Development 127, 4825-4835 (2000)
Printed in Great Britain © The Company of Biologists Limited
DEV7837

Evolutionary conservation of redundancy between a diverged pair of forkhead transcription factor homologues

Laurent Molin, Andrew Mounsey, Sobia Aslam, Petra Bauer, Jane Young, Martyn James, Archana Sharma-Oates and Ian A. Hope*

School of Biology, University of Leeds, Leeds LS2 9JT, UK
*Author for correspondence (e-mail: i.a.hope@leeds.ac.uk)

Accepted 29 August; published on WWW 24 October 2000

SUMMARY

The *Caenorhabditis elegans* gene *pes-1* encodes a transcription factor of the forkhead family and is expressed in specific cells of the early embryo. Despite these observations suggesting *pes-1* to have an important regulatory role in embryogenesis, inactivation of *pes-1* caused no apparent phenotype. This lack of phenotype is a consequence of genetic redundancy. Whereas a weak, transitory effect was observed upon disruption of just T14G12.4 (renamed *fkh-2*) gene function, simultaneous disruption of the activity of both *fkh-2* and *pes-1* resulted in a penetrant lethal phenotype. Sequence comparison suggests these two forkhead genes are not closely related and the functional association of *fkh-2* and *pes-1* was only explored because of the similarity of their expression patterns.

Conservation of the *fkh-2/pes-1* genetic redundancy between *C. elegans* and the related species *C. briggsae* was demonstrated. Interestingly the redundancy in *C. briggsae* is not as complete as in *C. elegans* and this could be explained by alterations of *pes-1* specific to the *C. briggsae* ancestry. With overlapping function retained on an evolutionary time-scale, genetic redundancy may be extensive and expression pattern data could, as here, have a crucial role in characterization of developmental processes.

Key words: *Caenorhabditis elegans, pes-1, sloppy-paired, Caenorhabditis briggsae*, Genetic redundancy

INTRODUCTION

The *C. elegans* genome is essentially completely sequenced and contains approximately 19,000 protein coding genes (*C. elegans* sequencing consortium, 1998). This figure is much higher than the number of essential genes predicted by classical genetic studies (Sulston et al., 1992; Waterston and Sulston, 1995; Johnsen and Baillie, 1997). Reverse genetic analysis by double-stranded RNA mediated interference (RNAi; Fire et al., 1998) seems to confirm that the majority of *C. elegans* genes fail to show any obvious phenotype when inactivated (P. Gönczy et al., personal communication). Similar conclusions have been reached for plants (Martienssen, 1998), yeast (Oliver et al., 1992), *Drosophila melanogaster* (Ashburner et al., 1999) and the mouse (Cooke, 1997) and so may be true for multicellular organisms in general.

There are various explanations for why a gene may appear non-essential, yet be retained within a genome. A gene may have a relatively minor role or may only be required under specific environmental conditions, such that inactivation would cause a phenotype difficult to detect in the laboratory. Alternatively, a gene may be redundant, in that, if inactivated other genes can perform the same function. Genetic redundancy of this type has been demonstrated experimentally (Johnson et al., 1981; Krause et al., 1989) and, theoretically, could be evolutionarily stable (Thomas, 1993; Nowak et al., 1997).

The *C. elegans* gene *pes-1* was originally identified through a molecular strategy, promoter trapping (Hope, 1991). *pes-1* encodes a transcription factor of the forkhead family and is expressed during embryogenesis in the descendants of several founder cells (Hope, 1994). *pes-1* expression in descendants of the AB blastomere is dependent on *glp-1* (Molin et al., 1999) and is first observed shortly after the *glp-1*-dependent signal to ABalp and ABara takes place. Reverse genetic analysis, described here, reveals that *pes-1* is not essential in embryogenesis but this is because of genetic redundancy and the gene does function during embryonic development. The redundancy involves two diverged members of the forkhead transcription factor family and is conserved in the related nematode *Caenorhabditis briggsae*. Such conservation demonstrates that redundancy can be maintained on an evolutionary time-scale, although the variation that was found may reveal how genetic redundancy can serve as a substrate for evolution.

