1996). Amino acid identities are indicated with a dash and differences are specified. Percentage identity with COG-2 is indicated to the right of each sequence. The conserved alanine affected by the *ku207* allele is underlined. The GenBank accession number for the *cog-2* cDNA is AF097319.

**B**

```
                                   ku207
                                     ↓
COG-2    RPQAKSP NHIKRPMNAF MVWARDERRK ILKAYPDMHN SNISKILGSR WKGMSNSEKQ PYYEEQSRLS KLHMEQHPDY RYRPRPKRTC VIDGKKVRVN EYKTIMKTKK DLM
Sox-5    -GRGSNE P--------- ----K----- --Q-F----- ---------- --A-T-L--- ------A--- -Q-L-KY--- K-K------- LV----L-IG ---A--RNRR QE-           71%
Sox-6    -GR-S-E P--------- ----K----- --Q-F----- ---------- --S---Q--- ------A--- -I-L-KY-N- K-K------- IV----L-IG ---QL-RSRR QE-           72%
SRY      DRV----- I--S--Q--- MALEN-R-R- -E---Q--YQ --MLTEA--W -FFQ-AQK-Q AM-R-KY-N- K----R-AKM                                            49%
SOX CONS.        RPMNAF VW   R          P   N    SK LG     W         K  P   E L    H    P Y KY                                          96%
```
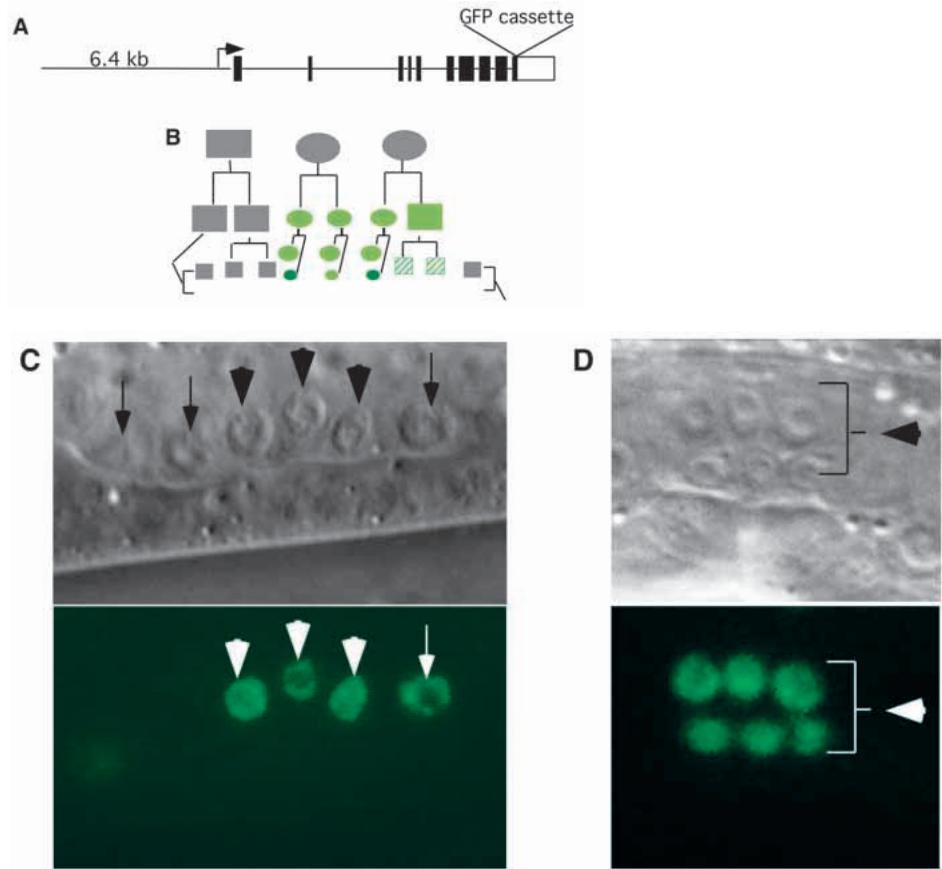
**Fig. 8**. COG-2 expression pattern.
(A) Schematic of the *cog-2*::gfp translational fusion. Filled boxes represent all ten *cog-2* exons and the open box represents the *cog-2* 3′ end. (B) Diagram depicting the dynamic expression pattern of *cog-2*::gfp as shown in C and D. Squares indicate cells of the ρ lineage and circles indicate cells of the π lineage. Cells with no expression are gray. Expressing cells are a solid light green. Cells with higher level expression at later stages are dark green, and the hatched squares indicate cells that usually do not express *cog-2*::gfp but have been seen to fluoresce slightly. Nomarski photomicrographs and corresponding fluorescence images of (C) L3 and (D) early L4 hermaphrodite carrying *cog-2*::gfp on an extrachromosomal array. Arrows, ρ precursors. Arrowheads, π precursors and π cells.

