Available online at www.sciencedirect.com
SCIENCE DIRECT
Developmental Biology 289 (2006) 444 – 455
DEVELOPMENTAL BIOLOGY
www.elsevier.com/locate/ydbio

Genomes & Developmental Control

Med-type GATA factors and the evolution of mesendoderm specification in nematodes

Cristian Coroian, Gina Broitman-Maduro, Morris F. Maduro *

Department of Biology, UC Riverside, 3380 Spieth Hall, Riverside, CA 92521, USA

Received for publication 18 May 2005, revised 7 October 2005, accepted 22 October 2005
Available online 1 December 2005

Abstract

In the nematode, *C. elegans*, the divergent GATA-type transcription factors MED-1 and MED-2 are encoded by an unlinked, redundant pair of intronless genes. The *med-1,2* genes are among the first to be activated in the embryo and are critical for the specification of the 7-cell stage MS (mesoderm) and E (endoderm) precursor cells. We have previously shown that the binding site recognized by MED-1 is a noncanonical RAGTATAC site that is not expected from the resemblance of its single C4-type zinc finger to those of other known GATA factors, which recognize the consensus HGATAR. To date, no MED-like zinc fingers have been described outside of *C. elegans*. In order to understand the evolution of these transcription factors, and the evolution of gene networks that specify early cell fates in *Caenorhabditis*, we have identified *med* sequence homologs in the related nematodes *C. briggsae* and *C. remanei*. While *C. briggsae* encodes two *med*-like genes similar to *C. elegans*, we find evidence for seven distinct *med*-like genes in *C. remanei*. Somewhat unexpectedly, the coding regions of all *med* genes appear to lack introns. We report that the *med* homologs have similar expression in their respective species. We further show that the *C. briggsae* homologs, and at least five of the seven *C. remanei* homologs, can fully complement the embryonic lethal phenotype of a *C. elegans med-1,2(−)* strain. We conclude that Med function and expression have been conserved over tens of millions of years of evolution, and that there may be a mechanism that selects against the acquisition of introns in these genes.

Keywords: Intronless genes; *C. remanei*; med-1; Gene duplication; GATA factors; Zinc fingers; Transcription factors; Cell fate specification

Introduction

Deployment of appropriate gene regulatory networks is the means by which different cells acquire different fates during metazoan development (Levine and Davidson, 2005). Gene comparison among closely-related species is one method that can be used to learn more about such networks, as conserved features can reveal fundamental properties that change very little over evolutionary time (Yuh et al., 2002). Such studies contribute to our understanding of how the generation of diversity among metazoans is related to changes in the interactions among genes.

We are interested in the evolution of the gene regulatory network that operates in mesendoderm specification in *Caenorhabditis*. At the 7-cell stage of embryonic development in *C. elegans*, the precursor cells MS and E are born (Sulston et al., 1983). The E cell clonally establishes the entire intestine (midgut), which consists of 20 cells at hatching, while its sister cell, MS, generates 80 cells that are primarily mesodermal in character, including body wall muscle and cells of the posterior half of the feeding organ, the pharynx.

The transcriptional regulatory cascade that controls specification of MS and E in *C. elegans* is well understood (Fig. 1) (Maduro and Rothman, 2002). The maternal gene *skn-1* encodes a bZIP/homeodomain transcription factor required to specify MS and E: loss of *skn-1* function results in a transformation of both MS and E into the mesectodermal precursor C, which makes body wall muscle and hypodermis (Bowerman et al., 1992). In the mother of MS and E, called EMS, SKN-1 activates the expression of the *med-1* and *med-2* genes (Bowerman et al., 1993; Maduro et al., 2001). Depletion of *med-1,2* by RNAi also results in a transformation of MS and E to C, although E is correctly specified in approximately half of mutant embryos (Maduro et al., 2001). In the E cell, MED-1,2 activate the genes *end-1* and *end-3*, which encode a pair of

* Corresponding author. Fax: +1 951 827 4286.
E-mail address: mmaduro@citrus.ucr.edu (M.F. Maduro).

0012-1606/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2005.10.024

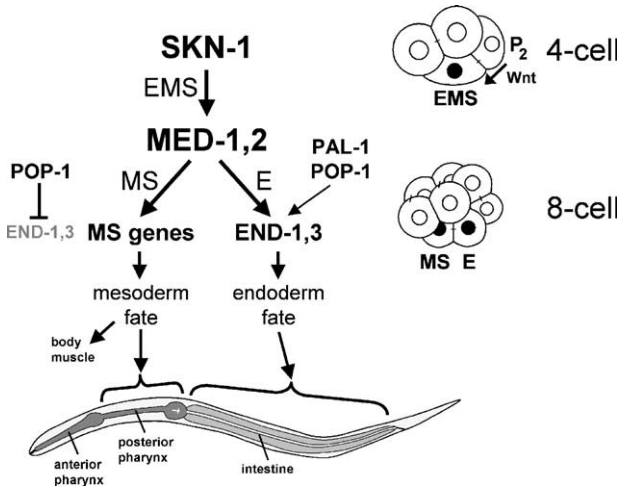


Fig. 1. Simplified representation of the regulatory cascade specifying MS and E in *C. elegans*. Diagrams of a 4-cell and 8-cell embryo are shown for reference. The maternal SKN-1 transcription factor activates zygotic expression of *med-1* and *med-2* in EMS (Maduro et al., 2001). In turn, MED-1,2 activate target genes in MS and E to specify their fates. In the E cell, MED-1,2 activate *end-1,3* to specify an endodermal fate (Maduro et al., 2002). The Wnt effector POP-1 is a coregulator along with the MEDs: in MS, POP-1 represses *end-1,3*, while in the E cell, the repressive activity of POP-1 is blocked as a result of the Wnt signaling event that polarizes EMS at the 4-cell stage (Goldstein, 1992; Maduro et al., 2002; Rocheleau et al., 1997; Thorpe et al., 1997). In E, POP-1 and PAL-1 also function as *end-1,3* activators in parallel with the MEDs (Maduro and Rothman, 2002; Maduro et al., 2005b). MS gives rise to primarily mesodermal cell types, including body wall muscle and the posterior half of the pharynx. The E cell clonally establishes the entire midgut (intestine). Sister cells are joined by a short line, and the nuclei of EMS, MS, and E are shaded black. In this and other figures, embryos are shown with anterior to the left and dorsal up.

GATA factors that specify an endoderm fate (Maduro and Rothman, 2002; Maduro et al., 2005a). In the MS cell, the corepressor POP-1 blocks the ability of MED-1,2 to activate *end-1,3*, allowing the *med*s to promote a mesodermal fate (Maduro et al., 2002). The repressive function of POP-1 is blocked in the E cell by transduction of a Wnt/MAPK/Src signal that originates from a cell–cell interaction between EMS and its sister cell, P_2 (Bei et al., 2002; Goldstein, 1992; Rocheleau et al., 1997, 1999; Shin et al., 1999; Thorpe et al., 1997). POP-1 then functions as an activator, and along with the Caudal-like protein PAL-1, also contributes to E specification in parallel with *med-1,2* (Maduro et al., 2005b). The *med-1* and *med-2* genes therefore play a critical role in the specification of the MS and E progenitors, as they function at the interface of SKN-1 activation and transduction of a signal initiated by the P_2–EMS interaction.

The *med-1* and *med-2* genes encode nearly-identical single C4 zinc finger proteins with similarity to the GATA family of transcription factors, which recognize the canonical binding site HGATAR (Lowry and Atchley, 2000; Maduro et al., 2001). MED-1 (and presumably MED-2, whose DNA-binding domain differs by only a single amino acid) recognizes RAGTATAC sites, suggesting that MED-1,2 have diverged in function from the GATA family (Broitman-Maduro et al., 2005). No proteins similar to MED-1 are known to exist outside of *Caenorhabditis* that are not themselves more similar to canonical GATA factors, suggesting that they may be a nematode invention. Because of the uniqueness of the MEDs among eukaryotic regulators, little is known of their evolutionary origin, and structure–function properties of the MED DNA-binding domain have yet to be elucidated.

In order to expand the collection of known MED-like regulators, and as first step toward studying the mesendoderm gene network in other species, we have mined the genome sequences of the related nematodes, *C. briggsae* (a hermaphroditic species) and *C. remanei* (a male–female species). *C. remanei* and *C. briggsae* are more closely related to each other than *C. elegans* is from either (Cho et al., 2004; Kiontke et al., 2004). By some estimates, *C. elegans* is 80–110 million years diverged from the last common ancestor to *C. briggsae*, and hence *C. remanei* (Coghlan and Wolfe, 2002; Stein et al., 2003). In this study, we compare the structure and genome organization of the *med* homologs from *C. briggsae* (two *med* genes) and *C. remanei* (seven *med* genes), and provide evidence that their expression and function have been conserved.

Materials and methods

Strains and genetics

The following strains were used in this work: PB4641 (wild-type *C. remanei* strain), AF16 (wild-type *C. briggsae*), N2 (wild-type *C. elegans*), MS240 [*unc-119(ed4) III; irEx102[unc-119(+), pMM808]*], MS247 [*med-1(ok804) X; med-2(cx9744) III; irDp1(III;f)*]. Transgenic strains used to test rescue are listed in Table 1. The free duplication *irDp1* (carried by MS247) was derived from the free duplication *sDp3*, which carries a single copy of *med-2*, after spontaneous integration of an array (*irEx14*) consisting of the *Ce-med-1* plasmid pMM277, the *unc-32*-rescuing plasmid pAIE5 (Pujol et al., 2001), and an *unc-119*::YFP reporter plasmid (pMM531). A more detailed description of the properties of MS247 will appear elsewhere.

