The Evolutionary Duplication and Probable Demise of an Endodermal GATA Factor in *Caenorhabditis elegans*

Tetsunari Fukushige,[1] Barbara Goszczynski, Helen Tian and James D. McGhee[2]

Genes and Development Research Group, Department of Biochemistry and Molecular Biology,
University of Calgary, Alberta T2N 4N1, Canada

Manuscript received March 23, 2003
Accepted for publication June 4, 2003

ABSTRACT

We describe the *elt-4* gene from the nematode *Caenorhabditis elegans. elt-4* is predicted to encode a very small (72 residues, 8.1 kD) GATA-type zinc finger transcription factor. The *elt-4* gene is located ~5 kb upstream of the *C. elegans elt-2* gene, which also encodes a GATA-type transcription factor; the zinc finger DNA-binding domains are highly conserved (24/25 residues) between the two proteins. The *elt-2* gene is expressed only in the intestine and is essential for normal intestinal development. This article explores whether *elt-4* also has a role in intestinal development. Reporter fusions to the *elt-4* promoter or reporter insertions into the *elt-4* coding regions show that *elt-4* is indeed expressed in the intestine, beginning at the 1.5-fold stage of embryogenesis and continuing into adulthood. *elt-4* reporter fusions are also expressed in nine cells of the posterior pharynx. Ectopic expression of *elt-4* cDNA within the embryo does not cause detectable ectopic expression of biochemical markers of gut differentiation; furthermore, ectopic *elt-4* expression neither inhibits nor enhances the ectopic marker expression caused by ectopic *elt-2* expression. A deletion allele of *elt-4* was isolated but no obvious phenotype could be detected, either in the gut or elsewhere; brood sizes, hatching efficiencies, and growth rates were indistinguishable from wild type. We found no evidence that *elt-4* provided backup functions for *elt-2.* We used microarray analysis to search for genes that might be differentially expressed between L1 larvae of the *elt-4* deletion strain and wild-type worms. Paired hybridizations were repeated seven times, allowing us to conclude, with some confidence, that no candidate target transcript could be identified as significantly up- or downregulated by loss of *elt-4* function. *In vitro* binding experiments could not detect specific binding of ELT-4 protein to candidate binding sites (double-stranded oligonucleotides containing single or multiple WGATAR sequences); ELT-4 protein neither enhanced nor inhibited the strong sequence-specific binding of the ELT-2 protein. Whereas ELT-2 protein is a strong transcriptional activator in yeast, ELT-4 protein has no such activity under similar conditions, nor does it influence the transcriptional activity of coexpressed ELT-2 protein. Although an *elt-2* homolog was easily identified in the genomic sequence of the related nematode *C. briggsae*, no *elt-4* homolog could be identified. Analysis of the changes in silent third codon positions within the DNA-binding domains indicates that *elt-4* arose as a duplication of *elt-2,* some 25–55 MYA. Thus, *elt-4* has survived far longer than the average duplicated gene in *C. elegans*, even though no obvious biological function could be detected. *elt-4* provides an interesting example of a tandemly duplicated gene that may originally have been the same size as *elt-2* but has gradually been whittled down to its present size of little more than a zinc finger. Although *elt-4* must confer (or must have conferred) some selective advantage to *C. elegans*, we suggest that its ultimate evolutionary fate will be disappearance from the *C. elegans* genome.

DEVELOPMENT of the endoderm or intestine lineage in the nematode *Caenorhabditis elegans* depends crucially on a series of GATA-type transcription factors (for recent review, see Maduro and Rothman 2002). A current model of the regulatory hierarchy controlling gut development can be summarized as follows. The pair of small redundant GATA factors, MED-1 and MED-2, responds to the maternally provided factor SKN-1 and is involved in the distinction between the endoderm (intestinal or E lineage) and its mesodermal sister lineage MS (Maduro *et al.* 2001). The MED-1/MED-2 pair activates the genes encoding a second redundant pair of GATA factors, called END-1 and END-3; expression of *end-1* and *end-3* is endoderm specific but transient, beginning when the gut lineage has only a single cell (the 1E cell stage) and declining by the ~8E cell stage (Zhu *et al.* 1997, 1998). The END-1/END-3 pair in turn activates the GATA-factor *elt-2* gene, probably directly; *elt-2* expression begins midway through the 2E cell stage and continues throughout the life of the worm (Fukushige *et al.* 1998). ELT-2 activates, again probably directly, genes associated with terminal intestinal differentiation, such as the gut-specific carboxylesterase gene *ges-1* and the gene encoding the gut-specific intermedi-

[1] *Present address:* Laboratory of Molecular Biology, NIDDK, National Institutes of Health, Bethesda, MD 20892-0510.

[2] *Corresponding author:* Health Sciences Centre, Room 2205, University of Calgary Faculty of Medicine, 3330 Hospital Dr., N.W., Calgary, AB T2N 4N1, Canada. E-mail: jmcghee@ucalgary.ca

ate filament protein containing the epitope MH33 (Fukushige *et al.* 1998; T. Fukushige and J. D. McGhee, unpublished observations). ELT-2 also activates its own promoter (Fukushige *et al.* 1998, 1999). The absence of the *elt-2* gene causes lethality; *elt-2* null worms hatch but die with malformed intestines (Fukushige *et al.* 1998), suggesting that *elt-2* is necessary for expression of some particular gene or genes associated with the formation of a functioning intestine. Ectopic expression experiments demonstrate that ELT-2 is sufficient for expression of early gut markers, such as *ges-1* (Fukushige *et al.* 1998). However, these same markers are still expressed in the *elt-2* null mutants, indicating that at least one additional factor can activate these early gut genes in the absence of ELT-2 (Fukushige *et al.* 1998). One plausible candidate for an ELT-2 backup is ELT-7, a GATA factor that was identified from the genomic sequence and that is indeed expressed in the gut (K. Strohmaier and J. Rothman, personal communication). A second plausible candidate is the subject of this article: ELT-4 is a very small GATA factor encoded by a gene lying immediately upstream of the *elt-2* gene. Thus, this article addresses the following questions. What is the function of *elt-4* in the development of the *C. elegans* intestine? What is the evolutionary relation between *elt-4* and *elt-2*?

MATERIALS AND METHODS

Genetics and molecular biology: *C. elegans* was grown and maintained by standard methods (Brenner 1974). Unless otherwise noted, recombinant DNA manipulations also followed standard procedures (Sambrook and Russell 2001). The 5′-rapid amplification of cDNA ends (RACE) reaction to define the 5′-end of the *elt-4* transcript used the FirstChoice RLM-RACE kit from Ambion (Austin, TX) and the following two primers: ELT4R1 (5′-CTGCATGTTTCTTGTTTTTCTTC-3′) and ELT4R2 (5′-CAAGCCGTTTCCGATGAGAAGC-3′). To determine if the *elt-4* and *elt-2* coding sequences are present on the same transcript, reverse transcriptase (RT)-PCR was performed using the following two primers: RELT4F (5′-GTTAAGAATGGATAATAACTACTTAG-3′) corresponding to the *elt-4* 5′-untranslated region (UTR) and MH-13 (5′-GTAGG GTACACATGTTTG-3′) annealing to the *elt-2* 3′-UTR. Insertion of PCR-amplified green fluorescent protein (GFP) coding sequences from plasmid pPD95.67 (kindly provided by A. Fire, Carnegie Institute, Baltimore) immediately upstream of the *elt-4* termination codon (to produce either pJM156 or pJM188; see Figure 2A) was performed using the Stratagene (La Jolla, CA) QuikChange site-directed mutagenesis kit as suggested by Geiser *et al.* (2001); all coding regions in the final constructs were sequenced. Transgenic *C. elegans* was produced by standard methods (Mello and Fire 1996), using rescue of either *unc-119(ed4)* or *lin-15(n765ts)* as a transformation marker; reporter constructs were injected at concentrations of 50–100 μg/ml. The transforming array for one selected strain expressing pJM188 was integrated into the genome using γ-irradiation, as described previously (Egan *et al.* 1995); two independent stable lines (JM118 and JM119) were produced and both showed the same expression pattern.

RNA-mediated interference: RNA-mediated interference (RNAi) to the *elt-2* gene generally followed the procedures described by Fire and co-workers (Montgomery *et al.* 1998; Fire 1999). Synthesis of sense and antisense RNA was performed in separate reactions, using either T3 or T7 RNA polymerases (Promega, Madison, WI) and the appropriately cleaved *elt-2* cDNA plasmid (pJM68) as template. Transcripts were purified by phenol extraction, chloroform extraction, and ethanol precipitation and resuspended in diethylpyrocarbonate-treated 10 mm phosphate buffer, 1 mm EDTA (pH ~7.5); concentrations were determined spectrophotometrically. Equal amounts of the two strands were mixed in 1 m ammonium acetate, placed in boiling water for 2 min, allowed to cool slowly overnight, ethanol precipitated, and resuspended in diethylpyrocarbonate-treated 10 mm phosphate buffer, 1 mm EDTA (pH ~7.5) at a final concentration of ~1 mg/ml. Young hermaphrodite worms were injected once in a gonad and once in the gut/body cavity and then allowed to recover for 12–24 hr at ~22° before they were transferred to a fresh plate and observation of progeny was begun.

Miscellaneous methods: We have previously described the histochemical assay for endogenous GES-1 activity (Edgar and McGhee 1986) as well as the antibody staining protocols to detect ELT-2 protein and the MH33-reactive gut-specific intermediate filament (Fukushige *et al.* 1998). Electrophoretic mobility shift assays ("band shifts") were performed essentially as described previously (Kalb *et al.* 1998; Mains and McGhee 1999). The growth curves shown in Figure 4C were obtained as described previously (McGhee *et al.* 1990), except that nose-to-tail-tip lengths were measured using the ImageJ program, applied to converted image files obtained on a Zeiss Axiovision 2i microscope (×5 lens).

ELT-4 and ELT-2 proteins were expressed in *Saccharomyces cerevisiae* by cloning their respective cDNA sequences into the YCpGAL series of vectors (Bonner 1991; Shim *et al.* 1995; Kalb *et al.* 2002); constructs in which the cDNAs had been inserted in the antisense orientation were used as controls. The cotransformed reporter plasmid contained the tandem pair of GATA sites from the *C. elegans ges-1* gene (sequence provided in Table 1 below) inserted into the *Xho*I site of plasmid pLGΔ178. Yeast manipulations and the assay for β-galactosidase activity were performed as described previously (Kalb *et al.* 2002).