MATERIALS AND METHODS

Strains and general methods

C. elegans and *C. briggsae* were cultured on agar plates by standard methods as described previously (Sulston and Hodgkin, 1988). Unless otherwise specified, standard protocols (Sambrook et al., 1989; Ausubel et al., 1993) were used for all molecular biology techniques.

Double-stranded RNA-mediated interference (RNAi)

RNAi was carried out according to the method of Fire et al. (1998). RNA was synthesized in vitro using T3 and T7 RNA polymerase and the templates listed below. For templates generated by PCR, T3 or T7 promoters were included at the 5′ end of the primers. Injected animals were placed onto individual plates and transferred to a new plate after 4 hours. All the progeny laid on this second plate, up until the parent was removed 24 hours later, were followed to observe the effects of RNAi.

ce-pes-1 template: a 650 bp DNA fragment containing the last four exons of the gene was generated by PCR from a *ce-pes-1* cDNA clone using the primers T3PES and T7PES.

ce-fkh-2 template: a 1.1 kb *Eco*RV-*Bgl*II DNA fragment from the T14G12 cosmid, containing the last three exons of the gene, was subcloned between the *Eco*RV and *Bam*HI sites of pBluescript (Stratagene).

cb-pes-1 template: a 560 bp *Xba*I-*Hin*dIII restriction fragment from a *cb-pes-1* cDNA clone (see below) was subcloned between the *Xba*I and *Hin*dIII sites of pBluescript (Stratagene).

cb-fkh-2 template: an 830 bp fragment, containing the second and third exons of the gene, was generated by PCR from genomic DNA in a worm lysate (Williams et al., 1992) using the primers T3T14 and T7T14.

Tc1 insertion and excision

All procedures were performed as described by Zwaal et al. (1993). Briefly, a library of MT3126 (*mut-2*), a strain of *C. elegans* with a high level of transposition, was screened by PCR for individuals with insertion of a Tc1 transposable element in the *pes-1* gene. A strain, homozygous for a Tc1 insertion in the first intron, was identified. A PCR-based screen for deletions in *pes-1*, arising from incomplete repair following excision of this Tc1 element in this strain, was performed and independent *pes-1* deletion alleles were recovered. PCR reactions were performed on worm lysates using primers 24C7.8 and 24C7.5, located 2.7 kb apart, 50 bp upstream of the first translation start site and 10 bp downstream of the stop codon, respectively.

Reporter gene fusions

ce-pes-1::gfp: a 5.1 kb *Pst*I-*Xma*I fragment from the plasmid pUL#24C7 (Hope, 1991), was subcloned between the *Pst*I and *Xma*I sites of pPD95.70 (Fire et al., 1990) to generate pUL#MJA1. The insert contains 3 kb of upstream sequence, with fusion to *gfp* within the fourth exon of *ce-pes-1*.

ce-fkh-2::gfp: this gene fusion was made using a PCR-based strategy (Hobert et al., 1999; A. Nathoo and A. Hart, personal communication). A 1.9 kb PCR product, containing the whole GFP coding sequence, was generated from the vector pPD95.67 (Fire et al., 1990), using the primers pPDGFP and GFPA. A 3 kb PCR product was generated from a T14G12 cosmid preparation using primers T14A and T14B. This product contains 2.95 kb of *fkh-2* upstream region and ends within the first exon of the gene. The last 21 nucleotides of the 3′ primer T14B are complementary to the first 21 nucleotides of the 1.9 kb GFP PCR product. A third, nested PCR, was performed from a mixture of the two initial PCR products using primers T14C (immediately downstream of T14A) and GFPB, to join the fragments together.

cb-pes-1::gfp: A 6.5 kb *Msc*I-*Nsi*I fragment from the BAC clone CB038P19 was subcloned between the *Msc*I and *Pst*I sites of pPD95.69 (Fire et al., 1990) to generate pUL#LM1. The resulting construct contains 1.5 kb of upstream sequence, with fusion to *gfp* within the sixth exon of *cb-pes-1*.