Table 1
Rescue of a *C. elegans med-1,2(−)* mutant with *med* homolog transgenes

Strain(s)	Transgene	Rescue[a]	Viable	Arrested embryo		Total
				With gut	Without gut	
MS247	None[b]	n/a	0% (0)	58% (70)	42% (50)	120
MS290, MS291	*Ce-med-1*	Yes	97% (203)	1% (3)	1% (3)	209
MS402, MS403	*Ce-med-2*	Yes	98% (353)	2% (6)	0% (1)	360
MS298, MS299	*Cb-med-1*	Yes	80% (237)	18% (53)	3% (8)	298
MS420	*Cb-med-2*	Yes	33% (14)	57% (24)	10% (4)	42
MS293, MS295	*Cr-med-1*	Yes	76% (148)	24% (46)	0% (0)	194
MS335, MS336	*Cr-med-2*	No	0% (0)	61% (57)	39% (37)	94
MS314, MS315	*Cr-med-3*	Yes	86% (128)	13% (19)	1% (2)	149
MS317, MS393	*Cr-med-4*	Yes	64% (319)	35% (176)	1% (7)	502
MS319, MS320	*Cr-med-5*	Yes	87% (290)	11% (38)	2% (5)	333
MS322, MS323	*Cr-med-6*	Yes	93% (169)	6% (11)	1% (2)	182
MS337, MS338	*Cr-med-7*	No	0% (0)	90% (53)	10% (6)	59

[a] MS247 [*med-1(ok804); med-2(cx9744); irDp1*] animals were made transgenic for a plasmid or PCR product containing the indicated gene and an *unc-119*::CFP reporter. Lines that could be propagated in the absence of the *irDp1* balancer were considered to be rescued. For those that did not rescue, lines were constructed that also contained the *rol-6*^{D} marker plasmid (pRF4) to facilitate maintenance of the transgenes. Only animals expressing *unc-119*::CFP (indicative of the test array) but not *unc-119*::YFP (indicative of *irDp1*) were scored. For those genes that rescued, only *unc-119*::CFP embryos were scored.

[b] Non-*irDp1* MS247 segregants (i.e., embryos lacking *unc-119*::YFP) were scored. n/a, not applicable.

Identification of C. briggsae and C. remanei med genes

The *C. briggsae med* homologs correspond to CBG18154 (*Cb-med-1*) and CBG18155 (*Cb-med-2*) in WormBase (http://www.wormbase.org, release WS140). To identify candidate *med* genes from *C. remanei*, we used a TBLASTN search of the preliminary genome assembly of *C. remanei* (as of 7/29/2004) with *Ce*-MED-1 using the Washington University BLAST server (http://www.genome.wustl.edu/blast/client.pl). Sequences were extended in the 5′ and 3′ directions by correlating the aligned segment pairs to the sequence files downloaded from the Washington University server (ftp://www.genome.wustl.edu/pub/seqmgr/remanei/plasmid_assembly/). Gene sequences were compiled, annotated, and analyzed using Vector NTI 6 (Informax) on a Dell PC. The Accession numbers for the *Cr-med* genes are as follows: *Cr-med-1*, *-2*, and *-3*, DQ288162; *Cr-med-4* and *-5*, DQ288163; *Cr-med-6*, DQ288164; and *Cr-med-7*, DQ288165.

RNA interference (RNAi)

For RNAi experiments, double-stranded RNA (dsRNA) was synthesized using the T7 MEGAscript kit (Ambion) from a PCR product containing the *Cr-med-1*, *Cr-med-3*, or GFP coding regions tagged with the T7 RNA polymerase recognition sequence on both ends. dsRNA at 4 μg/μL was injected into both gonad arms of *C. remanei* females as was described for *C. elegans* hermaphrodites (Fire et al., 1998). Depletion of *skn-1* for testing the *Cr-med-1*::GFP translational fusion strain MS240 was performed by placing L2 animals on *E. coli* HT115 bacteria induced to accumulate *skn-1* dsRNA (Montgomery, 2004), and scoring progeny embryos in the uterus of these animals 3 days later. All RNAi experiments were performed at 20°C.

Transgenic C. elegans strains and plasmid construction

Transgenic animals were generated as described (Mello et al., 1991). The *Cr-med-1*::GFP::MED-1 fusion pMM808 was constructed as follows. The *Cr-med-1* promoter and coding region/3′UTR were sequentially cloned into a pBluescript KS(-) backbone as separate PCR products amplified from PB4641 genomic DNA. An open-ended GFP cassette, amplified from vector pPD95.67 (a gift from Andrew Fire, Stanford University), was inserted at the join between these segments. PCR and cloning were performed according to standard protocols; oligonucleotide sequences and cloning details are available upon request.

For rescue assays, plasmids carrying *Ce-med-1* (pMM277), *Ce-med-2* (pMM493), *Cb-med-1* (pMM556), and *Cr-med-1* (pMM807), and PCR products carrying the remaining *Cr-med* genes were generated containing ~0.5 kbp–1 kbp 5′ to the coding region and ~0.3 kbp 3′ sequence. Plasmid DNA (50–100 ng/μL) or gel-purified PCR products (20 ng/μL) were coinjected with *unc-119*::CFP (pMM809) into MS247. Lines that carried *unc-119*::CFP but lacked *unc-119*::YFP (indicative of *irDp1*) were identified using appropriate filter sets (Miller et al., 1999) on a Leica dissecting microscope. The absence of *irDp1* could also be confirmed because these strains lacked an adult egg-laying defect associated with *irDp1*. Thereafter, rescued lines were self-maintaining, and transgenic embryos could be scored by pan-neuronal *unc-119*::CFP expression from midembryogenesis to adulthood (Maduro and Pilgrim, 1995). For *Cr-med-2* and *Cr-med-7*, which did not rescue, lines were made in which the arrays were also marked with *rol-6*^{D} (Mello et al., 1991). PCR confirmed the presence of *Cr-med-2* and *Cr-med-7* in these strains.

In situ hybridization

Detection of *med* transcripts in whole-mount embryo preparations was performed essentially as described (Seydoux and Fire, 1995), except that embryos were fixed using Streck Tissue Fixative (STF; Streck Laboratories, Inc.) and antisense RNA probes were used. A more detailed protocol will appear elsewhere.

RT-PCR analysis

Embryos were obtained from *C. remanei* females by bleaching. After homogenization in Trizol (Invitrogen), total RNA was purified from the aqueous phase using the Qiagen RNeasy Kit according to the manufacturer's instructions. Residual genomic DNA was removed by treatment with DNaseI (Ambion). First-strand and PCR primers were designed to target specific *Cr-med* genes. Superscript II Reverse Transcriptase (Invitrogen) was used according to the manufacturer's instructions, except that 0.5 μL of T4gp32 (USB) was added to the RT reactions in 10 μL total volume.

Results

C. elegans and C. briggsae have two med genes, while C. remanei has seven

In *C. elegans*, the *med-1* and *med-2* genes are found on separate chromosomes (Fig. 2) (Maduro et al., 2001). In *C. briggsae*, there are two *med*-like genes found within 2 kbp of each other in an inverted orientation. In both *C. elegans* and *C. briggsae*, the intraspecific MED paralogs are nearly identical (98% for *Ce*-MED-1/2 and 97% for *Cb*-MED-1/2), while *Ce*-MED-1 and *Cb*-MED-1 are only 29% identical overall, consistent with recent duplication. We used the sequence of *C. elegans* MED-1 to search contigs from the preliminary genome assembly of *C. remanei*. Using the TBLASTN algorithm (Altschul et al., 1990), seven distinct coding regions were identified (Figs. 2 and 3). Like their *C. elegans* and *C. briggsae* counterparts, all seven *Cr-med* genes appear to lack introns, as we could find no good candidate splice donor and acceptor sites similar to those used in *C. elegans* (Blumenthal and Steward, 1997). The conservation of amino acids and lack of in-frame stop codons across the coding regions further support an intronless structure (Fig. 3). Putative polyadenylation sites (AATAAA) occur within 73–108 base pairs downstream of predicted *Cb-med*s and *Cr-med*s (Fig. 2). The *C. remanei med* genes display a combination of linked and unlinked duplications: *Cr-med-1*, *-2*, and *-3* are located within 5 kbp of each other, while *Cr-med-4* and *-5* appear in adjacent, parallel orientation within 2 kbp (Fig. 2). The remaining genes, *Cr-med-6* and *Cr-med-7*, appear on separate contigs. As these identifications were based on a preliminary sequence assembly, further synteny among all the *Cr-med* genes, and the existence of additional *med* paralogs, cannot be ruled out.

Most of the *Cr-med* genes encode proteins of a similar size (164–197aa), while *Cr-med-2* contains a 5′ truncated coding region and is predicted to encode a 93-aa polypeptide that includes the DNA-binding domain (Figs. 2 and 3). An inspection of the protein alignments (Fig. 3) and a phylogenetic tree generated from their DNA-binding domains (Fig. 4C) strongly suggest that the intraspecific *med* duplications arose from a single ancestral gene by duplication. Hence, the formation and retention of new *med* paralogs in *Caenorhabditis* appear to be a common occurrence for this gene family.