Ectopic expression of the *elt-4* gene in embryos: An *Eco*RV/*Sac*I fragment from the *elt-4* cDNA clone was inserted into *Sma*I/*Sac*I-cleaved vector pPD49.83 (kindly provided by A. Fire); the resulting construct (pJM402) has the *elt-4* coding sequence in the correct orientation downstream of the *C. elegans* heat-shock promoter (Stringham *et al.* 1992), exactly as had been done previously for the *elt-2* cDNA (Fukushige *et al.* 1998). Transformed strains were produced and the transforming array from one such line was integrated into the genome as described above. Embryos from this integrated transformed line (JM92) were isolated at the 1- to 4-cell stage, incubated at room temperature (~22°) for 75 min, heat-shocked at 34° for 30 min, and then incubated at 20° overnight before testing for marker expression was done. Controls included similar strains expressing *elt-2* cDNA under heat-shock control, as described previously (Fukushige *et al.* 1998).

Isolation of a chromosomal deletion in the *elt-4* gene: The library of ethyl methanesulfonate-mutagenized *C. elegans* strains described by Tsang *et al.* (2001) was screened with the following pairs of nested primers: outside pair, oJM60 (TGGGTG TTCCGATCTGAAACC) and oJM63 (GATTGCGTAGCATGA TCCAGC); inside pair, oJM61 (TGCGGTCTACTGGTTTTAC CTAGC) and oJM62 (ACATAGAACATTGCGACCAACG). A population producing a strong deletion band was subjected to four rounds of sib selection, at which point single worms could be demonstrated to be homozygous for a deletion completely removing the *elt-4* gene. This strain was then outcrossed

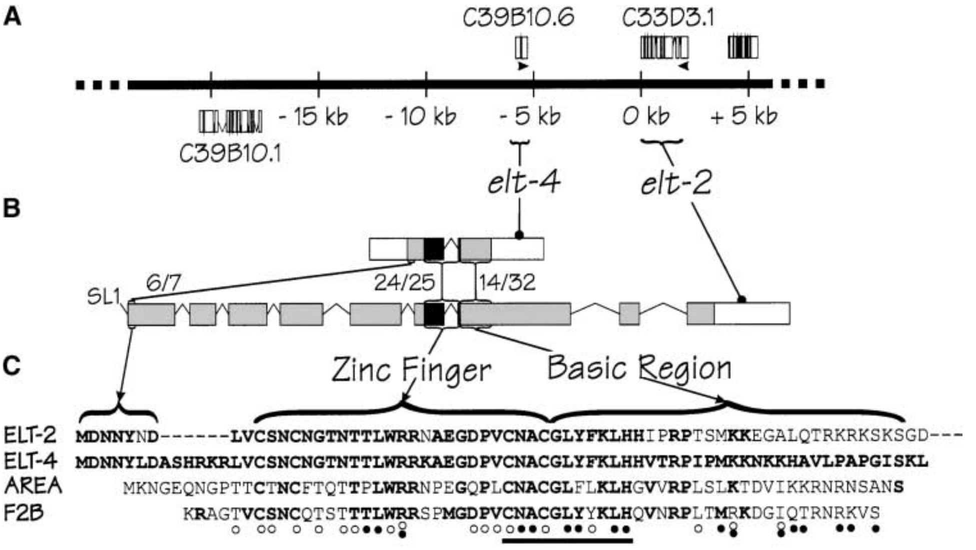
A
C39B10.6
C33D3.1
- 15 kb
- 10 kb
- 5 kb
0 kb
+ 5 kb
C39B10.1
elt-4
elt-2
B
SL1
6/7
24/25
14/32
C
Zinc Finger
Basic Region
ELT-2 MDNNYND------LVCSNCNGTNTTLWRRNAEGDPVCNACGLYFKLHHIPRPTSMKKEGALQTRKRKSKSGD---
ELT-4 MDNNYLDASHRKRLVCSNCNGTNTTLWRRKAEGDPVCNACGLYFKLHHVTRPIPMKKNKKHAVLPAPGISKL
AREA MKNGEQNGPTTCTNCFTQTTPLWRRNPEGQPLCNACGLFLKLHGVVRPLSLKTDVIKKRNRNSANS
F2B KRAGTVCSNCQTSTTTLWRRSPMGDPVCNACGLYYKLHQVNRPLTMRKDGIQTRNRKVS

FIGURE 1.—*elt-4* is a separate gene, lying upstream of *elt-2* and encoding a very small GATA factor. (A) Overall view of the *elt-4*/*elt-2* chromosomal locus (cosmids C39B10 and C33D3). The Genefinder program initially predicted that the ELT-2 protein contains two zinc finger DNA-binding domains. Arrowheads indicate the position of PCR primers used to test if sequences encoding the two zinc finger domains exist on a single transcript. Scale is in kilobases, centered on the *elt-2* initiation codon. (B) Expansion and alignment of the *elt-4* and *elt-2* regions of the *C. elegans* genome, showing the overall gene structure and the position of conserved amino acids. Solid rectangles correspond to the zinc finger DNA-binding domains; shaded rectangles depict coding sequences and open rectangles depict either 5′- or 3′-untranslated regions. (C) Sequence alignment of ELT-2 with the entire ELT-4 protein. Also included in the alignment are the sequences of a peptide from the Aspergillus GATA factor AREA (STARICH *et al.* 1998a) and F2B, a peptide from chicken GATA-1 (OMICHINSKI *et al.* 1993a,b). Following the designation of OMICHINSKI *et al.* (1993a), open circles represent residues involved in maintaining the three-dimensional structure of the DNA-binding domain, solid circles represent residues involved in DNA contact, and the underlined region represents the highly conserved α-helix that inserts into the DNA major groove.

six times to wild-type worms to produce the final deletion strain JM116 *elt-4(ca16)*, which was used in all experiments with the following exception. When we tried to perform *elt-2* RNAi on JM116, we realized that our laboratory "wild-type" strain (to which the *elt-4* deletion strain had been repeatedly outcrossed) had apparently picked up a mutation conferring RNAi resistance. We thus crossed JM116 to an independently obtained (RNAi sensitive) wild-type strain and verified that RNAi sensitivity had indeed been introduced back into JM116 (strain now designated JM124).

Microarray analysis: To produce L1 larvae from the *elt-4 (ca16)* null strain (JM116) and from wild-type (N2) controls, parallel cultures were grown at 20° on enriched growth medium (standard NGM plates containing a 10-fold higher concentration of peptone) and gravid adult worms were isolated using a 40-μm nylon mesh. Embryos were released by alkaline-hypochlorite treatment (WOOD 1988) and incubated overnight in M9 buffer without added food. The hatched L1 larvae were harvested, washed with water, and frozen at −70°. Total RNA was extracted using Trizol (Invitrogen, San Diego) and poly(A)$^+$ RNA was isolated using an mRNA isolation kit from QIAGEN (Valencia, CA). Seven paired poly(A)$^+$ RNA pools were sent to Stuart Kim (Department of Genetics, Stanford University) for microarray analysis (KIM *et al.* 2001).

RESULTS

***elt-4* encodes a very small GATA factor:** The GeneFinder program of AceDB (STEIN *et al.* 2001) initially predicted that the *elt-2* gene encodes two zinc finger GATA-factor-type DNA-binding domains. In contrast, our previous analysis (HAWKINS and MCGHEE 1995) indicated that *elt-2* encodes a significantly smaller protein with only one zinc finger, corresponding essentially to the 3′-region of the Genefinder prediction. Northern analysis supports the shorter size of *elt-2* determined from the cDNA (HAWKINS and MCGHEE 1995) but it would be difficult to rule out infrequent transcripts corresponding to a two-finger variant. We used RT-PCR with mixed stage cDNA as template to search for such a longer transcript but were unsuccessful (data not shown). We thus suspected that the upstream zinc finger sequence might encode a separate protein, hereafter referred to as ELT-4. Indeed, a cDNA clone corresponding to a separate upstream gene was subsequently identified by Y. Kohara (National Institute of Genetics, Mishima, Japan) and the present view of the *elt-4*/*elt-2* genomic locus is shown in Figure 1A. We used 5′-RACE to determine that the *elt-4* transcript begins 117 bp upstream of the *elt-4* ATG codon (data not shown). RT-PCR produced no evidence that *elt-4* was *trans*-spliced to the SL1 leader (KRAUSE and HIRSH 1987; BLUMENTHAL *et al.* 2002). The single intron in *elt-4* occurs precisely at the same point in the zinc finger domain as does a corresponding intron in the *elt-2* gene. The ELT-4/ELT-2 alignments in Figure 1, B and C, show that three blocks of sequence have been conserved: 6/7 amino acid residues at the N terminus, 24/25 residues in the zinc finger and 14/32 residues in the basic C-terminal region following the zinc finger. Overall, ELT-4 is predicted to contain only 72 amino acids (predicted molecular weight, 8130 D), which would make it by far the smallest GATA factor so far reported in any organism (see, for example, LOWRY and ATCHLEY 2000 and references therein).