Plasmids or PCR products were co-injected into the germline syncytium of *C. elegans* or *C. briggsae* with pRF4 (Mello et al., 1991), a plasmid containing a dominant *rol-6* mutation, which confers a strong roller phenotype to transgenic animals. At least two, independent, transformed strains were examined for each reporter gene fusion. GFP expression was observed by fluorescence microscopy.

Cloning and sequencing *cb-pes-1*

The *cb-pes-1* gene was cloned using a *C. briggsae* genomic DNA gridded fosmid library provided by Genome Systems. An initial screen of this library, using an almost complete *ce-pes-1* cDNA as the probe, failed because of hybridization to repetitive DNA within the *C. briggsae* genome. A second screen of the library using only the region of the *ce-pes-1* gene encoding the forkhead domain as the probe, was successful. This probe was generated by PCR with the primers FHD3 and FHD5 (see below) using a *ce-pes-1* cDNA clone as the template. The most strongly recognized fosmid clones were covered by the BAC (Bacterial Artificial Chromosome) clone CB038P19. A 2.2 kb *Hin*dIII-*Pst*I restriction fragment from this BAC, to which the *ce-pes-1* probe hybridized, was subcloned and sequenced, confirming that this was likely to be the *C. briggsae* orthologue of *pes-1*. The 6.5 kb *Msc*I-*Nsi*I fragment from CB038P19, containing the promoter region of *cb-pes-1* used in a *gfp* reporter gene fusion, was identified in a Southern hybridization with a probe generated by PCR using primers CBPESA and CBPESB.

C. briggsae cDNA was prepared from 3 μg of total RNA isolated from a mixed-stage *C. briggsae* population using Trizol Reagent (GIBCO BRL). Reverse transcription was performed with the first-strand synthesis kit from Pharmacia, using primer 5A to obtain the 5′ end of the *cb-pes-1* mRNA and a dT17-adaptor primer to obtain the 3′ end. The 5′ end of *cb-pes-1* cDNAs was amplified by nested PCR using an SL1 *trans*-spliced leader oligonucleotide (SL1) as the forward primer, and primers 5A and 5B successively as the reverse primers. The nested PCR generated three *cb-pes-1*-specific products of 566, 466 and 220 bp. Isolation of the 3′ end of *cb-pes-1* cDNA was attempted by nested PCR using the adaptor primer as the reverse primer and primers 3A and 3B successively as the forward primers. This nested PCR generated one *cb-pes-1*-specific product of 253 bp that lacked the last 102 nucleotides of the transcript, as compared to the predicted splicing pattern, and did not include a stop codon. Amplification of such a product, missing the 3′ end of the transcript, is probably due to mis-annealing of the dT17-adaptor primer to an A-rich region upstream of the stop codon. All PCR products were cloned into pCR2.1-TOPO vector (Invitrogen) and sequenced. The *C. briggsae* cDNA sequences have been deposited in GenBank under the accession numbers AF260299, AF260300, AF260301 and AF260302.

45 kb of genomic DNA of the BAC clone CB038P19 has since been sequenced by the Washington University Genome Sequencing Center and is available at ftp://genome.wustl.edu/pub/gscl/sequence/st.louis/briggsae/.

Primers

T3PES: AATTAACCCTCACTAAAGGGGCTTCAACATCTCGGACTTG
T7PES: TAATACGACTCACTATAGGGAACCATGGGGATATTCTGG
T3T14: AATTAACCCTCACTAAAGGCTTATCTGCAGATATCAACG
T7T14: TAATACGACTCACTATAGGGCGGGTTACTGTAGTTTAGC
24C7.8: TAATATTCCGCAGTCGGCTTTC
24C7.5: GTCGGTCGACAAAAACTCAGAAGGCTATTC
pPDGFP: GCTTGCATGCCTGCAGGTCG
GFPA: AAGGGCCCGTACGGCCGACTAGTAGG
GFPB: GGAAACAGTTATGTTTGGTATATTGGG
T14A: GTTTTCTCATAAGATCGCCG
T14B: CGACCTGCAGGCATGCAAGCTGTTGTTCAATCTTGACCGCC
T14C: GTTGGATCCATTGGATTATG
FHD3: GCCACTTTGCTTTTTTGGC
FHD5: GAATCACCAACCAAAAGACC