Conserved structural features of the MED class of GATA factors

The *Ce*-MED proteins contain a single C4-type zinc finger and basic domain which recognize a non-GATA consensus sequence (Broitman-Maduro et al., 2005). By alignment with GATA DNA-binding domains whose solution structures have

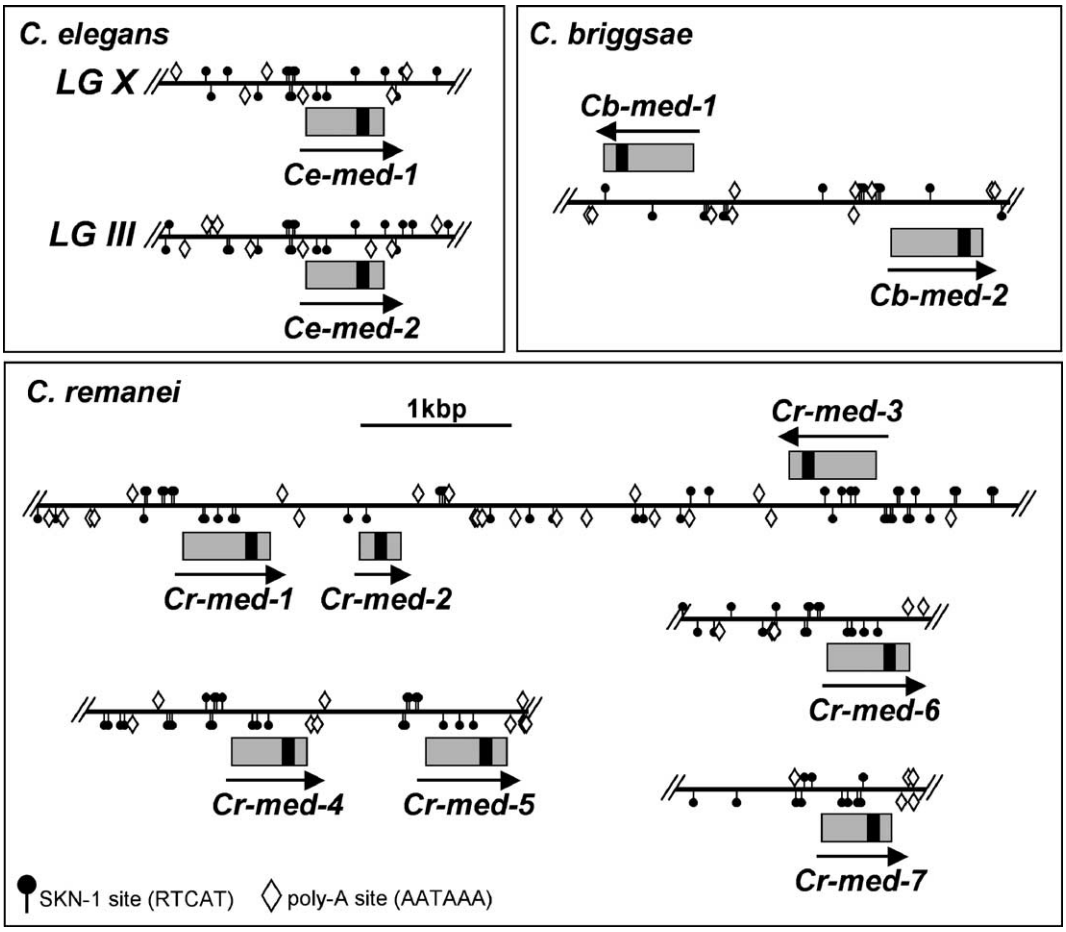


Fig. 2. The arrangement of *med* genes in *C. elegans*, *C. briggsae*, and *C. remanei*. SKN-1 sites (lollipops) and poly-A sites (diamonds) are shown above the DNA if they occur on the top strand and below the DNA if they occur on the complementary strand. The coding regions for the *med* genes are shown as shaded rectangles, with the C4 zinc fingers indicated by darker shading.

been determined, MED-1,2 contain apparent amino acid differences at positions known to make base-specific contacts (Broitman-Maduro et al., 2005; Omichinski et al., 1993; Starich et al., 1998). We examined the entire suite of 11 MED-like coding regions to identify amino acids that may define this divergent GATA subtype. An alignment of the predicted DNA-binding domains of the *C. elegans* and *C. remanei* MEDs, as well as those of other GATA factors, reveals structural similarities (Fig. 4A). First, the MED zinc fingers all include an extra amino acid between the cysteine pairs: The fungal *AreA*, chicken cGATA-1, and *C. elegans* ELT-1 and ELT-3 GATA factors contain 17 residues between the 2nd and 3rd cysteines, while all of the MED proteins contain 18, raising the intriguing possibility that the additional amino acid accounts for the GTATA core of the MED binding site. Second, we can identify nine amino acids that occur among all MED proteins, but which do not occur in other GATA factors (indicated by asterisks in Fig. 4A). Although the functional significance of these amino acids is not yet known, their high degree of conservation suggests that at least some of them may mediate protein–DNA interactions that determine MED binding specificity. Finally, a phylogenetic tree generated from the alignment shown in Fig. 4A validates the notion that the MEDs do indeed form a distinct subgroup of GATA factors (Fig. 4B).

5′ flanking gene sequences predict similar expression of Cr-med genes

In addition to similar coding regions as their *C. elegans* counterparts, the *Cb-med* and *Cr-med* genes display multiple recognition sequences for the SKN-1 transcription factor in the 5′ flanking sequence (Fig. 2). SKN-1 has structural features found in both bZIP and homeodomain transcription factors, and its core binding site has been determined as RTCAT (Blackwell et al., 1994). Of the known direct SKN-1 targets in *C. elegans*, regulation by SKN-1 in vivo appears to require clusters of SKN-1 sites in close proximity (An and Blackwell, 2003; Maduro et al., 2001). Both *C. elegans med* genes contain clusters of SKN-1 sites ~90 bp upstream of the start codon. These sites mediate binding by SKN-1 in vitro and are required for activation of a *med-1* reporter gene in vivo (Maduro et al., 2001). For both *Cb-med* genes, and five of the seven *Cr-med* genes, there are at least six SKN-1 sites within the 250 bp directly upstream of the predicted start codon. As this is expected to occur by chance in less than 0.5% of all DNA segments of similar size and base composition, the presence of these sites suggests that they have undergone positive selection. The occurrence of multiple SKN-1 sites in the *Cb-med* and *Cr-med* genes, therefore, suggests that most of them are direct targets of a SKN-1-like activator. Two of