The small size of the ELT-4 peptide does not necessarily preclude sequence-specific binding to DNA. The alignments of Figure 1C include the sequences of a 66-amino-acid fragment from the fungal GATA factor AREA (STARICH *et al.* 1998a) and a 59-amino-acid peptide (called F2B) encompassing the C-terminal zinc finger

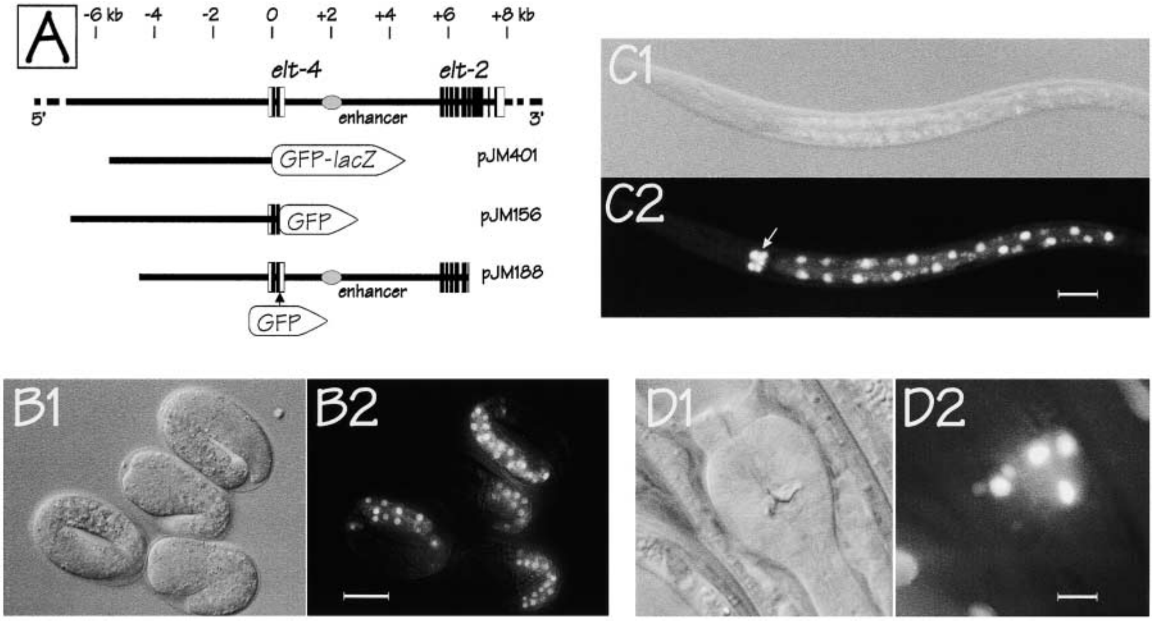

Figure 2.—*elt-4* is expressed in all cells of the intestine plus nine cells in the posterior pharynx.(A) Schematic representation of the three reporter gene fusions used to determine the *elt-4* expression pattern. At the top the *elt-4/elt-2* locus is shown, with scale centered on the initiation codon of *elt-4*. The three constructs are: pJM401, in which 5.5 kb of the *elt-4* 5′-flanking region are fused to GFP immediately after the *elt-4* ATG codon; pJM156, in which 6.8 kb of the *elt-4* 5′-flanking region plus the *elt-4* coding region are fused to GFP immediately before the *elt-4* termination codon; and pJM188, a construct whose 11.2-kb insert contains 4.5 kb of the *elt-4* 5′-flanking region, the entire *elt-4* coding region, the region between *elt-4* and *elt-2* (which includes the *elt-2* enhancer), and approximately half of the *elt-2* coding region (but not including the *elt-2* DNA-binding domain), into which a GFP coding sequence has been inserted immediately before the *elt-4* termination codon. (B) GFP fluorescence observed in embryos transformed with pJM188. B1, differential interference contrast (DIC) optics; B2, GFP fluorescence. Two embryos are at the ∼1.5-fold stage and two embryos are at the ∼3-fold stage. GFP expression is detected in all gut nuclei. Fluorescent images represent a maximum point projection of an aligned stack of nine deconvolved images taken at focal planes spaced at 1-μm intervals. Bar, 20 μm. (C) GFP fluorescence observed in an L1 larva transformed with pJM188. C1, DIC; C2, GFP fluorescence. GFP expression is detected in all nuclei of the intestine plus nine nuclei in the posterior bulb of the pharynx (arrow). Bar, 20 μm. (D) Pharynx of an adult worm transformed with pJM188. D1, DIC; D2, GFP fluorescence. Fluorescent images represent a maximum point projection of an aligned stack of 15 deconvolved images taken at focal planes spaced at 1-μm intervals. A full through-focus series reveals nine expressing nuclei (see text). Bar, 10 μm.

domain of chicken GATA-1 (Omichinski *et al.* 1993a,b). Both of these peptides have been shown to bind sequence specifically to DNA and, in fact, three-dimensional NMR structures have been determined for both peptides, complexed to their cognate binding sites. Beneath the F2B sequence on Figure 1C are indicated residues involved in maintaining the structure of the DNA-binding domain (open circles) and residues involved in DNA contact (solid circles; Omichinski *et al.* 1993a). The majority of both types of residues are conserved in ELT-4; in particular, the α-helix involved in major DNA contacts (underlined in Figure 1C) is highly conserved. Only in the C-terminal half of the basic region, which contains residues that contact the minor groove of the binding site, are residues less conserved. However, in spite of these conserved features, ELT-4 must be close to the minimum size required for sequence-specific binding to DNA: a peptide lacking six residues from the C terminus of F2B does not bind DNA (Omichinski *et al.* 1993b) and the arginine residue six positions from the C terminus of the AREA peptide is the last residue to contact DNA and the last residue required for AREA activity (Starich *et al.* 1998a).

***elt-4* is expressed in the intestine:** To determine where and when the *elt-4* gene is expressed, as well as to determine if regulatory signals that control *elt-2* also influence the expression of *elt-4*, we constructed three different *elt-4*::reporter gene fusions as diagrammed in Figure 2A. The expression patterns determined for the three different transforming reporter constructs are highly similar and within the variation normally seen with multiple independently transformed strains expressing the same construct. Thus, Figure 2 shows only images obtained with the longest construct, pJM188, which contains the entire *elt-4* locus with GFP inserted in frame at the *elt-4* C terminus.

The large majority of GFP signal, at all stages of development, is in the intestine. As shown in Figure 2B, the first GFP signal can be detected at the ∼1.5-fold stage of embryogenesis; by the 3-fold stage, GFP expression is easily detected in all cells of the gut. Late in embryogenesis, GFP expression can be detected in nine

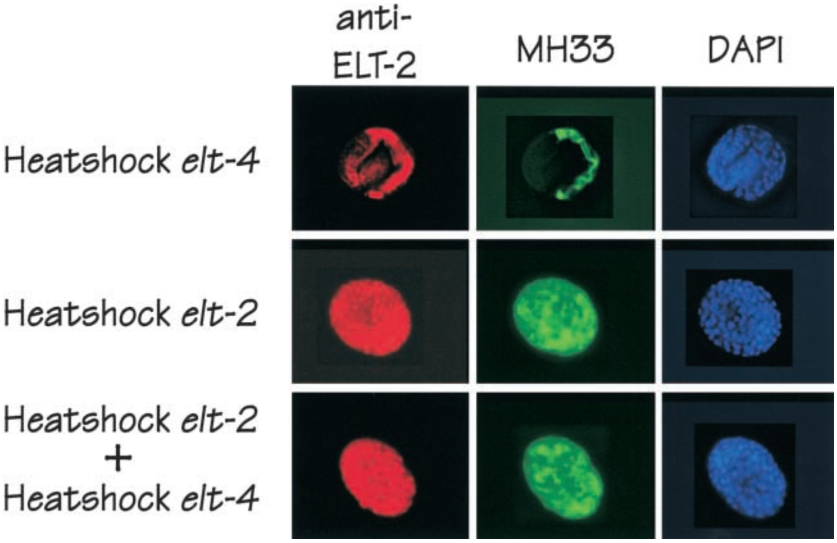

Figure 3.—Ectopic expression of *elt-4* in the early embryo does not cause ectopic expression of early gut markers. Embryos transformed with a heat-shock promoter::*elt-4* cDNA construct (top row) or a heat-shock promoter::*elt-2* cDNA construct (middle row) or transformed independently with both constructs (bottom row) were heat-shocked as described in materials and methods and then stained with an affinity-purified anti-ELT-2 antibody (left column), the monoclonal antibody MH33 (middle column), or the DNA-specific dye 4′,6-diamidino-2-phenylindole (right column). (In the heat-shock promoter::*elt-4* embryo stained with the anti-ELT-2 antibody, fluorescent intensity outside of the gut is nonnuclear and "nonspecific.")

nuclei in the posterior bulb of the pharynx, bracketing the pharyngeal grinder; the relative intensity of the intestinal and pharyngeal expression is shown for an L1 larva in Figure 2C. Both gut and pharynx expression continue throughout the remaining stages of development. A higher magnification view of GFP expression in an adult pharynx is shown in Figure 2D; on the basis of nuclear position, the nine expressing cells are the two triads of m6 and m7 muscle cells, as well as the immediately posterior triad of marginal cells (Albertson and Thomson 1976).

From the expression patterns directed by the three different constructs diagrammed in Figure 2A, we can conclude that: (i) the 5.5-kb fragment 5′ to the *elt-4* gene is sufficient to direct embryonic and larval gut (and pharynx) expression and (ii) the *elt-2* promoter, which is present in pJM188 but lacking in pJM156 and pJM401, does not appear to have a major influence on *elt-4* expression. The ELT-2 protein does however appear to be the major activator of *elt-4* as shown by the following experiment. Double-stranded RNA corresponding to the *elt-2* cDNA was injected into a strain (JM117) carrying an integrated transgenic array containing the construct pJM401 (see Figure 2A); the majority (>75%) of reporter gene expression was abolished (data not shown). Thus, the 5′-flanking region of *elt-4* is currently our best candidate for a promoter for which *elt-2* is necessary.

We attempted to verify the reporter gene expression patterns by producing ELT-4 specific antibodies. However, the similarity between ELT-2 and ELT-4 sequences provides only a limited number of peptides that could be used as distinctive antigens and our attempts to produce histochemically useful antibodies using the most promising of these peptides were unsuccessful.

Ectopic *elt-4* does not activate ectopic expression of gut markers in the early *C. elegans* embryo: We previously demonstrated that expression of a number of early gut markers (*ges-1*, gut granules, the gut-specific MH33-reactive intermediate filament, and the *elt-2* gene itself) can be driven ectopically by forced ectopic expression of *elt-2* (Fukushige *et al.* 1998). To determine whether *elt-4* had similar abilities, we produced an integrated transgenic strain expressing *elt-4* cDNA under control of the *C. elegans* heat-shock promoter and tested a range of induction conditions in an attempt to optimize expression. Ectopic *elt-4* does indeed cause arrest of embryonic development, usually after the beginning of morphogenesis; this is significantly later than the stage of arrest caused by ectopic expression of *elt-2* (Fukushige *et al.* 1998). However, under no conditions could we detect significant ectopic expression of early gut markers (Figure 3). We performed similar heat-shock experiments on a strain that also contained the heat shock::*elt-2* construct. We observed that ectopic expression of *elt-2* causes approximately the same level of ectopic marker expression in the presence or absence of ectopic *elt-4* (compare the middle and bottom rows of Figure 3). We conclude that ELT-4 neither greatly inhibits nor greatly augments the *in vivo* action of ELT-2.