disposing of the AC, which appears to be simply in the way after it has carried out its previous functions. Since *cog-2* mutants have a rather specific defect affecting primarily uterine and vulval morphogenesis, it is most likely that the role of COG-2 in cell fusion may be to allow identification of a specific cell fusion partner rather than as a general factor required for multiple cell fusion events. However, the identification of an important regulator of a cell fusion event opens the door for further investigation into the mechanism of cell fusion. Perhaps identification of targets of the COG-2 transcription factor will allow elucidatation of cell fusion processes.

## COG-2 is a putative transcription factor

*cog-2* encodes a putative transcription factor containing a Sox domain. Sox family members contain an HMG box that is greater than 50% identical to the HMG box of SRY. Sox genes, which are present in the genomes of many vertebrates, have diverse tissue-specific expression patterns during early development and have been implicated in numerous developmental processes (Denny et al., 1992a; Laudet et al., 1993; Pevny and Lovell-Badge, 1997; Prior and Walter, 1996). Examples include the murine genes Sox2, Sox4 and Sox9. Sox2 is expressed embryonically in the nervous system as well as other tissues and acts in association with Oct-3 to regulate fibroblast growth factor 4 in embryonic carcinoma cell lines (Ambrosetti et al., 1997; Collignon et al., 1996; Yuan et al., 1995). Sox4 is expressed in the embryonic heart and neural tissue and in the thymus after birth (Wetering et al., 1993). Sox9, which plays roles in both sex determination and skeletal formation, is expressed in the developing gonad and in chondrocytes (Morais-da-Silva et al.,

1996; Wright et al., 1995). Additionally, a Sox gene from the *Drosophila* genome, *fish-hook/Dichaete,* is a segmentation gene that specifically regulates the *even-skipped* gene (Ma et al., 1998; Nambu and Nambu, 1996; Russell et al., 1996).

The Sox domain of COG-2 is most similar to that of the Sox5 and Sox6 proteins. Sox5 is a site-specific DNA-binding protein that binds to the minor groove of the double helix and severely bends DNA as do many other Sox proteins (Connor et al., 1994). Sox5 is likely an architectural type transcription factor. Architectural factors distort the DNA structure and may act to bring other factors into close proximity after binding distant sequences, to recruit components that bind specific structural elements or to prevent the binding of other factors. Based on the sequence similarity to Sox5 and the expression pattern of COG-2, we speculate that COG-2 is an architectural type transcription factor that targets a gene or genes required for terminal utse cell fate, perhaps acting to allow recognition of the utse cell fusion partner, the AC, or otherwise promote fusion.

*cog-2* mutants display a partially penetrant protruding vulva phenotype. We speculate that this phenotype may be the result of defective connections between uterine and vulval cells resulting in abnormal eversion of the vulva at the last stage of morphogenesis. While the basis for any effect of *cog-2* on brood size is unknown, *cog-2* may play a role in early embryogenesis since it is expressed in the embryo prior to elongation (data not shown). Expression of *cog-2* reporters in body wall muscles suggests a possible explanation for the gonad migration defects since the hermaphrodite gonad migrates along the body wall muscles (Kimble and Hirsh, 1979).

Expression of COG-2 in the nervous system of *C. elegans* is

intriguing in light of the large number of Sox genes expressed during mammalian CNS development (Pevny and Lovell-Badge, 1997). *cog-2* mutants have no obvious neural or behavioral defects. However, the *C. elegans* genome has multiple Sox genes, and COG-2 may act redundantly with any of these factors as has been postulated for Sox genes in mouse and human.