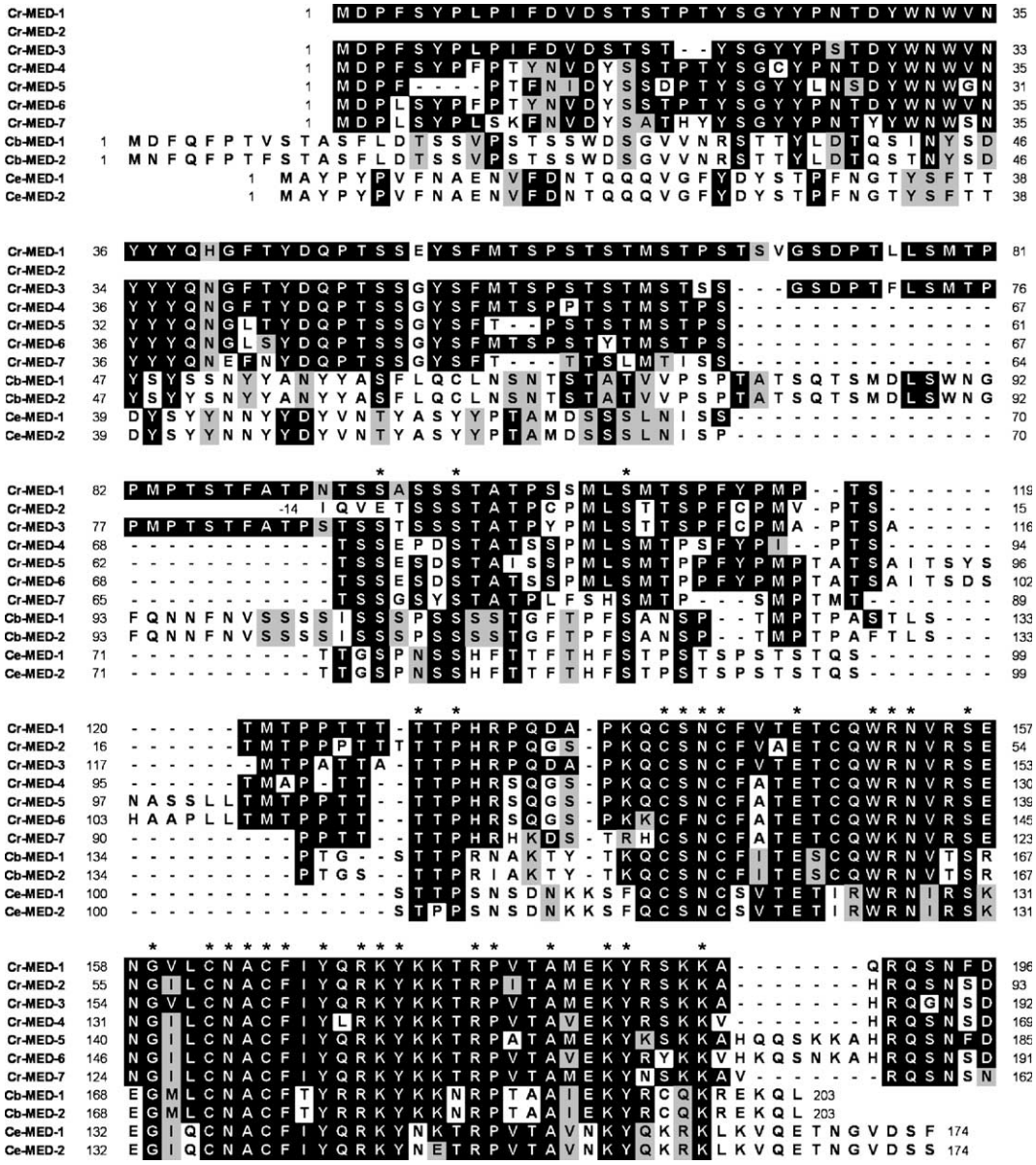


Fig. 3. Alignment of all known MED proteins. White text on a black background indicates identities with *Cr*-MED-1, while black text on a gray background indicates a conservative change as defined by the BLOSUM62mt2 matrix (Henikoff and Henikoff, 1992). Amino acid positions relative to the start codon are indicated. For Cr-MED-2, an additional 14 codons specified upstream of the putative start codon have been included in the alignment. Asterisks indicate residues that are conserved in all MED coding regions (excepting those in *Cr*-MED-2 that occur upstream of its putative start codon).

the *Cr-med* homologs have anomalies that suggest that they may not be expressed at high efficiency downstream of SKN-1. While all other *med* genes except *Cr-med-2* and *Cr-med-7* contain overlapping SKN-1 sites of the form RTC*ATCAT* (one site underlined, the other overlapping site in italics), *Cr-med-2* contains only a single SKN-1 site in its upstream region, and while *Cr-med-7* contains four SKN-1 sites in close proximity, none of these is present in an overlapping fashion. Hence, there is reason to suspect that *Cr-med-2* and *Cr-med-7* may not efficiently recruit SKN-1 (discussed below).

Conservation of Cr-med expression

Embryonic development of *C. remanei* has not been characterized in detail, but is superficially similar to that of *C. elegans* (Fig. 5). We therefore sought to determine if the molecular basis for mesendoderm specification was also conserved. Previous evidence suggested that *Ce-med-1,2* are transcribed at the 4-cell stage (Maduro et al., 2001). We confirmed *Ce-med-1,2* expression by in situ hybridization using an antisense *Ce-med-1* probe, which should detect both *med* genes due to their high degree of nucleotide similarity (Maduro et al., 2001). As expected, we detected nascent *Ce-med* mRNA transcripts in the EMS nucleus (Fig. 6A). To determine whether the *Cr-med* genes are similarly transcribed, we performed in situ hybridization on early *C. remanei* embryos using an antisense *Cr-med-3* probe. As the *Cr-med* genes share several regions of as high as 90% sequence identity (data not shown), this probe is capable of detecting all seven putative *Cr-med* genes simultaneously. We observed strong signal in the *C. remanei* EMS

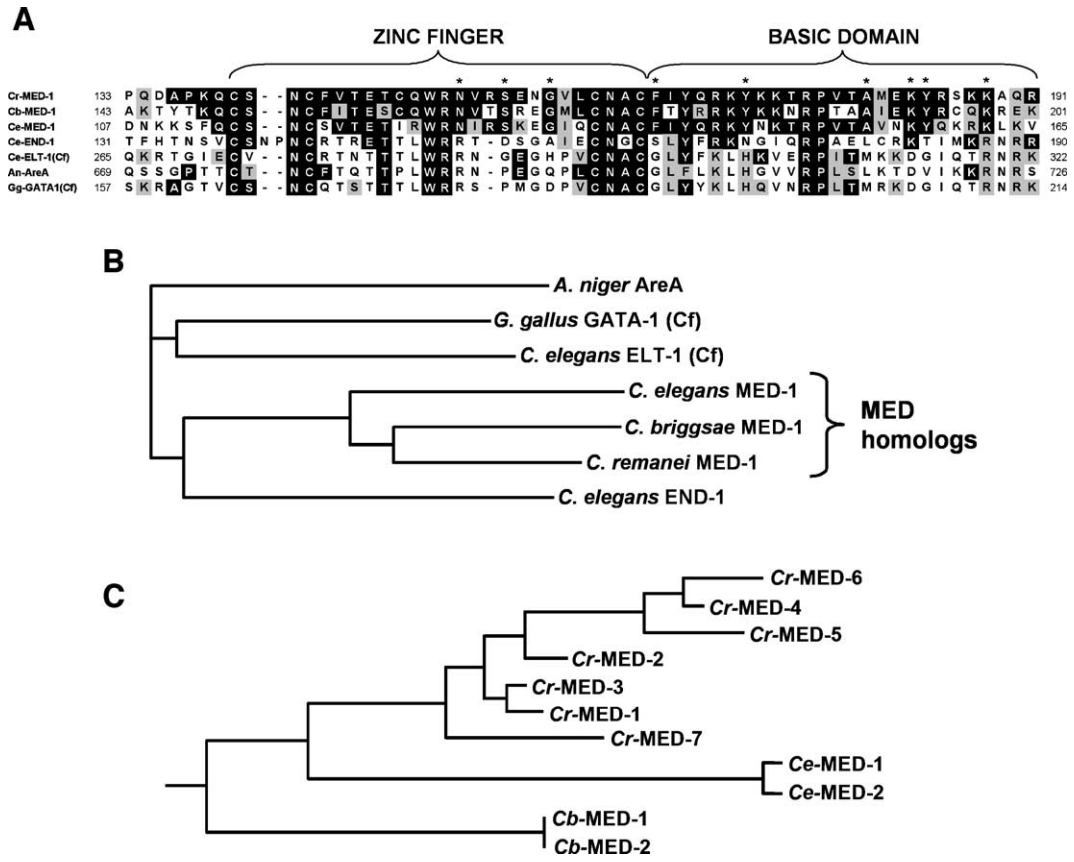


Fig. 4. Relatedness of the MED DNA binding domains to those of other GATA factors. (A) Alignment of DNA-binding domains (DBDs) of the MEDs, two other *C. elegans* GATA factors (END-1 and ELT-1; sequences from Wormbase, http://www.wormbase.org), a fungal GATA factor (*A. niger* AreA; Accession number CAA57524), and chicken GATA1 (*Gg*-GATA1; Accession number A32993). For the two-fingered proteins ELT-1 and GATA1, only the carboxyl finger (Cf) and basic domain are shown. (B) Relatedness tree generated by AlignX (Vector NTI 6, ClustalW algorithm) from the alignment shown in panel A, showing that the MED proteins form a subgroup diverged from other GATA factors. Note that this particular alignment (which includes only one MED DBD from each species) suggests that the homologs of *C. remanei* and *C. briggsae* are more closely related to each other than to *C. elegans* MED-1, consistent with their apparent evolutionary relationship (Cho et al., 2004; Kiontke et al., 2004). (C) In an alignment of the DBDs of all MEDs from all three species (rooted using *Ce*-END-1 and chicken GATA-1 as outgroups, not shown), the *Ce*-MEDs are found to be more closely related to the *Cr*-MEDs than the *Cb*-MEDs, and support is found for a single ancestral MED gene. The horizontal time scales in panels B and C are not identical.

blastomere (Fig. 6D). We did not detect expression in other embryonic stages (not shown). In parallel experiments with *Cb-med-1*, we observed expression in EMS in *C. briggsae* embryos (data not shown). Therefore, at least one *med* gene is transcribed in EMS in *C. remanei* and *C. briggsae.* To evaluate whether multiple *C. remanei med* genes are expressed, we performed RT-PCR on early embryo RNA. We found evidence that at least five of the *Cr-med* genes (*Cr-med-1*, *-3*, *-5*, *-6*, and *-7*) are expressed (Fig. 6G). Taken with the result that we did not see in situ hybridization outside the EMS cell, we conclude that at least five of the *C. remanei med* genes are expressed in the early embryo.

We next evaluated the ability of a *C. remanei med* gene reporter to express appropriately when introduced into *C. elegans*. A *C. elegans med-1*::GFP::MED-1 fusion (pMM280) expresses in the nucleus of E and MS and their daughters (Figs. 6B–C) (Maduro et al., 2001). We constructed a *Cr-med-1*::GFP::MED-1 construct, pMM808. *C. elegans* embryos transgenic for pMM808 as an extrachromosomal array (strain MS240) display an expression pattern indistinguishable from the *C. elegans* reporter (Figs. 6E, F). With both the *C. elegans* and *C. remanei* fusions, no significant expression was observed outside the early E and MS lineages (data not shown). In similar experiments, *C. elegans* embryos transgenic for a *C. briggsae med-1*::YFP fusion showed expression in the early MS and E lineages (not shown).