Production and characterization of a null mutation in the *elt-4* gene: To determine whether the absence of *elt-4* in a worm produces an observable phenotype, we screened a library of deletion strains (Tsang *et al.* 2001), using PCR to detect a population in which the *elt-4* gene had been entirely deleted but the adjacent *elt-2* enhancer was left intact. One homozygous strain was isolated and backcrossed six times to wild-type worms [final strain designated as JM116 *elt-4(ca16)*]; the details of the deletion are given in Figure 4A; a Southern blot confirming that the strain is homozygous for the deletion is shown in Figure 4B. The *elt-4* knockout strain JM116 has no obvious phenotype, either in the gut or elsewhere. As shown in Figure 4C, the growth rate of JM116 is essentially indistinguishable from that of wild-type worms grown in parallel. Brood sizes, measured

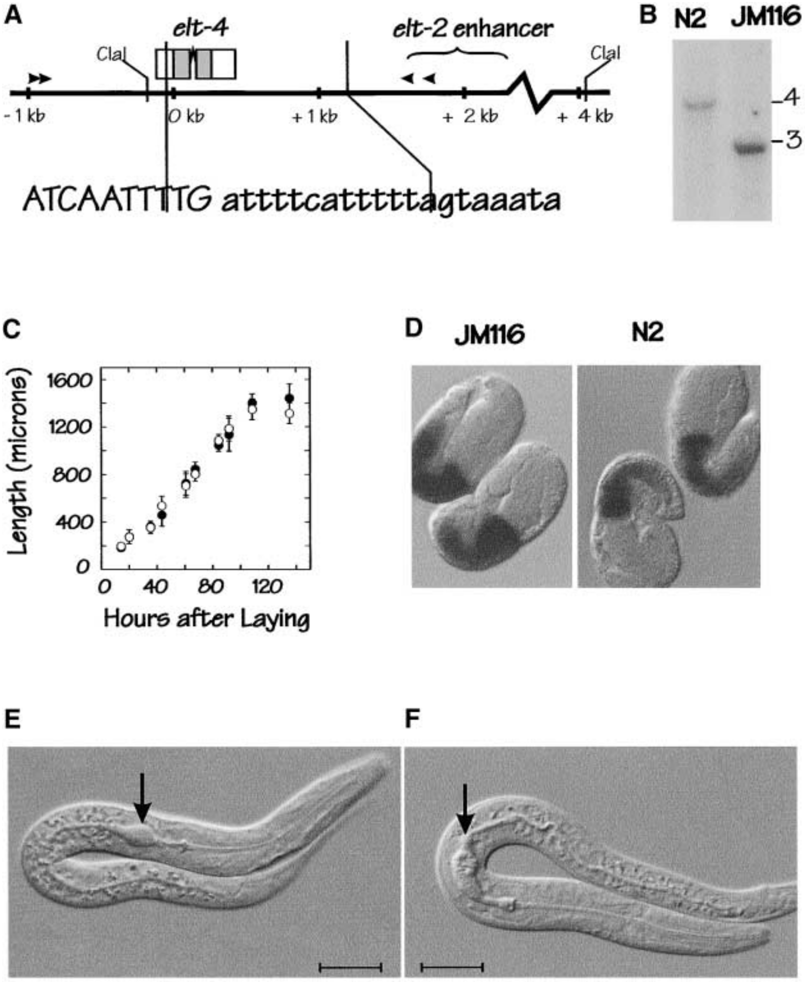

FIGURE 4.—Production of a deletion allele (*ca16*) of the *elt-4* gene. (A) Schematic diagram of the *elt-4* locus (coding sequences in black), showing the two nested pairs of PCR primers used to detect the deletion (arrowheads). The DNA sequence across the deleted region is shown below; the deletion removes the sequence from 42 bp upstream of the *elt-4* ATG down to 1196 bp downstream of the *elt-4* ATG. The sequence shown in lowercase is inserted at the junction. The position of the major *elt-2* enhancer is shown downstream of *elt-4*. *Cla*I sites are indicated. (B) Southern blot confirming complete deletion of the *elt-4* gene in strain JM116. Genomic DNA was produced from either wild-type worms (N2) or the *elt-4* deletion strain (JM116) and digested with *Cla*I prior to Southern blotting. Size standards are shown to the right. (C) Growth curve (length in micrometers plotted as a function of hours after laying) for wild-type worms (N2; open circles) and *elt-4* null worms (JM116; solid circles) at 19.9 ± 0.2°. Error bars represent sample standard deviations estimated from measuring 9–11 animals for each data point. (D) Embryos (~1.5-fold stage) stained for endogenous *ges-1* gene activity. Left, *elt-4* null embryos (JM116); right, wild-type embryos (N2). (E) Arrested larva produced by wild-type hermaphrodite injected with double-stranded *elt-2* RNA. Arrow points to the obstruction at the gut anterior. Bar, 20 μm. (F) Arrested larva produced by *elt-4* deletion hermaphrodite injected with double-stranded *elt-2* RNA. Arrow points to the obstruction at the gut anterior. Bar, 20 μm.

under a variety of conditions and temperatures, are also essentially normal: (brood size of JM116)/(brood size of N2) = 1.01 ± 0.16 (SD). Hatching efficiency is >99% (data not shown). Biochemical markers of early intestinal development (*ges-1*, gut granules, the MH33-reactive intermediate filament, and *elt-2*) are expressed in the mutant at apparently normal levels; the particular example of *ges-1* is shown in Figure 4D.

To determine if *elt-4* and *elt-2* are redundant, we performed RNAi to *elt-2* in the *elt-4* deletion strain (see MATERIALS AND METHODS). Injection of double-stranded *elt-2* RNA into wild-type worms produces arrested larvae that develop an obstructed gut phenotype (Figure 4E), as described previously for the *elt-2* knockout (FUKUSHIGE *et al.* 1998). Injection of the same *elt-2* RNA into the *elt-4* deletion strain produces arrested larvae with phenotypes essentially indistinguishable from those produced in the control strain (Figure 4F). Both wild-type and *elt-4* deletion worms, when injected with double-stranded *elt*-2 RNA, produce embryos that stain for GES-1 activity (data not shown). In other words, there is no evidence that loss of *elt-4* exacerbates the *elt-2* null phenotype.

The fact that ELT-4, a GATA factor, is expressed in at least some cells of the pharynx raises the possibility that ELT-4 could be involved in the GATA site-dependent switch of *ges-1* expression from the gut into the pharynx (AAMODT *et al.* 1991; KENNEDY *et al.* 1993; EGAN *et al.* 1995; FUKUSHIGE *et al.* 1996; MARSHALL and MCGHEE 2001). However, when the full *ges-1* construct pJM15 was introduced into the *elt-4(ca16)* background and transgenic embryos were assayed for *ges-1* activity, no unusually high level of staining in the pharynx was observed, compared to the pJM15-transformed wild-type controls (data not shown).

Use of microarrays to search for *elt-4*-regulated genes: The fact that *elt-4(ca16)* worms are essentially indistinguishable from wild type provides a situation in which DNA microarrays can be employed to search more exhaustively for differences in gene expression, without complications arising from mutation-derived differences in population structure. Embryos were prepared from parallel cultures of JM116 and N2 worms and allowed to hatch in the absence of food. Poly(A)$^+$ RNA was extracted from these matched samples of L1 larvae and used as template to produce either Cy3- or Cy5-labeled cDNA. Hybridization to a microarray containing essentially all of the coding sequences from the *C. elegans* genome was carried out by Stuart Kim through the Stanford Microarray Facility (Kim *et al.* 2001). Analyses were repeated on RNA samples isolated from seven independent pairs of L1 populations; in four of these pairs, Cy3 was used to label the JM116 cDNA and in the other three pairs, the dye assignment was reversed. We first removed data for all genes that did not show a minimum spot intensity in each of the 14 RNA preparations; we set the minimum spot intensity as 1000 (arbitrary units); this level is ~1% of the maximum spot intensity seen for any gene on the array and is two- to five-fold above background, depending on the hybridization experiment; 1871 different genes survived this first test of reproducibility.