# REFERENCES

**Ambrosetti, D. C., Basilico, C. and Dailey, L.** (1997). Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. *Mol. Cell. Biol.* **17,** 6321-6329.

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

**Clark, S. G., Chisholm, A. D. and Horvitz, R. H.** (1993). Control of cell fates in the central body region of C. elegans by the homeobox gene lin-39. *Cell* **74,** 43-55.

**Collignon, J., Shanthini, S., Hacker, A., Cohen-Tannoudji, M., Norris, D., Rastan, S., Stevanovic, M., Goodfellow, P. N. and Lovell-Badge, R.** (1996). A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2. *Development* **122,** 509-520.

**Connor, F., Cary, P. D., Read, C. M., Preston, N. S., Driscol, l. P. C., Denny, P., Crane-Robinson, C., and Ashworth, A.** (1994). DNA binding and bending properties of the post-meiotically expressed Sry-related protein Sox-5. *Nucl. Acids Res.* **22,** 3339-3346.

**Connor, F., Wright, E., Denny, P., Koopman, P. and Ashworth, A.** (1995). The *SRY*-related HMG box-containing gene *Sox6* is expressed in the adult testis and developing nervous system of the mouse. *Nucl. Acids Res.* **23,** 3365-3372.

**Denny, P., Swift, S., Brand, N., Dabhade, N., Barton, P. and Ashworth, A.** (1992a). A conserved family of genes related to the tesis determining gene, SRY. *Nucl. Acids Res.* **20,** 2887.

**Denny, P., Swift, S., Connor, F. and Ashworth, A.** (1992b). An SRY-related gene expressed during spermatogenesis in the mouse encodes a sequence-specific DNA-binding protein. *EMBO J* **11,** 3705-3712.

**Freyd, G., Kim, S. K. and Horvitz, H. R.** (1990). Novel cysteine-rich motif and homeodomain in the product of the *Caenorhabditis elegans* cell lineage gene *lin-11*. *Nature* **344,** 876-879.

**Gu, T., Orita, S. and Han, M.** (1998). *Caenorhabditis elegans* SUR-5, a novel but conserved protein, negatively regulates LET-60 Ras activity during vulval induction. *Mol. Cell. Biol.* **18,** 4556-4664.

**Hernandez-Verdun, D.** (1974). Morphogenesis of the syncytium in the mouse placenta. Ultrastructural study. *Cell Tissue Res.* **148,** 381-96.

**Hill, R. J. and Sternberg, P. W.** (1992). The gene *lin-3* encodes an inductive signal for vulval development in *C. elegans*. *Nature* **358,** 470-476.

**Hobert, O., D'Alberti, T. and Ruvkun, G.** (1998). Control of neural development and function in a thermoregulatory network by the LIM homeobox gene *lin-11*. *J. Neurosci.* **18,** 2034-2096.

**Horvitz, H. R. and Sternberg, P. W.** (1991). Multiple intercellular signalling systems control the development of the C. elegans vulva. *Nature* **351,** 535-541.

**Kimble, J., and Hirsh, D.** (1979). Post-embryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*. *Dev. Biol.* **70,** 396-417.

**Laudet, V., Stehelin, D. and Clevers, H.** (1993). Ancestry and diversity of the HMG box superfamily. *Nucl. Acids Res.* **21,** 2493-2501.

**Ma, Y., Niemitz, E. L., Nambu, P. A., Shan, X., Sackerson, C., Fujioka, M., Goto, T. and Nambu, J. R.** (1998). Gene regulatory functions of drosophila fish-hook, a high mobility group domain sox protein. *Mech. Dev.* **73,** 169-182.

**Maduro, M. and Pilgrim, D.** (1995). Identification and cloning of *unc-119*, a gene expressed in the *Caenorhabditis elegans* nervous system. *Genetics* **141,** 977-988.

**Mello, C. and Fire, A.** (1995). DNA Transformation. *In Caenorhabditis elegans*: Modern Biological Analysis of an Organism. Vol. 48 (ed. H. F. Epstein and D. C. Shakes), pp. 452-482. Academic Press, San Diego.

**Mello, C. C., Kramer, J. M., Stinchcomb, D. and Ambros, V.** (1991). Efficient gene transfer in *C. elegans* after microinjection of DNA into germline cytoplasm: Recombination drives the assembly of heritable transgenic structures. *EMBO J.* **10,** 3959-3970.