As embryonic activation of *Ce-med-1,2* is dependent on activity of *skn-1*, we tested whether this was true of the *Cr-med-1* reporter expressed in *C. elegans*. Such dependence is expected because of the presence of SKN-1 site clusters in the *Cr-med-1* promoter. Indeed, while 51% (n = 106) of untreated MS240 embryos at the 8- to 16-cell stage express the *Cr-med-1* reporter, expression was not detected in any similarly-staged embryos (n = 122; $P < 0.0001$) when *skn-1* was targeted by RNAi. We conclude that like *Ce-med-1*, a *Cr-med-1* reporter transgene is activated in a SKN-1-dependent manner when introduced into *C. elegans*.

Chromosomal Ce-med mutants show that the meds are essential genes

RNA interference (RNAi) can be used to produce entire broods of phenocopy mutants for many targeted genes (Fire et

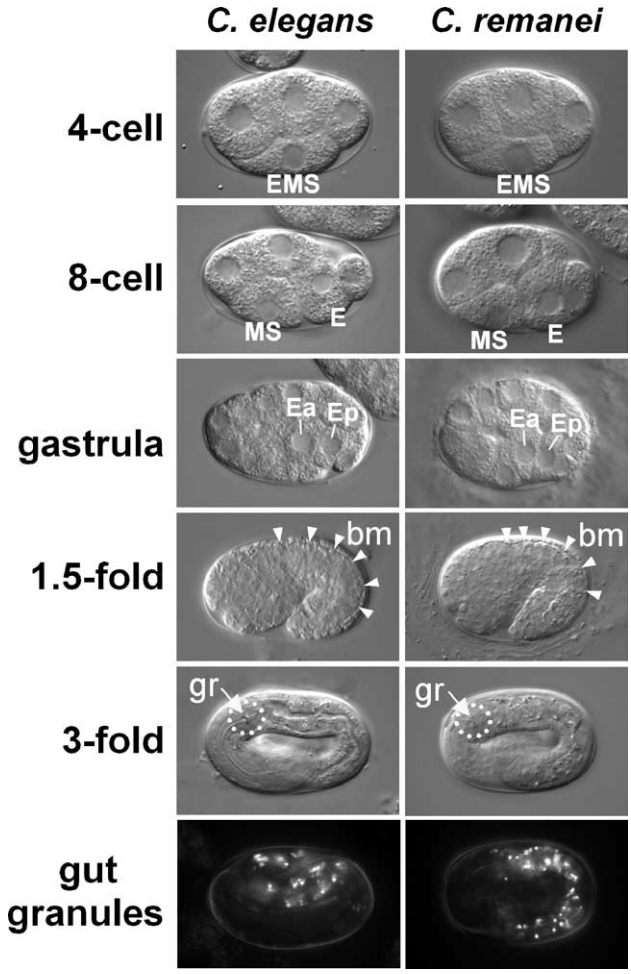


Fig. 5. *C. elegans* and *C. remanei* undergo similar early embryonic development. DIC images of living embryos were obtained at various stages. Early cleavages are asymmetric and establish a similar appearance of the size and position of early blastomeres at the 4-cell, 8-cell, and gastrula stages. At the 1.5-fold stage, a prominent basement membrane (bm) outlines the developing digestive tract. At the 3-fold stage, the posterior pharynx (made by MS) can be seen by the presence of the grinder (gr). Gut granules, as seen by polarized light, are indicative of differentiated endoderm. *C. elegans* and *C. remanei* embryos are approximately 50 μm along their long axis.

al., 1998). RNAi targeted to *Ce-med-1,2* produces a small number of arrested embryos that show defects in MS and E specification, along with many viable animals (Maduro et al., 2001). As the only means used to assess *med-1,2* function was RNAi, an absolute requirement for *med-1,2* function was not proven, nor was there a good means of assessing the *med* function of exogenous transgenes (i.e. by rescue). We constructed a strain, MS247, that generates *bona fide med-1,2(−)* mutant embryos. MS247 is homozygous for *med-1(ok804)*, a deletion removing the entire *Ce-med-1* coding region, and *med-2(cx9744)*, an insertion of a heterologous transposable element that is predicted to abrogate *Ce-med-2* function (see Fig. 8) (Granger et al., 2004). The strain is balanced by a large free duplication, *irDp1*, that provides *med* function and expresses an *unc-119*::YFP marker that enables identification of animals that have inherited it. All *med-1(ok804); med-2(cx9744)* embryos made by MS247 undergo embryonic arrest. These embryos lack posterior pharynx, which is made by MS, and a further 42% also lack endoderm (Figs. 7A–D). In isolation, >95% of single *med-1(ok804)* or *med-2(cx9744)* mutant embryos are completely viable. Hence, the *med* genes together are indeed essential for embryogenesis as originally proposed (Maduro et al., 2001), and the small number of mutants that result from *med-1,2(RNAi)* results from poor RNAi effectiveness with these genes.

We attempted to obtain *Cr-med(RNAi)* embryos by injecting high concentrations of *Cr-med-1* or *Cr-med-3* dsRNA directly into female gonad arms, which is expected to produce the strongest possible RNAi effects (Fire et al., 1998; Maduro et al., 2001; Montgomery, 2004). While several *med*-like progeny (many of which appeared to also lack endoderm) were observed, similar embryos were also found after injection of control GFP dsRNA and even in untreated animals (data not shown). The high background of arrested embryos (5–20%) in our *C. remanei* strain, coupled with the apparent low efficacy of RNAi targeted to the *Cr-med*s (as occurs in *C. elegans*), leaves us unable to correlate a specific phenotype with loss of *Cr-med* function.

Most of the Cr-med and Cb-med genes can substitute for Ce-med-1,2

As another approach to evaluate the functional conservation of the *Cr-med* genes, we tested the ability of the *Cr-med*s to complement the *C. elegans med-1,2(−)* mutant strain. We coinjected the *Cr-med-1*::GFP::MED-1 plasmid with the *rol-6^{D}* transformation marker (Mello et al., 1991) into MS247. Among the progeny were Rol animals that had lost both expression of the *irDp1*-associated *unc-119*::YFP marker, as well as an adult egg-laying defect associated with *irDp1*, suggesting that the rescuing function of *irDp1* was now being provided by an array containing *Cr-med-1*(+). Two lines were obtained that appeared to have lost *irDp1*, and which produced broods consisting of arrested Med embryos and viable embryos (Figs. 7E–F) that developed into fertile adults. Using PCR, we tested genomic DNA of one line (MS268) and confirmed that *irDp1* was no longer present (Fig. 8). This confirms that the *Cr-med-1*::GFP::MED-1 transgene is able to rescue *med-1(ok804); med-2(cx9744)* embryos to complete viability.

We next tested each of the *med* homologs for their ability to complement the lethality of *C. elegans med-1(ok804); med-2(cx9744)* embryos. For each homolog, plasmids or PCR products containing each gene were coinjected with an *unc-119*::CFP reporter into MS247. We note that as extrachromosomal arrays typically consist of many copies of the injected DNA (Mello et al., 1991), this type of rescue assay is at best semi-quantitative in nature. For all *med* homologs except *Cr-med-2* and *Cr-med-7*, we obtained lines that had lost the parent *irDp1* balancer and were now completely dependent upon the transgene.

Using the cotransformed *unc-119*::CFP marker, which can be detected at all stages after midembryogenesis, we quantified the degree of rescue by scoring the proportion of array-carrying animals that were restored to embryonic viability (Table 1). We found that while 97–98% of embryos carrying *Ce-med-1(+)* or

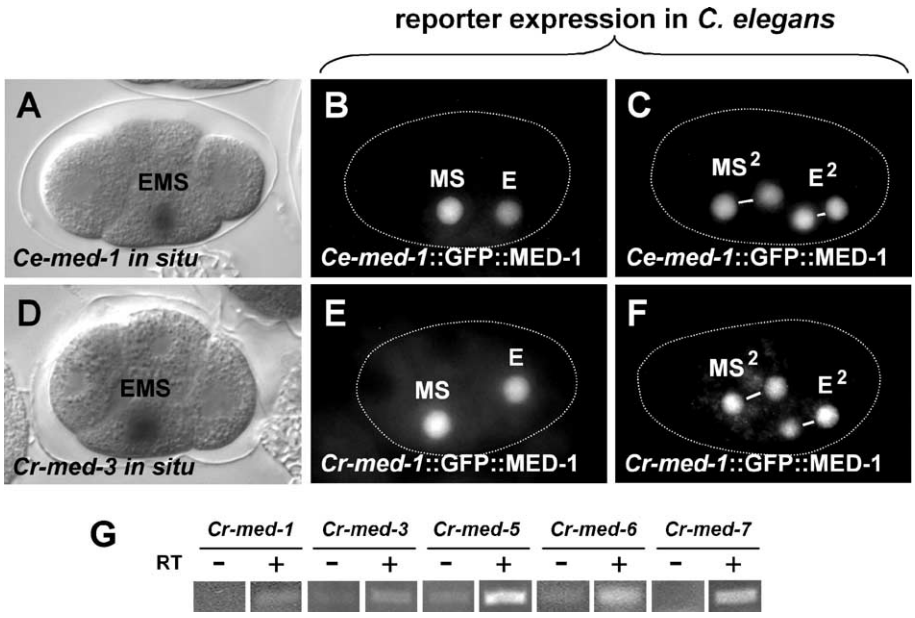


Fig. 6. Expression of *C. elegans* and *C. remanei med* genes is conserved. Top row: in situ hybridization of a 4-cell *C. elegans* embryo with a *Ce-med-1* probe reveals transcription in the nucleus of EMS (A). A *C. elegans med-1*::GFP::MED-1 translational fusion shows nuclear signal in the EMS daughters MS and E (B), and in their daughters (C). Bottom row: in situ hybridization of a *C. remanei* embryo with a *Cr-med-3* probe reveals transcription in the EMS nucleus (D). The nucleotide sequence similarity among the *C. remanei med* paralogs predicts that this probe will detect all *med* mRNAs simultaneously. A translational *Cr-med-1*::GFP::MED-1 fusion, introduced into *C. elegans*, displays a similar pattern of activation in E and MS (E) and their daughters (F). (G) Gene-specific RT-PCR products from early *C. remanei* embryos, showing detection of transcripts from five *Cr-med* genes (lanes marked '+'). Lanes corresponding to control reactions lacking reverse transcriptase are shown as '−'. Amplification was not observed for *Cr-med-2* and *Cr-med-4* (not shown).