We analyzed the data in two ways. The first approach was a straightforward scheme based on the ratios of the hybridization intensities in the two channels, with the aim of quickly assessing whether the two RNA populations differed significantly. The second approach was a more discriminating analysis based on intensity differences between the two channels [significance analysis of microarrays (SAM; Tusher *et al.* 2001)] and will be discussed below. In the first approach, all the data passing the preliminary reproducibility criterion were arrayed in a table as diagrammed in Figure 5A; each row contains the set of seven replicate measurements of Ln[(intensity of JM116 channel)/(intensity of N2 channel)] for one particular gene spot; each column lists the Ln(ratio) of all spots measured in one single hybridization experiment using one of the seven pairs of matched RNA samples. [Ln(ratios), rather than ratios, are used because replicated measurements are more likely to be normally distributed (Nadon and Shoemaker 2002) and errors are more likely to be independent of the magnitude of the intensity ratios, which we verified from our data.] Our analysis is based on the recognition that variation within each row reflects the experimental precision with which the Ln(ratio) can be measured for any particular gene but that variation within each column reflects both the precision of the experimental measurements and any real changes in gene expression between wild-type and the *elt-4* null larvae. For each row, the Ln(ratio) for a particular gene spot, averaged over the seven replicate hybridizations, was calculated, as was the sample (unbiased) standard deviation. Under the null hypothesis that there is no significant change in gene expression between JM116 and N2 L1 larvae, the frequency distribution of the observed average Ln(ratios) should be accurately predicted by a distribution whose center is determined by the overall average Ln(ratio) for all gene spots and all replicates but whose width is determined solely by the experimental variation inherent in measuring Ln(ratio) for individual spots. Figure 5B (bars) shows the frequency distribution of the Ln(ratio) for each gene spot averaged over the seven replicates. The continuous line is the predicted normal distribution centered on the overall average Ln(ratio) of −0.1 (averaged over all the 1871 gene spots; this small deviation from zero reflects good but not perfect normalization of the intensities between the two channels, over the seven replicates); the standard deviation used to calculate the width of the predicted distribution shown in Figure 5B is computed as (standard deviation of the sample of seven replicates, averaged over all gene spots)/$\sqrt{7}$; the factor $1/\sqrt{7}$ is introduced to convert the standard deviation of the sample to the standard deviation of the mean of the sample (Snedcor and Cochran 1980). As can be seen from Figure 5B, there is excellent overall agreement between the observed frequency distribution of the Ln(ratio) and the distribution calculated on the assumption that there is no difference in gene expression between L1 larvae of *elt-4(ca16)* and wild-type worms. To reiterate, the observed spread of measured Ln(ratio) is due essentially entirely to experimental variability in measuring spot intensities and no genes appear to be dramatically up- or downregulated in the absence of *elt-4*. Only one gene was identified as being upregulated by >2-fold and this is *hsp-70* (2.15-fold increase). Seven genes were identified as being downregulated by >2-fold (average decrease, 2.3-fold; maximum decrease, 2.8-fold) and six of these are cuticular collagens; the remaining putatively downregulated gene is fructose bis-phosphate aldolase. We interpret these results to mean that the genes identified as being up- or downregulated are not likely to be gut genes that could be specific targets of *elt-4*. Rather, the identified genes appear to be highly expressed, with the additional complication in the case of the collagens that they belong to a multicopy family (well over 100 genes).

The second way in which we analyzed our data was to use the SAM method, which emphasizes the use of differences, not ratios, in spot intensities (Tusher *et al.* 2001). The data from the same 1871 genes used in the previous analysis were entered as seven sets of paired intensities. To normalize signal intensities between the different hybridization experiments in such a way as to avoid dominance by highly expressed genes, intensities for single hybridization experiments were normalized using the slope of a graph plotting the cube root of spot intensity for a single hybridization experiment *vs.* the cube root of spot intensity averaged over all seven experiments (Tusher *et al.* 2001). The output of the SAM program is a plot relating the intensity differences observed for each gene spot between JM116 and N2 RNA (averaged over the seven replicates and normalized to experimental variability) *vs.* the differences predicted if the data sets are permuted. An adjustable parameter, Δ, is used as a criterion to judge whether gene expression is "significantly" different between the two RNA populations. The smaller the value of Δ, the greater the number of genes are judged to be significantly different; at the same time, however, the program returns a

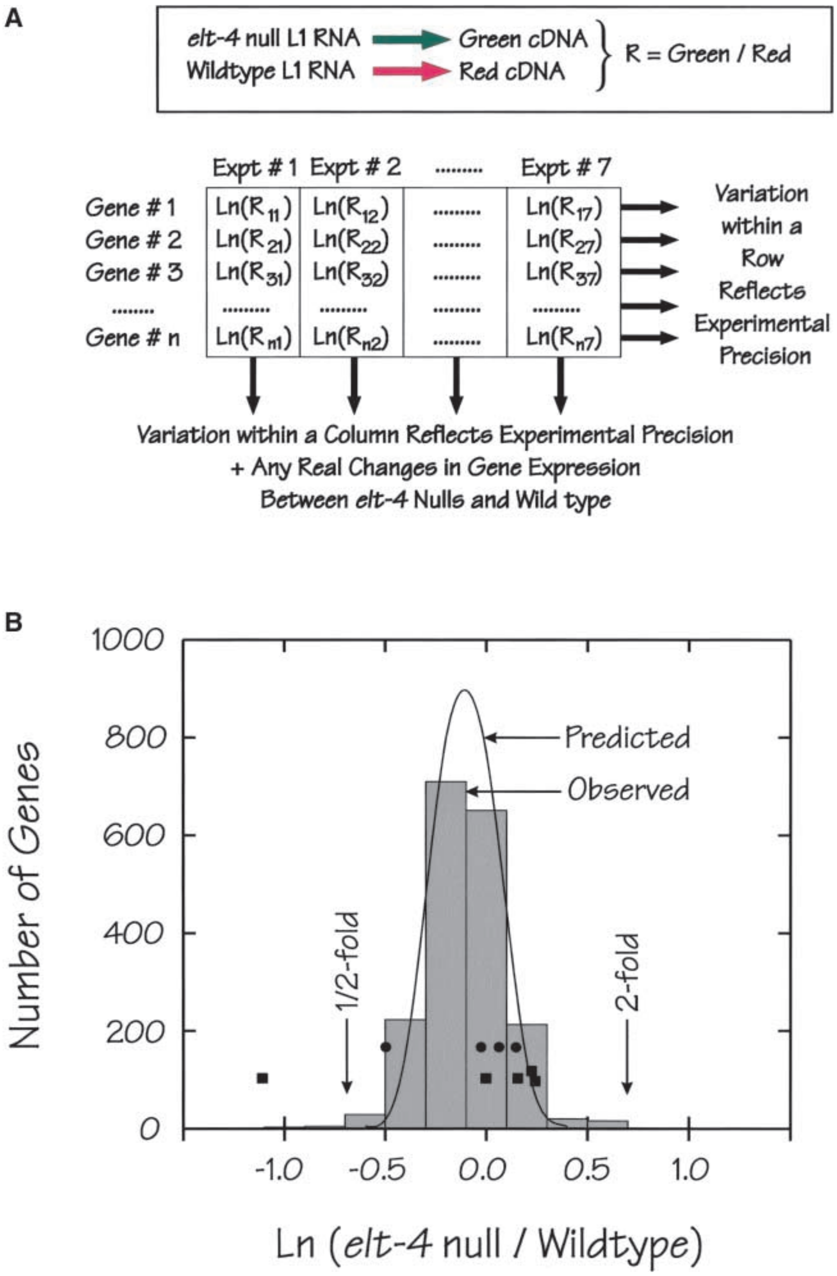

FIGURE 5.—Use of microarrays to assess the differences in transcript levels between *elt-4 (ca16)* L1 larvae and wild-type L1 larvae. (A) Schematic outline of the analysis procedure. Further details are provided in the text. (B) Frequency distribution of Ln[(Hybridization intensity of a particular gene spot using *elt-4* null cDNA)/(Hybridization intensity of the same gene spot using wild-type cDNA)]. Observed gene numbers are depicted as shaded bars. The solid line is the normal frequency distribution predicted on the basis that the only variation in intensity ratio is due to experimental error. Further details are provided in the text. Solid circles represent replicate measurements for the *elt-2* gene; solid squares represent replicate measurements for the *ges-1* gene.

greater number of false positives. To be sensitive to any differences in gene expression between the two RNA populations, we use a value of $\Delta = 0.4$, the smallest value used in the original publication to differentiate between two cell populations; under these conditions, roughly half of the identified genes were estimated to be false positives (TUSHER *et al.* 2001). We further specify the modest criterion that a gene must be either upregulated or downregulated by 20%. Even with this nonstringent choice of parameters, only seven genes are judged to be expressed differentially between the two RNA populations (as indicated by the arrows in Figure 6). One gene is judged to be upregulated and this is an acyl-carrier protein expressed in mitochondria. Of the six genes judged to be downregulated, one is ubiquitin and a second is a small novel open reading frame of which little is known. Three of the "downregulated" genes are cuticular collagen genes and the last is fructose-bis-phosphate aldolase; these last four genes were also identified by the previous analysis. Aldolase is central to glycolysis and is expected to be present in all cell types. The same widespread distribution undoubtedly holds for ubiquitin and likely holds for the mitochondrial carrier protein as well. As noted above, collagens are highly expressed in the *C. elegans* hypodermis and belong to a multigene family. Thus, we believe that all of these identified genes are false positives. Indeed, the SAM analysis predicts that, with this choice of parameters, three of the seven returned genes are expected to be falsely identified.

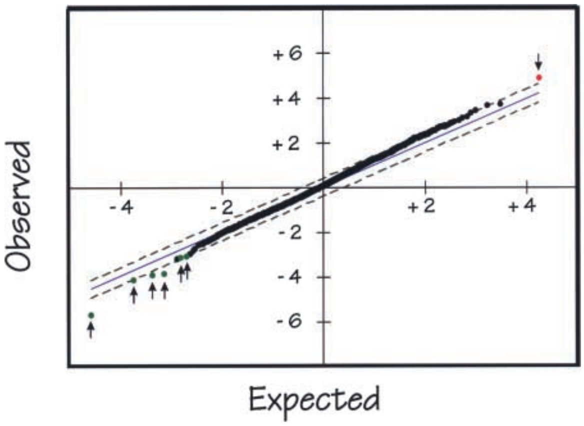

FIGURE 6.—SAM plot relating the (normalized) differences in hybridization intensity of each gene spot observed between the *elt-4(ca16)* and wild-type RNA *vs.* the expected average difference if the gene spot intensities are permuted. Dashed lines correspond to a "significance level" associated with a value of the adjustable parameter $\Delta = 0.4$ and a change in transcript level of ±20%. Only seven genes (indicated by arrows) meet these criteria. Further details are provided in the text.

We draw the same basic conclusion from these two different approaches to microarray analysis, namely that there is no evidence that the RNA population in the *elt-4(ca16)* larvae is significantly different from the RNA in wild-type L1 larvae. There is, of course, the possibility that genes could be expressed differentially in other stages of the life cycle. However, judging from *elt-4* expression patterns (Figure 2), the L1 larvae would seem to be a stage at which any *elt-4* dependent differences would be apparent.

Although our inability to identify *elt-4*-regulated genes was disappointing, nonetheless we were encouraged by the overall consistency of the replicated data and feel confident that significant differences could have been detected if they indeed existed. In any event, we wish to emphasize the importance of multiple independent replicates of the hybridization experiments. If we had performed the hybridizations only twice, an average of 136 (SD = 152) genes would have been identified as up- or downregulated by twofold (averaged over the 21 possible pairs of our hybridization data). With seven replicates, this list is reduced to roughly a half-dozen, all of which we interpret as being false positives.