**Morais-da-Silva, S., Hacker, A., Harley, V., Goodfellow, P., Swain, A. and Lovell-Badge, R.** (1996). Sox9 expression during gonadal development implies a conserved role for the gene in testis differentiation in mammals and birds. *Nature Genet.* **14,** 62-68.

**Nambu, P. A. and Nambu, J. R.** (1996). The *Drosophila fish-hook* gene encodes a HMG domain protein essential for segmentation and CNS development. *Development* **122,** 3467-3475.

**Newman, A. P. and Sternberg, P. W.** (1996). Coordinated morphogenesis of epithelia during development of the *Caenorhabditis elegans* uterine-vulval connection. *Proc. Natl. Acad. Sci. USA* **93,** 9329-9333.

**Newman, A. P., White, J. G. and Sternberg, P. W.** (1995). The *Caenorhabditis elegans lin-12* gene mediates induction of ventral uterine specialization by the anchor cell. *Development* **121,** 263-271.

**Newman, A. P., White, J. G. and Sternberg, P. W.** (1996). Morphogenesis of the *C. elegans* hermaphrodite uterus. *Development* **122,** 3617-3626.

**Pettitt, J., Wood, W. B. and Plasterk, R. H. A.** (1996). *cdh-3*, a gene encoding a member of the cadherin superfamily, functions in epithelial cell morphogenesis in *Caenorhabditis elegans*. *Development* **122,** 4149-4157.

**Pevny, L. H. and Lovell-Badge, R.** (1997). *Sox* genes find their feet. *Curr. Op. Gen. Dev.* **7,** 338-344.

**Podbilewicz, B. and White, J. G.** (1994). Cell fusions in the developing epithelia of C. elegans. *Dev. Biol.* **161,** 408-424.

**Prior, H. M. and Walter, M. A.** (1996). Sox genes: Architects of development. *Mol. Med.* **2,** 405-412.

**Riddle, D. L., Blumenthal, T., Meyer, B. J., and Priess, J. R.** (1997). *C. elegans II*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

**Roodman, G. D.** (1991). Osteoclast differentiation. *Crit. Rev. Oral. Biol. Med.* **2,** 389-409.

**Russell, S. R. H., Sanchez-Soriano, N., Wright, C. R. and Ashburner, M.** (1996). The Dichaete gene of Drosophila melanogaster encodes a SOX-domain protein required for embryonic segmentation. *Development* **122,** 3669-3676.

**Salser, S. J., Loer, C. M. and Kenyon, C.** (1993). Multiple HOM-C gene interactions specify cell fates in the nematode central nervous system. *Genes Dev.* **7,** 1714-1724.

**Seydoux, G., Savage, C. and Greenwald, I.** (1993). Isolation and characterization of mutations causing abnormal eversion of the vulva in *Caenorhabditis elegans*. *Dev. Biol.* **157,** 423-436.

**Sternberg, P. and Han, M.** (1998). Genetics of RAS signaling in *Caenorhabditis elegans*. *Trends Genet.* (in press).

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

**Sulston, J. E., Schierenberg, E., White, J. G. and Thomson, J. N.** (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100,** 64-119.

**Wakelam, M. J.** (1985). The fusion of myoblasts. *Biochem. J.* **228,** 1-12.

**Wetering, M. v. d., Oosterwegel, M., Norren, K. v. and Clevers, H.** (1993). Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes. *EMBO J.* **12,** 3847-3854.

**Wilson, N. F., and Snell, W. J.** (1998). Microvilli and cell-cell fusion during fertilization. *Trends Cell Biol.* **8,** 93-96.

**Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J. et al.** (1994). 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans. *Nature* **368,** 32-38.

**Wood, W. B.** (1988). *The nematode* Caenorhabditis elegans. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

**Wright, E., Hargrave, M. R., Christiansen, J., Cooper, L., Kun, J., Evans, T., Gangadharan, U., Greenfield, A. and Koopman, P.** (1995). The *Sry*-related gene Sox9 is expressed during chondrogenesis in mouse embryos. *Nature Genet.* **9,** 15-20.

**Yuan, H., Corbi, N., Basilico, C. and Dailey, L.** (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev.* **9,** 2635-2645.