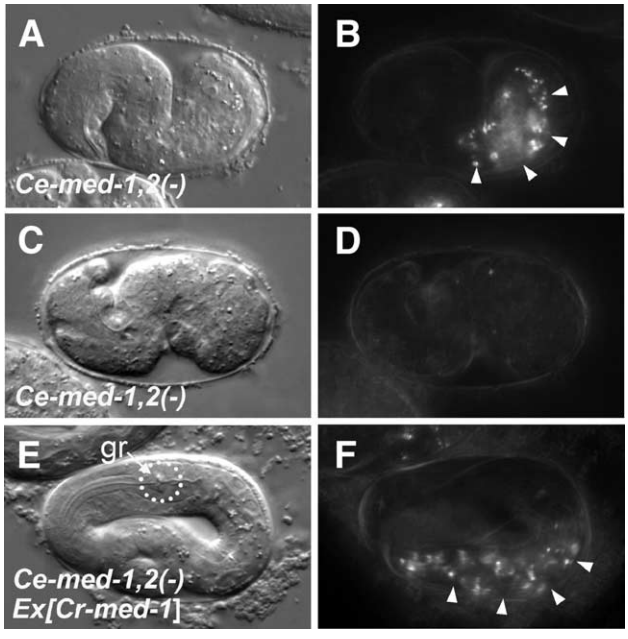
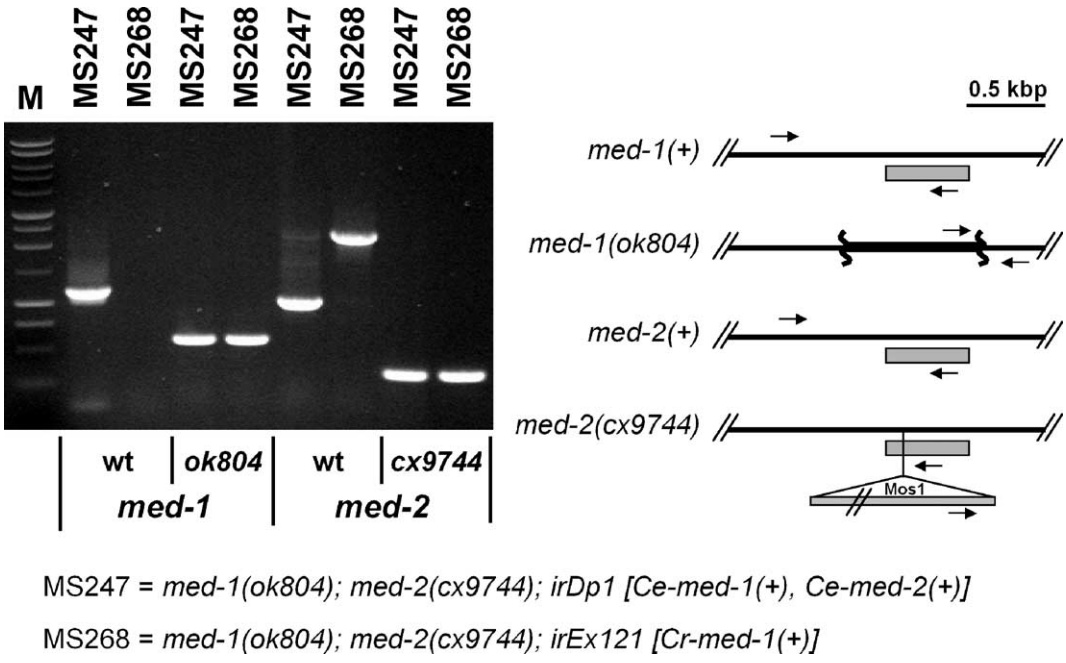


Fig. 7. The function of the MEDs is conserved between *C. elegans* and *C. remanei*. (A–D) *C. elegans med-1(ok804); med-2(cx9744)* mutant embryos, which fail to elongate beyond 2-fold and lack posterior pharynx as seen by Nomarski optics (A, C). 58% (n = 120) of *med-1,2(−)* mutants make endoderm (arrowheads in panel B), while the remaining 42% do not (D) as assayed by the presence of gut granules under polarized light. (E–F) *C. elegans med-1,2(−)* embryo rescued by a *Cr-med-1*::GFP::MED-1 transgene array. The animal has undergone complete elongation, posterior pharynx is present as evidenced by the prominent grinder (gr), and endoderm is present (arrowheads in panel J).

Ce-med-2(+) were rescued, the proportion of fully-rescued animals carrying an interspecific *med* transgene was generally lower. For all homologs except *Cr-med-2* and *-7*, a significant fraction of transgenic embryos (between 7 and 67%) was inviable. By Nomarski optics, many of these embryos elongated beyond two-fold, but had apparent defects in the morphology of the posterior pharynx, with some having only a slightly irregular grinder (data not shown), suggesting that a failure to completely specify MS (rather than E) accounts for the lack of complete rescue. Indeed, endoderm was made by a higher proportion of these arrested embryos (between 85% and 100%) than by arrested nontransgenic embryos (58%; $P < 0.01$ for all except *Cr-med-6*, which has a $P = 0.07$), consistent with a greater ability of the *med* homologs to rescue the partial E defect of *med-1,2(−)* embryos.

While the aforementioned *med* homologs rescued the vast majority of transgenic embryos, *Cb-med-2*, *Cr-med-2*, and *Cr-med-7* showed marked differences in their abilities to complement a *C. elegans med-1,2(−)* strain. First, while we were able to obtain a single rescuing line for *Cb-med-2*, the majority of transgenic embryos (67%) were inviable. This may be due only to the line itself, as one line (MS317) for *Cr-med-4* also showed 32% viability, while the other *Cr-med-4* line (MS393) showed 97% (the data in the table are pooled). Hence, we can make no strong conclusion about *Cb-med-2* function other than it is capable of rescue. Second, the genes *Cr-med-2* and *Cr-med-7* were unable to rescue to viability. Among arrested embryos, 61% (57/94) of *med-1,2(−); Ex[Cr-med-2]* embryos made endoderm, which is comparable to the 58% that did so in nontransgenic *med-1,2(−)* embryos (P = 0.7). In contrast, 90% (53/59) of *med-1,2(−); Ex[Cr-med-7]* embryos



MS247 = *med-1(ok804); med-2(cx9744); irDp1 [Ce-med-1(+), Ce-med-2(+)]*

MS268 = *med-1(ok804); med-2(cx9744); irEx121 [Cr-med-1(+)]*

Fig. 8. A *C. elegans med-1,2(−)* strain rescued by *Cr-med-1(+)* carries no intact *Ce-med* gene. The *ok804* allele of *med-1* deletes 2.5 kbp of genomic DNA, including the entire coding region and promoter of *med-1*, and contains a heterologous insertion of 0.8 kbp. The agarose gel shows results of PCR amplification of *C. elegans med-1* and *med-2* wild-type and mutant alleles from MS247, a strain rescued by *irDp1*, and MS268, a strain in which *irDp1* has been replaced by an extrachromosomal array containing *Cr-med-1(+)*. To the right of the gel, diagrams (approximately to scale) show the arrangement of PCR primers (arrows) relative to the *C. elegans med-1* and *med-2* genes. The gray rectangles directly beneath the DNA show the *med-1* or *med-2* coding regions. For *med-1(ok804)*, the heterologous insertion is shown by a thicker black line between the deletion breakpoints. For *med-2(cx9744)*, the heterologous Mos1 insertion (Granger et al., 2004) is shown as a large gray rectangle.

made endoderm ($P < 0.0001$). Hence, both *Cr-med-2* and *Cr-med-7* fail to rescue the MS specification defect, but where *Cr-med-2* also fails to rescue E specification, there is evidence that *Cr-med-7* can rescue E fate. Perhaps not surprisingly, these two genes are the very ones whose promoters differ from the other (rescuing) genes in that they lack clusters of *overlapping* SKN-1 sites as discussed earlier. The presence of a conserved, intact open reading frame in *Cr-med-7*, and its ability to complement the endoderm defect, suggest that when introduced into *C. elegans*, the *Cr*-MED-7 gene product is made from this gene (perhaps at a low level) and that it has at least partial function. In the case of *Cr-med-2*, the apparent amino truncation of the *Cr*-MED-2 coding region, and its inability to rescue either the MS or E defects in *C. elegans*, suggest that the transgene may either not be expressed, or that the short *Cr*-MED-2 gene product lacks an essential component of MED function.