ELT-4 binds weakly and nonspecifically to DNA and has no transcriptional activity in yeast: Up to this point in our analysis, we have been unable to uncover any function of *elt-4* in controlling gut genes. We thus decided to investigate whether ELT-4 does indeed bind to DNA. Recombinant protein was produced in bacteria, with either a polyhistidine tag at the N terminus or a GST tag at the C terminus; proteins were purified on the corresponding affinity columns and were used either with or without proteolytic removal of the affinity tag. ELT-4 protein was also produced by *in vitro* transcription-translation, in either the presence or the absence of cotranslated ELT-2. ELT-4 DNA interactions were investigated primarily by electrophoretic mobility shift assays (band shifts). As double-stranded DNA probes, we used the tandem pair of WGATAR sites that control the *C. elegans ges-1* gene (EGAN *et al.* 1995), as well as various candidate WGATAR-containing oligonucleotides identified in the *elt-2* enhancer. We also used a panel of WGATAR-containing oligonucleotides kindly provided by C. Trainor (National Institutes of Health, Bethesda, MD), including one particular oligonucleotide from the chicken α-globin promoter that has been found to bind strongly to every GATA factor yet investigated (C. TRAINOR, personal communication).

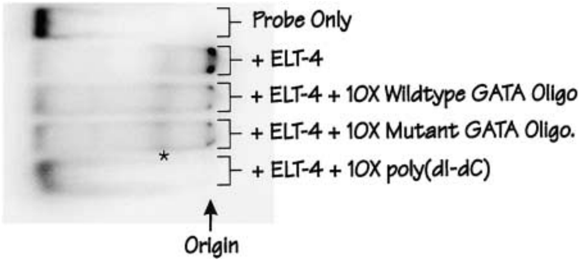
Probe Only
+ ELT-4
+ ELT-4 + 10X Wildtype GATA Oligo
+ ELT-4 + 10X Mutant GATA Oligo.
+ ELT-4 + 10X poly(dI-dC)
*
Origin

FIGURE 7.—Electrophoretic mobility shift assay to investigate the binding of ELT-4 protein to a double-stranded oligonucleotide probe containing the pair of tandem WGATAR sites from the *ges-1* promoter. Unlabeled double-stranded competitors are indicated on the various lanes. A smear (indicated by asterisk) is often seen on the autoradiograph but never forms a distinct band expected for a stable discrete protein-DNA complex.

Typical results are shown in Figure 7. Modest levels of ELT-4 protein cause all of the probe to collect at the top of the gel but this "binding" is both weak and nonspecific. Binding is largely abolished either by a 10-fold molar excess of the wild-type (double stranded) oligonucleotide, the same oligonucleotide but in which the WGATAR sites have been mutated, or by a 10-fold mass excess of nonspecific competitor poly(dIdC::dIdC). No reproducible band of intermediate migration that could correspond to a specific stable ELT-4::DNA complex was ever observed at any level of protein input. We estimate that, even if ELT-4 had a specific binding affinity 2–3 orders of magnitude lower than that measured with peptide F2B (OMICHINSKI *et al.* 1993b) or 10-fold lower than that measured with AREA (STARICH *et al.* 1998b), we would nonetheless have detected complex formation. No significant "extra" bands were observed when the experiments were repeated in the presence of a range of concentrations of purified ELT-2 protein (produced in baculovirus). Band shifts were performed over a wide range of experimental conditions, varying temperature, binding buffer, divalent cations (zinc, iron, etc.), electrophoresis buffer, the presence or absence of ELT-2 protein, and the level of nonspecific competitor polynucleotide [poly(dIdC:dIdC)].

We also renatured the protein from trifluoroacetic acid, exactly as used by Omichinski *et al.* (1993a) to produce effective binding in similar sized peptides from chicken GATA-1; renaturations were conducted in the presence of zinc, iron, or magnesium ions, all without success. As additional controls, we showed that both GST-tagged ELT-2 and *in vitro* translated ELT-2, produced under similar conditions, bind DNA tightly and specifically (data not shown). In separate experiments, we could find no evidence that GST-ELT-4, bound to glutathione-agarose beads, was able to interact with purified ELT-2 protein (data not shown).

Although ELT-4 appears to lack detectable sequence-specific DNA-binding activity when *in vitro* biochemical assays are used, it is possible that, under conditions more closely approximating an intracellular environment, ELT-4 could bind DNA and perhaps also act as a transcriptional activator. Shim *et al.* (1995) described an experimental system in which a *C. elegans* GATA factor (in their case, ELT-1) could be examined for its ability to activate transcription in *S. cerevisiae.* We have recently used the same system to explore transcriptional activation properties of the *C. elegans* PHA-4 protein in combination with a putative cofactor PEB-1 (Kalb *et al.* 2002). The system involves two cotransformed and independently selected plasmids: (i) a reporter plasmid in which the candidate *cis*-acting regulatory site (in this case, five copies of the tandem pair of GATA sites that control the *ges-1* gene; Egan *et al.* 1995) is placed in the position of upstream activating sequence (UAS) adjacent to a basal promoter driving transcription of a *lacZ* reporter gene and (ii) a second plasmid in which a cDNA for the candidate transcriptional activator (in our case, *elt-4* or *elt-2*) is transcribed under control of a galactose-inducible promoter (GAL1). As one set of negative controls, *elt-2* and *elt-4* coding sequences are cloned in the antisense orientation but still transcribed under GAL1 control. In a second set of negative controls, the reporter vector is "empty"; *i.e.*, no *cis*-acting sites have been inserted as UAS. Table 1 summarizes our results. ELT-2 confers high levels of β-galactosidase activity, several hundredfold above background. In contrast, ELT-4 produces no significant activity above background.

Although ELT-4 may have no activity by itself, it might nonetheless augment or inhibit the activation properties of ELT-2. Thus, we repeated the experiment using a construct in which both *elt-2* and *elt-4* coding sequences were expressed from the same plasmid, under independent GAL1 control and transcribed in the same direction. The negative control contains the correctly transcribed *elt-2* sequence but with the *elt-4* coding cDNA transcribed in the antisense direction relative to its galactose-regulated promoter. From the results shown in Table 1, it is clear that ELT-4 has no significant influence, either positive or negative, on the transcriptional activity produced by ELT-2.

DISCUSSION

In this article, we have identified *elt-4* as a new GATA-factor gene in the nematode *C. elegans.* In *C. elegans*, specification and differentiation of major tissue types such as the intestine and hypodermis depend critically on GATA transcription factors (Maduro and Rothman 2002; Patient and McGhee 2002) and thus the analysis of a new member of the class becomes an important step in understanding the overall regulatory hierarchy of embryonic development. In addition, *elt-4* is interesting because the encoded protein is exceptionally small, consisting of little beyond the DNA-binding domain.

The GATA factor Serpent plays a critical role in development of the Drosophila endoderm (Reuter 1994; Rehorn *et al.* 1996) and has recently been shown to produce an alternatively spliced transcript that encodes a protein with two zinc finger DNA-binding domains (Waltzer *et al.* 2002). The zinc finger domains of ELT-4 and ELT-2 are highly similar to the C-terminal zinc finger of Serpent (72–76% identity), raising the possibility that ELT-4 might actually be part of a two-fingered (possibly "homologous") variant of ELT-2. However, RT-PCR and our previous Northern analysis (Hawkins and McGhee 1995), together with the existence of a distinct *elt-4* cDNA clone, provided no evidence that *elt-4* sequences are transcribed as an alternatively spliced two-finger variant of the downstream *elt-2* gene.

elt-4 is expressed in the developing intestine (plus a few cells in the posterior pharynx). However, we could detect no function for *elt-4* by ectopic expression experiments, by analysis of an *elt-4* deletion mutant, or by genome-wide microarray analysis of potentially affected transcripts. We could detect no evidence that *elt-4* provided backup functions for *elt-2* and, indeed, we were unable to demonstrate sequence-specific ELT-4 binding. Thus, we are led to the following questions: where did *elt-4* come from, when did it arise, and why has the sequence of the zinc finger domain been so highly conserved?

To estimate when the *elt-4*/*elt-2* duplication event took place, we compared sequences between *C. elegans* and the related nematode *C. briggsae.* The *elt-2* homolog in *C. briggsae* was readily identified in the available genomic sequence (designated CBG17257). Sequences of the two ELT-2 proteins are highly conserved: 25/25 residues are identical in the zinc finger domain and 24/25 residues are identical in the basic region immediately adjacent. Overall, the two protein sequences are 68% identical (73% similar). The two chromosomal regions are at least locally syntenic in the two nematodes: that is, the C39B10.1 gene, a G protein-coupled receptor lying ~18 kb upstream of the *C. elegans elt-2* gene (see Figure 1A), has a clear homolog lying approximately the same distance upstream of the *C. briggsae elt-2* gene. However, no sequence that could potentially be the *C. briggsae* homolog of *elt-4* could be identified in the sequence be-

TABLE 1

Test for transcriptional activity of the *elt-4* and *elt-2* GATA factors in yeast

Plasmid	Transcriptional activator(s)	Reporter	Activity	SD	*n*
pJM169	*elt-4* sense	Empty vector	0.5	0.4	3
		5 × GATA pair	0.4	0.5	6
pJM170	*elt-4* antisense	Empty vector	0.6	0.7	4
		5 × GATA pair	0.2	0.4	5
pJM200	*elt-2* sense	Empty vector	−0.4	0.6	6
		5 × GATA pair	224.1	110.2	13
pJM201	*elt-2* antisense	Empty vector	0.3	0.4	6
		5 × GATA pair	0.3	0.2	6
pJM202	*elt-2* sense + *elt-4* sense	Empty vector	0.4	0.3	6
		5 × GATA pair	441.8	193.2	16
pJM204	*elt-2* sense + *elt-4* antisense	Empty vector	0.4	0.4	9
		5 × GATA pair	403.4	269.9	15
YCpGal3	Empty vector	Empty vector	0.5	0.3	6
		5 × GATA pair	0.4	0.4	6

Assay for β-galactosidase activity is essentially as described by Kalb *et al.* (2002); activity is recorded as "units × 1000." The control "Empty vector" is pLGΔ178 by itself; 5 × GATA pair refers to pLGΔ178 containing five copies of the tandem pair of GATA sites from the control region of the *ges-1* gene (sequence is ATGCATGCAAC**T GATAG**CAAAAC**TGATAA**GGGTCAA). As a negative control for the transcriptional activators, the vector YCp-GAL3 was used with no inserted cDNA. As an additional control, we verified that essentially all of the activity produced by pJM202 and pJM204 with the 5 × GATA pair reporter was dependent on the addition of galactose (data not shown). *n*, the number of independent cultures that were assayed (pooled from 2–3 completely independent replicates of the overall experiments).

tween the *C. briggsae* homologs of *elt-2* and C39B10.1 (or elsewhere in the currently available genomic sequence).

Thus, the two simplest models for the evolutionary history of the *elt-4/elt-2* gene pair are that: (i) the *elt-4/elt-2* duplication was present in the last common ancestor of *C. elegans* and *C. briggsae* but the *elt-4* homolog disappeared in the *C. briggsae* lineage or (ii) the *elt-4/elt-2* duplication event occurred only in the *C. elegans* lineage, after *C. elegans* and *C. briggsae* had diverged. To distinguish between these two alternatives, we aligned the sequences for the zinc finger DNA-binding domains of all three sequences, considered only amino acid positions that are identical in all three species (thereby attempting to avoid complications introduced by evolutionary selection), and counted the number of third-position synonymous codon changes that occurred in the three pairwise combinations. Multiple replacements were corrected using the simple Jukes-Cantor one-parameter model (Li 1997). The results are shown in Figure 8. The data clearly favor the model in which the *elt-4/elt-2* duplication event occurred only in the *C. elegans* lineage, after the point at which *C. elegans* and *C. briggsae* diverged 50–120 MYA (Coghlan and Wolfe 2002). Assuming uniform molecular clock rates (Li 1997), we estimate that the *elt-4/elt-2* gene duplication occurred ~25–55 MYA. The average lifetime of a duplicated gene in *C. elegans* is estimated to be only a few million years (Lynch and Conery 2000); hence, the *elt-4* gene, in spite of its lack of obvious function, has survived far longer than the average.

How did the *elt-4* gene arrive at its current abbreviated form? We suggest that *elt-4* arose as a complete duplication of the *elt-2* gene, not as a partial duplication of only the DNA-binding domain. Not only are the *elt-2* and *elt-4* N termini highly conserved but also a region in the 5′-flanking DNA 100–500 bp upstream of the *elt-4* ATG is highly conserved with a DNA region 1.8–2.2 kb upstream of the *elt-2* ATG, which in turn is highly conserved with a DNA region 2.2–2.6 kb upstream of the initiation codon of the *elt-2* gene in *C. briggsae*. Within these regions,

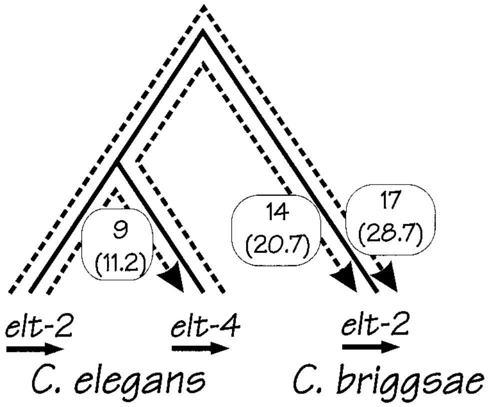

Figure 8.—Proposed evolutionary history of the *elt-4* gene. As described in more detail in the text, we suggest that *elt-4* arose as a duplication of the *elt-2* gene after *C. elegans* and *C. briggsae* had diverged from each other. The numbers placed on the proposed phylogenetic tree represent the numbers of third-position synonymous codon changes within the conserved DNA-binding domain for each of the three pairwise comparisons. Numbers in parentheses are corrected for multiple replacements.

there is an ~60-bp core sequence that is >90% conserved among *C. elegans elt-4, C. elegans elt-2,* and *C. briggsae elt-2,* a prime candidate for a *cis*-acting regulatory region. Thus we are confident that the original duplication involved the *elt-2* 5′-flanking region together with the majority of the coding region. We cannot make an equally definitive statement whether the full 3′-end of the *elt-2* gene was included in the original duplication but it seems most likely that it was: the region between the conserved DNA-binding domain and the 3′-end of the *elt-2* gene would be a much smaller recombinational target than the region between *elt-2* and the adjacent downstream gene. Thus we propose that *elt-4* was whittled down to its present size by internal deletions. However, this must have occurred in a very particular manner, retaining almost complete conservation of the zinc finger DNA-binding domain and with the size diminution presently at (possibly stalled at) close to the minimum size required for sequence-specific binding.

Thus, *elt-4* presents the interesting example of a duplicated gene for which no obvious biological function could be discerned but which, judging from the high degree of sequence conservation in the zinc finger, must have been under selective pressure in the past, if not at present. Of course, *elt-4* could have a subtle or infrequently required function that the present experiments would have overlooked completely. Indeed, it is well recognized that effects of a magnitude that could never be detected by current laboratory methods could nonetheless produce strong selective advantages in the natural environment (Li 1997; Nowak *et al.* 1997). Thus, a huge challenge will be to connect the essentially qualitative data produced by even the most sophisticated experiment in developmental biology to the quantitative data required to understand how alleles spread through populations. Only when this connection is established will we be able to test models proposing selective advantages conferred on a particular developmental variant.

We end by pointing out a further feature of the *elt-4/elt-2* duplication and presumably of tandem duplication events in general. The analysis of Semple and Wolfe (1999) indicates that the most probable configuration of duplicated genes in *C. elegans* is as a simple tandem duplication, *i.e.*, a duplication of a single gene with no intervening genes. Local duplications of two or three or more tandem genes occur with decreasing likelihood but the following argument would also apply to these cases as well, or at least to the genes on the borders of the duplication. Figure 9 considers the simplest model for how such local tandem duplications might be produced, namely as a result of misalignment of chromosome homologs, followed by unequal crossing over. Even if the complete gene coding sequences were to be duplicated, it would seem unlikely that all the gene regulatory sequences would also be duplicated in their entirety. In other words, the act of tandem duplication does not necessarily lead to two identical genes, one of which is now free to diverge. Rather, the duplication may well produce two genes, neither of which is controlled in the same manner as the original gene; right from their birth, the duplicate genes could have different expression patterns. It is certainly the general impression that *cis*-acting control sequences are more likely to be situated in the 5′-flanking region of a gene than in the 3′-flanking region (although we are not aware of a comprehensive compilation). Thus, other things being equal, it might be expected that the expression pattern of the 5′ member of a tandem gene duplication would be controlled more like the original parent gene than would its 3′ counterpart.

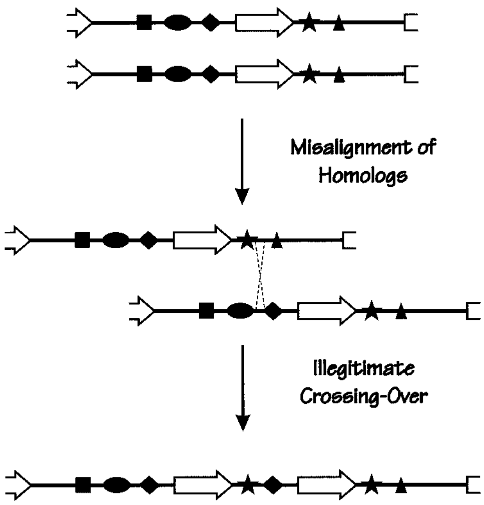

Figure 9.—Possible mechanism for generation of tandemly duplicated genes by illegitimate recombination between inaccurately paired chromosome homologs. The various solid symbols are meant to represent *cis*-acting control signals in the DNA. As described in more detail in the text, one potential consequence of such a duplication mechanism is that the *cis*-acting control signals that regulate either of the duplicated daughter genes may not include the full set of control signals that regulated the original parent gene. In other words, the act of tandem duplication *may* make the two duplicate genes unable to complement each other.

While the mechanism depicted in Figure 9 was not explicitly considered by Force *et al.* (1999), it certainly is in the spirit of their duplication-degeneration-complementation (DDC) model proposed to explain why genomes appear to have so many duplicate genes. In fact, it presents an extreme application of their DDC model: diverged functions are likely to appear immediately following tandem gene duplication, with no intervening time required for emergence of complementing functions. Similar considerations have been proposed by Averof (2002) for evolution of Hox genes. If *elt-4* had shown easily observable biological functions (as does *elt-2*), then it would have provided a fascinating experimental system in which to explore how the *elt-2* and *elt-4* regulatory regions are intertwined or have become extricated following the duplication event. Unfortunately, *elt-4* shows no obvious phenotype, even when measured quantitatively by whole-genome microarray analysis.

Whole-genome analysis (Lynch and Conery 2000) has focused attention on the wide spectrum of fates that await duplicated genes: the large majority of duplicates appear to become inactivated and rapidly disappear; rare duplicates have new functions and persist. The ultimate fate of *elt-4* would appear to lie between these two extremes: *elt-4* may have survived the initial postduplication culling but the odds are that it too will disappear. However, *elt-4* has obviously been resisting its demise; much like the grin of the Cheshire cat, the *elt-4* zinc finger could well be the last domain to disappear.

We thank Bernard Lemire (Department of Biochemistry, University of Alberta, Edmonton) and members of his laboratory for hospitality and for access to their *C. elegans* deletion library. We especially thank Stuart Kim (Department of Genetics, Stanford University) and the Stanford Microarray Facility for conducting the microarray hybridizations. We also thank Y. Kohara (National Institute of Genetics, Mishima, Japan) for providing a cDNA clone, A. Fire (Carnegie Institute, Baltimore) for providing *C. elegans* transformation vectors, and C. C. Trainor (National Institutes of Health, Bethesda, MD) for providing a set of GATA-site-containing oligonucleotide probes. We gratefully acknowledge support from the Japanese Society for Promotion of Science (T.F.) and from the Canadian Institutes of Health Research and the Alberta Heritage Foundation for Medical Research (J.D.M.).

LITERATURE CITED

Aamodt, E. J., M. A. Chung and J. D. McGhee, 1991 Spatial control of gut-specific gene expression during Caenorhabditis elegans development. Science **252:** 579–582.