Taken together with the expression data, we conclude that gene activation and function of the *med* genes have been conserved among *C. elegans*, *C. briggsae*, and *C. remanei*.

Discussion

In this study, we have shown that: (1) the *C. remanei* genome encodes seven structural homologs of *C. elegans med-1*; (2) the collective MED homologs predict functionally important amino acids in the MED zinc finger and basic domain; (3) expression of the *C. remanei* and *C. elegans med* genes occurs in the EMS nucleus; (4) a *Cr-med-1* translational GFP fusion introduced into *C. elegans* confers *med* function and similar expression as a *C. elegans* reporter; (5) five of the seven *C. remanei* homologs can complement the embryonic lethality associated with *C. elegans* embryos lacking *med-1,2* function; (6) expression and function of the *C. briggsae med* genes are similarly conserved. These results provide strong evidence that specification of mesendoderm in these species occurs by similar mechanisms.

Gene duplications among med genes are common

Comparison of the MED amino acid sequences suggests that all intraspecific paralogs arose from a single ancestral gene. The occurrence of seven highly similar *Cr-med* genes seems, therefore, to represent an extreme case of redundancy arising from gene duplication. In *C. elegans* and *C. briggsae*, the *med* genes exist in only two copies. Many of the non-MED GATA factors are also found in two to three paralogous copies (Maduro and Rothman, 2002). *Ce-end-1* and *Ce-end-3* are linked paralogs that function together to specify endoderm (E) fate (Maduro and Rothman, 2002; Maduro et al., 2005a). In *C. briggsae*, there is a single *Cb-end-1* homolog and two *Cb-end-3* homologs (Maduro et al., 2005a). Within *C. elegans*, other paralogous pairs are known: *Ce-elt-2* and *Ce-elt-7* (Maduro and Rothman, 2002) and *Ce-egl-18* and *Ce-elt-6* (Koh and Rothman, 2001). Our preliminary examination of the *C. remanei* genome sequence suggests that with the exception of the *med*s, a possible duplicate of *elt-1*, and the absence of the small *elt-4* gene, the remaining six *C. elegans* GATA factors (*Ce-elt-2*, *Ce-elt-3*, *Ce-egl-18*, *Ce-elt-6*, *Ce-end-1*, and

Ce-end-3) are each represented by a single homolog in *C. remanei* (data not shown).

Why might so many copies of *C. remanei med* genes have been maintained through evolution? One hypothesis is that these genes are in the process of acquiring new, essential functions that have caused the duplicated copies to be retained. However, the high degree of sequence relatedness, both within and outside the coding regions, suggests that most of the *Cr-med* genes share similar expression and function, as they are clearly recent duplicates. That the *Cr-med* genes have retained similar activity is substantiated by the ability of most to complement a *C. elegans* strain lacking endogenous *med* function. We did not detect *med* expression in postgastrulation *C. remanei* embryos, nor in similar *C. elegans* embryos expressing a *Cr-med-1* reporter, suggesting that the *Cr-med*s function only in MS and E specification. However, we cannot rule out additional functions of the *Cr*-MED proteins or postembryonic expression.

An alternative explanation is that the *med* genes have been maintained by a mechanism that selects for their additive contribution to a single function. If a threshold level of Med activity is critical for robust MS and E specification, gene duplication might have arisen as a means of compensating for decreased activation of individual *med* genes. If the mechanism that activates the *med*s became compromised in an ancestral species (e.g. from a reduction in number of SKN-1 sites in a single ancestral *med* gene), this might create a strong selection for additional Med activity arising from gene amplification. At least in *C. elegans*, the limiting step in *med* activation is not SKN-1 activity: in situ hybridization of embryos carrying a multicopy *Ce-med-1* transgene displays manyfold higher accumulation of *med* mRNA than wild-types (not shown), suggesting that expression of *med* genes is proportional to gene dosage. Hence, gene amplification may simply be a route to increase *med* expression, and assure that high amounts of MED activity, with the correct kinetics, are produced rapidly during a critical developmental time point.

Is Cr-med-2 a pseudogene?

The failure of *Cr-med-2* to complement a *C. elegans med-1,2(−)* double mutant, and the absence of a significant number of SKN-1 sites in its putative promoter, suggest that *Cr-med-2* may be a pseudogene. A similar situation exists with the *C. elegans* GATA factor gene *elt-2*: A 5′ truncated coding region, called *elt-4*, is found several kb upstream of its apparent full-size paralog, *elt-2* (Fukushige et al., 2003). ELT-2 is essential for the formation of the intestine and is known to function as a transcriptional activator (Fukushige et al., 1998, 2003). In contrast, while expression of *elt-4* has been demonstrated, there appears to be no phenotypic consequence to its deletion, nor can any biochemical function be assigned to the ELT-4 protein either in vitro or when expressed in yeast (Fukushige et al., 2003). There is precedent, therefore, for duplication and retention of a fragment of GATA family gene despite any detectable function. As none of our assays directly addressed function or expression of only *Cr-med-2* in *C. remanei*, a role for *Cr*-MED-2 distinct from early mesendoderm specification cannot be ruled out.

Differential requirements for the MEDs in MS and E fates

Our studies have shown that while the MED genes function in the specification of MS and E, there are different requirements for the MEDs in each of these sister cells. All embryos fail to specify MS in *C. elegans med-1,2(−)*, while only 42% fail to make endoderm. Clearly, the MEDs are critical for MS but dispensable much of the time for E fate. Indeed, we have shown that there are at least three other inputs into E specification: the Wnt effector POP-1, the Caudal-like homeodomain protein PAL-1, and maternal SKN-1 itself (Maduro and Rothman, 2002; Maduro et al., 2005b).

The critical requirement for MED function in MS rather than E is also apparent in the rescuing abilities of the various *med* homologs. With the exception of *Cr-med-2*, rescue of E occurred in more transgenic animals than did complete rescue of MS. As there are other inputs into endoderm specification, this is perhaps to be expected. Recent work has shown that the MED target genes in MS and E are largely nonoverlapping (that is, they function in MS or E but not both) (Broitman-Maduro et al., 2005). The ability of most of the *med* homologs to rescue the lethality of a *C. elegans med-1,2(−)* strain to viability implies that the entire suite of MED target genes can be activated by the *Cb*-MED and *Cr*-MED regulators (except *Cr*-MED-2/7). The reduced ability of these interspecific *med* homologs to fully rescue the MS defect suggests that there are at least some MS target genes that are exquisitely sensitive to stochastic differences in the levels or function of the transgenic MED proteins. Whether this arises from changes in recognition of *cis*-regulatory sites by the MEDs, or other aspects of MED expression and function, remains to be elucidated.

Why are all 11 med genes intronless?

In addition to the amino acid similarity and functional conservation, the *med* genes share one other characteristic, that of intronless coding regions. For the purposes of function, the presence of introns does not appear to be detrimental: both a *Cr-med-1*::GFP::MED-1 fusion and a *Ce-med-1*:GFP::MED-1 fusion can each rescue *Ce-med-1,2(−)* embryos, despite the presence of three introns in the GFP coding region (A. Fire, personal communication), and the fact that the inserted sequences result in a transgene mRNA that is approximately twice as long as the endogenous *med* transcripts. However, an effect operating at the level of a single gene in a chromosomal context, as opposed to a multicopy array, cannot be ruled out.

The *med*s appear to be unusual in their intronless gene structure, as such genes are comparatively rare in *C. elegans*. Among GATA factors in *C. elegans*, they are the only intronless ones. A recent study has determined that while 13% of annotated human genes and approximately 18% of *Drosophila* genes are intronless, only 3% of *C. elegans* genes lack introns (Sakharkar and Kangueane, 2004; Sakharkar et al., 2005). Some gene families, such as G-protein coupled

receptors (GPCRs) are encoded by primarily intronless genes, perhaps because many copies have arisen by retroposition (Gentles and Karlin, 1999; Roy and Gilbert, 2005); however, *C. elegans* appears to be the exception in this instance, as the majority of its GPCR genes do possess introns (Gentles and Karlin, 1999). The presence of conserved SKN-1 sites outside the *med* coding regions suggests that, with the exception of *Cr-med-2*, the *med*s have undergone amplification by a process that duplicates both the transcribed and flanking sequences, rather than a mechanism that requires an mRNA intermediate. Alternatively, a mechanism may exist that strongly selects against acquisition of introns in *med* genes. For example, gene conversion events among the *med*s could account for both the lack of introns and the apparent coevolution of *med* sequences within each species. Identification of more *med* homologs from other species, or demonstration of a gene conversion mechanism in *C. elegans*, may help resolve this question.

A conserved early mesendoderm network

The foregoing studies provide strong evidence that despite the divergence among the *C. elegans* and *C. remanei med* genes, they are activated in the same spatiotemporal pattern (i.e. immediately downstream of SKN-1), and most of these genes have retained the ability to activate the same sets of target genes as determined by inter-species rescue. Hence, while the *med* gene family has undergone much duplication in *C. remanei*, most of the paralogs still function in the context of a similar gene regulatory network that specifies the MS and E fates (Fig. 1). This suggests that despite tens of millions of years of evolution, the mechanism that specifies early mesendoderm fates in *Caenorhabditis* involves deployment of the same sets of genes. This is perhaps not that surprising, as more distant nematodes, such as *Halicephalobus* (Panagrolaimidae) and *Rhabditophanes* (Rhabditidae), show very similar early development as *C. elegans* in the establishment and position of the early founder cells, including MS and E (Borgonie et al., 2000). It will therefore be of interest to identify more *med* homologs from more distant species, and begin a directed approach to structure–function analysis of the MED class of zinc finger proteins.