Albertson, D. G., and J. N. Thomson, 1976 The pharynx of Caenorhabditis elegans. Philos. Trans. R. Soc. Lond. B Biol. Sci. **275:** 299–325.

Averof, M., 2002 Arthropod Hox genes: insights on the evolutionary forces that shape gene functions. Curr. Opin. Genet. Dev. **12:** 386–392.

Blumenthal, T., D. Evans, C. D. Link, A. Guffanti, D. Lawson *et al.*, 2002 A global analysis of Caenorhabditis elegans operons. Nature **417:** 851–854.

Bonner, J. J., 1991 Vectors for the expression and analysis of DNA-binding proteins in yeast. Gene **104:** 113–118.

Brenner, S., 1974 The genetics of *Caenorhabditis elegans*. Genetics **77:** 71–94.

Coghlan, A., and K. H. Wolfe, 2002 Fourfold faster rate of genome rearrangement in nematodes than in Drosophila. Genome Res. **12:** 857–867.

Edgar, L. G., and J. D. McGhee, 1986 Embryonic expression of a gut-specific esterase in Caenorhabditis elegans. Dev. Biol. **114:** 109–118.

Egan, C. R., M. A. Chung, F. L. Allen, M. F. Heschl, C. L. Van Buskirk *et al.*, 1995 A gut-to-pharynx/tail switch in embryonic expression of the Caenorhabditis elegans ges-1 gene centers on two GATA sequences. Dev. Biol. **170:** 397–419.

Fire, A., 1999 RNA-triggered gene silencing. Trends Genet. **15:** 358–363.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan *et al.*, 1999 Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151:** 1531–1545.

Fukushige, T., D. F. Schroeder, F. L. Allen, B. Goszczynski and J. D. McGhee, 1996 Modulation of gene expression in the embryonic digestive tract of C. elegans. Dev. Biol. **178:** 276–288.

Fukushige, T., M. G. Hawkins and J. D. McGhee, 1998 The GATA-factor elt-2 is essential for formation of the Caenorhabditis elegans intestine. Dev. Biol. **198:** 286–302.

Fukushige, T., M. J. Hendzel, D. P. Bazett-Jones and J. D. McGhee, 1999 Direct visualization of the elt-2 gut-specific GATA factor binding to a target promoter inside the living Caenorhabditis elegans embryo. Proc. Natl. Acad. Sci. USA **96:** 11883–11888.

Geiser, M., R. Cebe, D. Drewello and R. Schmitz, 2001 Integration of PCR fragments at any specific site within cloning vectors without the use of restriction enzymes and DNA ligase. Biotechniques **31:** 88–90, 92.

Hawkins, M. G., and J. D. McGhee, 1995 elt-2, a second Gata factor from the nematode Caenorhabditis elegans. J. Biol. Chem. **270:** 14666–14671.

Kalb, J. M., K. K. Lau, B. Goszczynski, T. Fukushige, D. Moons *et al.*, 1998 pha-4 is Ce-fkh-1, a fork head/HNF-3 homolog that functions in organogenesis of the C. elegans pharynx. Development **125:** 2171–2180.

Kalb, J. M., L. Beaster-Jones, A. P. Fernandez, P. G. Okkema, B. Goszczynski *et al.*, 2002 Interference between the PHA-4 and PEB-1 transcription factors in formation of the Caenorhabditis elegans pharynx. J. Mol. Biol. **320:** 697–704.

Kennedy, B. P., E. J. Aamodt, F. L. Allen, M. A. Chung, M. F. Heschl *et al.*, 1993 The gut esterase gene (ges-1) from the nematodes Caenorhabditis elegans and Caenorhabditis briggsae. J. Mol. Biol. **229:** 890–908.

Kim, S. K., J. Lund, M. Kiraly, K. Duke, M. Jiang *et al.*, 2001 A gene expression map for Caenorhabditis elegans. Science **293:** 2087–2092.

Krause, M., and D. Hirsh, 1987 A trans-spliced leader sequence on actin mRNA in C. elegans. Cell **49:** 753–761.

Li, W.-H., 1997 *Molecular Evolution*. Sinauer Associates, Sunderland, MA.

Lowry, J. A., and W. R. Atchley, 2000 Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain. J. Mol. Evol. **50:** 103–115.

Lynch, M., and J. S. Conery, 2000 The evolutionary fate and consequences of duplicate genes. Science **290:** 1151–1155.

Maduro, M. F., and J. H. Rothman, 2002 Making worm guts: the gene regulatory network of the Caenorhabditis elegans endoderm. Dev. Biol. **246:** 68–85.

Maduro, M. F., M. D. Meneghini, B. Bowerman, G. Broitman-Maduro and J. H. Rothman, 2001 Restriction of mesendoderm to a single blastomere by the combined action of SKN-1 and a GSK-3beta homolog is mediated by MED-1 and -2 in C. elegans. Mol. Cell **7:** 475–485.

Mains, P. E., and J. D. McGhee, 1999 *Biochemistry of C. elegans*. Oxford University Press, Eynsham, UK.

Marshall, S. D., and J. D. McGhee, 2001 Coordination of ges-1 expression between the Caenorhabditis pharynx and intestine. Dev. Biol. **239:** 350–363.

McGhee, J. D., J. C. Birchall, M. A. Chung, D. A. Cottrell, L. G. Edgar *et al.*, 1990 Production of null mutants in the major intestinal esterase gene (ges-1) of the nematode *Caenorhabditis elegans*. Genetics **125:** 505–514.

Mello, C., and A. Fire, 1996 DNA transformation, pp. 451–482 in *Methods in Cell Biology*, edited by H. F. Epstein and D. C. Shakes. Academic Press, San Diego.

Montgomery, M. K., S. Xu and A. Fire, 1998 RNA as a target of double-stranded RNA-mediated genetic interference in Caenorhabditis elegans. Proc. Natl. Acad. Sci. USA **95:** 15502–15507.

Nadon, R., and J. Shoemaker, 2002 Statistical issues with microarrays: processing and analysis. Trends Genet. **18:** 265–271.

Nowak, M. A., M. C. Boerlijst, J. Cooke and J. M. Smith, 1997 Evolution of genetic redundancy. Nature **388:** 167–171.

Omichinski, J. G., G. M. Clore, O. Schaad, G. Felsenfeld, C. Trainor *et al.*, 1993a NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. Science **261:** 438–446.

Omichinski, J. G., C. Trainor, T. Evans, A. M. Gronenborn, G. M. Clore *et al.*, 1993b A small single-"finger" peptide from the erythroid transcription factor GATA-1 binds specifically to DNA as a zinc or iron complex. Proc. Natl. Acad. Sci. USA **90:** 1676–1680.

Patient, R. K., and J. D. McGhee, 2002 The GATA family (vertebrates and invertebrates). Curr. Opin. Genet. Dev. **12:** 416–422.

Rehorn, K. P., H. Thelen, A. M. Michelson and R. Reuter, 1996 A molecular aspect of hematopoiesis and endoderm development common to vertebrates and Drosophila. Development **122:** 4023–4031.

Reuter, R., 1994 The gene serpent has homeotic properties and specifies endoderm versus ectoderm within the Drosophila gut. Development **120:** 1123–1135.

Sambrook, J., and D. W. Russell, 2001 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Semple, C., and K. H. Wolfe, 1999 Gene duplication and gene conversion in the Caenorhabditis elegans genome. J. Mol. Evol. **48:** 555–564.

Shim, Y. H., J. J. Bonner and T. Blumenthal, 1995 Activity of a C. elegans GATA transcription factor, ELT-1, expressed in yeast. J. Mol. Biol. **253:** 665–676.

Snedcor, G. W., and W. G. Cochran, 1980 *Statistical Methods*. The Iowa State University Press, Ames, Iowa.

Starich, M. R., M. Wikstrom, H. N. Arst, Jr., G. M. Clore and A. M. Gronenborn, 1998a The solution structure of a fungal AREA protein-DNA complex: an alternative binding mode for the basic carboxyl tail of GATA factors. J. Mol. Biol. **277:** 605–620.

Starich, M. R., M. Wikstrom, S. Schumacher, H. N. Arst, Jr., A. M. Gronenborn *et al.*, 1998b The solution structure of the Leu22→Val mutant AREA DNA binding domain complexed with a TGATAG core element defines a role for hydrophobic packing in the determination of specificity. J. Mol. Biol. **277:** 621–634.

Stein, L., P. Sternberg, R. Durbin, J. Thierry-Mieg and J. Spieth, 2001 WormBase: network access to the genome and biology of Caenorhabditis elegans. Nucleic Acids Res. **29:** 82–86.

Stringham, E. G., D. K. Dixon, D. Jones and E. P. Candido, 1992 Temporal and spatial expression patterns of the small heat shock (hsp16) genes in transgenic Caenorhabditis elegans. Mol. Biol. Cell **3:** 221–233.

Tsang, W. Y., L. C. Sayles, L. I. Grad, D. B. Pilgrim and B. D. Lemire, 2001 Mitochondrial respiratory chain deficiency in Caenorhabditis elegans results in developmental arrest and increased life span. J. Biol. Chem. **276:** 32240–32246.

Tusher, V. G., R. Tibshirani and G. Chu, 2001 Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA **98:** 5116–5121.

Waltzer, L., L. Bataille, S. Peyrefitte and M. Haenlin, 2002 Two isoforms of Serpent containing either one or two GATA zinc fingers have different roles in Drosophila haematopoiesis. EMBO J. **21:** 5477–5486.

Wood, W. B., 1988 *The Nematode Caenorhabditis elegans*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Zhu, J., R. J. Hill, P. J. Heid, M. Fukuyama, A. Sugimoto *et al.*, 1997 *end-1* encodes an apparent GATA factor that specifies the endoderm precursor in *Caenorhabditis elegans*. Genes Dev. **11:** 2883–2896.

Zhu, J., T. Fukushige, J. D. McGhee and J. H. Rothman, 1998 Reprogramming of early embryonic blastomeres into endodermal progenitors by a Caenorhabditis elegans GATA factor. Genes Dev. **12:** 3809–3814.

Communicating editor: P. Anderson