Note added in proof

We have recently been made aware of two additional candidate *C. briggsae med*-like genes: Wormbase CDG00186 (an apparent pseudogene) and CDG15489.

Acknowledgments

We are very grateful to Laurent Ségalat for sending us the *med-2(cx9744)* allele; John Gatesy for help with phylogenetic trees; and our anonymous reviewers for insightful comments. Some nematode strains used in this work were provided by the *Caenorhabditis* Genetics Center, which is funded by the NIH National Center for Research Resources (NCRR). This work was supported by an NSF grant (IBN#0416922) to M.M.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

An, J.H., Blackwell, T.K., 2003. SKN-1 links *C. elegans* mesendodermal specification to a conserved oxidative stress response. Genes Dev. 17, 1882–1893.

Bei, Y., Hogan, J., Berkowitz, L.A., Soto, M., Rocheleau, C.E., Pang, K.M., Collins, J., Mello, C.C., 2002. SRC-1 and Wnt signaling act together to specify endoderm and to control cleavage orientation in early *C. elegans* embryos. Dev. Cell 3, 113–125.

Blackwell, T.K., Bowerman, B., Priess, J.R., 1994. Formation of a monomeric DNA binding domain by Skn-1 bZIP and homeodomain elements. Science 266, 621–628.

Blumenthal, T., Steward, K., 1997. RNA processing and gene structure. In: Riddle, D.L., Blumenthal, T., Meyer, B.J., Priess, J.R. (Eds.), *C. elegans* II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 117–146.

Borgonie, G., Jacobsen, K., Coomans, A., 2000. Embryonic lineage evolution in nematodes. Nematology 2, 65–69.

Bowerman, B., Eaton, B.A., Priess, J.R., 1992. skn-1, a maternally expressed gene required to specify the fate of ventral blastomeres in the early *C. elegans* embryo. Cell 68, 1061–1075.

Bowerman, B., Draper, B.W., Mello, C.C., Priess, J.R., 1993. The maternal gene skn-1 encodes a protein that is distributed unequally in early *C. elegans* embryos. Cell 74, 443–452.

Broitman-Maduro, G., Maduro, M.F., Rothman, J.H., 2005. The noncanonical binding site of the MED-1 GATA factor defines differentially regulated target genes in the *C. elegans* mesendoderm. Dev. Cell 8, 427–433.

Cho, S., Jin, S.W., Cohen, A., Ellis, R.E., 2004. A phylogeny of *Caenorhabditis* reveals frequent loss of introns during nematode evolution. Genome Res. 14, 1207–1220.

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

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C., 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391, 806–811.

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

Fukushige, T., Goszczynski, B., Tian, H., McGhee, J.D., 2003. The evolutionary duplication and probable demise of an endodermal GATA factor in *Caenorhabditis elegans*. Genetics 165, 575–588.

Gentles, A.J., Karlin, S., 1999. Why are human G-protein-coupled receptors predominantly intronless? Trends Genet. 15, 47–49.

Goldstein, B., 1992. Induction of gut in *Caenorhabditis elegans* embryos. Nature 357, 255–257.

Granger, L., Martin, E., Segalat, L., 2004. Mos as a tool for genome-wide insertional mutagenesis in *Caenorhabditis elegans*: results of a pilot study. Nucleic Acids Res. 32, e117.

Henikoff, S., Henikoff, J.G., 1992. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. U. S. A. 89, 10915–10919.

Kiontke, K., Gavin, N.P., Raynes, Y., Roehrig, C., Piano, F., Fitch, D.H., 2004. *Caenorhabditis* phylogeny predicts convergence of hermaphroditism and extensive intron loss. Proc. Natl. Acad. Sci. U. S. A. 101, 9003–9008.

Koh, K., Rothman, J.H., 2001. ELT-5 and ELT-6 are required continuously to regulate epidermal seam cell differentiation and cell fusion in *C. elegans*. Development 128, 2867–2880.

Levine, M., Davidson, E.H., 2005. From the cover. Gene regulatory networks special feature: gene regulatory networks for development. Proc. Natl. Acad. Sci. U. S. A. 102, 4936–4942.

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

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

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

Maduro, M.F., Meneghini, M.D., Bowerman, B., Broitman-Maduro, G., Rothman, J.H., 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.

Maduro, M.F., Lin, R., Rothman, J.H., 2002. Dynamics of a developmental switch: recursive intracellular and intranuclear redistribution of *Caenorhabditis elegans* POP-1 parallels Wnt-inhibited transcriptional repression. Dev. Biol. 248, 128–142.

Maduro, M.F., Hill, R.J., Heid, P.J., Newman-Smith, E.D., Zhu, J., Priess, J.R., Rothman, J.H., 2005a. Genetic redundancy in endoderm specification within the genus *Caenorhabditis*. Dev. Biol. 284, 509–522.

Maduro, M.F., Kasmir, J.J., Zhu, J., Rothman, J.H., 2005b. The Wnt effector POP-1 and the PAL-1/Caudal homeoprotein collaborate with SKN-1 to activate *C. elegans* endoderm development. Dev. Biol. 285, 510–523.

Mello, C.C., Kramer, J.M., Stinchcomb, D., Ambros, V., 1991. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. EMBO J. 10, 3959–3970.

Miller III, D.M., Desai, N.S., Hardin, D.C., Piston, D.W., Patterson, G.H., Fleenor, J., Xu, S., Fire, A. (1999). Two-color GFP expression system for *C. elegans*. Biotechniques 26, 914–918, 920–921.

Montgomery, M.K., 2004. The use of double-stranded RNA to knock down specific gene activity. Methods Mol. Biol. 260, 129–144.

Omichinski, J.G., Clore, G.M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S.J., Gronenborn, A.M., 1993. NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. Science 261, 438–446.

Pujol, N., Bonnerot, C., Ewbank, J.J., Kohara, Y., Thierry-Mieg, D., 2001. The *Caenorhabditis elegans* unc-32 gene encodes alternative forms of a vacuolar ATPase a subunit. J. Biol. Chem. 276, 11913–11921.

Rocheleau, C.E., Downs, W.D., Lin, R., Wittmann, C., Bei, Y., Cha, Y.H., Ali, M., Priess, J.R., Mello, C.C., 1997. Wnt signaling and an APC-related gene specify endoderm in early *C. elegans* embryos. Cell 90, 707–716.

Rocheleau, C.E., Yasuda, J., Shin, T.H., Lin, R., Sawa, H., Okano, H., Priess, J.R., Davis, R.J., Mello, C.C., 1999. WRM-1 activates the LIT-1 protein kinase to transduce anterior/posterior polarity signals in *C. elegans*. Cell 97, 717–726.

Roy, S.W., Gilbert, W., 2005. The pattern of intron loss. Proc. Natl. Acad. Sci. U. S. A. 102, 713–718.

Sakharkar, M.K., Kangueane, P., 2004. Genome SEGE: a database for 'intronless' genes in eukaryotic genomes. BMC Bioinformatics 5, 67.

Sakharkar, M.K., Chow, V.T., Ghosh, K., Chaturvedi, I., Lee, P.C., Bagavathi, S.P., Shapshak, P., Subbiah, S., Kangueane, P., 2005. Computational prediction of SEG (single exon gene) function in humans. Front. Biosci. 10, 1382–1395.

Seydoux, G., Fire, A., 1995. Whole-mount in situ hybridization for the detection of RNA in *Caenorhabditis elegans* embryos. Methods Cell Biol. 48, 323–337.

Shin, T.H., Yasuda, J., Rocheleau, C.E., Lin, R., Soto, M., Bei, Y., Davis, R.J., Mello, C.C., 1999. MOM-4, a MAP kinase kinase kinase-related protein, activates WRM-1/LIT-1 kinase to transduce anterior/posterior polarity signals in *C. elegans*. Mol. Cell 4, 275–280.

Starich, M.R., Wikstrom, M., Arst Jr., H.N., Clore, G.M., Gronenborn, A.M., 1998. 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.

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., Coulson, A., D'Eustachio, P., Fitch, D.H., Fulton, L.A., Fulton, R.E., Griffiths-Jones, S., Harris, T.W., Hillier, L.W., Kamath, R., Kuwabara, P.E., Mardis, E.R., Marra, M.A., Miner, T.L., Minx, P., Mullikin, J.C., Plumb, R.W., Rogers, J., Schein, J.E., Sohrmann, M., Spieth, J., Stajich, J.E., Wei, C., Willey, D., Wilson, R.K., Durbin, R., Waterston, R.H., 2003. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. PLoS Biol. 1, E45.

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

Thorpe, C.J., Schlesinger, A., Carter, J.C., Bowerman, B., 1997. Wnt signaling polarizes an early *C. elegans* blastomere to distinguish endoderm from mesoderm. Cell 90, 695–705.

Yuh, C.H., Brown, C.T., Livi, C.B., Rowen, L., Clarke, P.J., Davidson, E.H., 2002. Patchy interspecific sequence similarities efficiently identify positive *cis*-regulatory elements in the sea urchin. Dev. Biol. 246, 148–161.