CBPESA: GCGCCTGAGAGCATTATTTTC
CBPESB: CGCGACTTTCTTACCGGAC
5A: CCAGAAACTTCCTTTTCCATCC
5B: GCGACAAATTATGACGAATAG
SL1: GGTTTAATTACCCAAGTTTGAG
dT17-adaptor primer: GACTCGAGTCGACATCTTTTTTTTTTTTT-TTTT
adaptor primer: GACTCGAGTCGACATCG
3A: GGATGGAAAAGGAAGTTTCTGG
3B: CAGTTAGGATCCGACGTG

Sequence analysis

Sequences were compared using the program ALIGN (http://vega.igh.cnrs.fr/bin/align-guess.cgi). The conserved region between *ce-pes-1* and *cb-pes-1*, outside the coding regions, was identified using the dot-plot function of MacVector 6.5. Gene splicing pattern predictions are based on FGENESH 1.0 (http://genomic.sanger.ac.uk/gf/gf.html), using the *C. elegans* option. Sequence similarities were found using the BLAST program (http://www2.ncbi.nlm.nih.gov/BLAST/).

RESULTS

Inactivation of *pes-1* does not cause any apparent phenotype

The gene's expression pattern, in specific cells in the early embryo, and the nature of the gene product, a transcription factor, suggested *pes-1* would have a role in controlling differential gene expression during embryogenesis. To explore the function of *pes-1*, the gene was inactivated by injecting *pes-1*-specific double-stranded RNA (dsRNA) into wild-type *C. elegans* adults, a technique known as RNAi (Fire et al., 1998). Almost all the embryos laid by these injected animals developed normally and gave rise to progeny with general morphology and behaviour apparently identical to the wild type (Table 1). The absence of an altered phenotype from RNAi suggested *pes-1* was not required for proper embryonic development. However, the absence of an altered phenotype from RNAi can result from insufficient, or lack of, inactivation of the targeted gene (Tabara et al., 1998; Tavernarakis et al., 2000). Therefore, a genetic deletion in *pes-1* was sought.

A Tc1 transposable element insertion into the first intron of *pes-1* was obtained. A PCR-based strategy was then used to screen for imprecise excisions, of this Tc1 insertion, that resulted in deletions of the *pes-1* gene. Four *pes-1* deletion alleles were isolated, and one of these, designated *leDf1*, was characterized further (Fig. 1A). *leDf1* removed 58 amino acids of the 100 amino acid forkhead domain and many of the deleted amino acids are known to be essential for forkhead domain DNA binding activity (Clevidence et al., 1993; Hacker et al., 1995). The *leDf1* deletion will have inactivated the gene. After backcrossing to wild type, no altered phenotype was apparent for *leDf1*. Indeed, no altered phenotype was apparent for the other three deletions also known to have removed varying extents of the protein coding region of *pes-1*. Growth rate at 15°C, 20°C and 25°C, brood size and male mating efficiency were apparently unaffected.

Different explanations can be advanced for the failure to detect an altered phenotype for the *pes-1* deletion mutant. The gene could be a non-functional pseudogene, but all aspects of the molecular characterization suggest *pes-1* is fully functional (Hope, 1994). Alternatively, the phenotype could be subtle or undetectable under laboratory conditions. However, this would not be expected for disruption of the function of a transcription factor, expressed in specific cells during early embryogenesis. Finally, genetic redundancy with another gene or regulatory system could explain why *pes-1* does not appear to be required for embryogenesis. Acceptance of this latter explanation would require identification of genes functionally redundant with *pes-1*.

Functional overlap between T14G12.4/*fkh-2* and *pes-1*

Genetic redundancy could be simply the consequence of an evolutionarily recent gene duplication event. A screen by degenerate PCR for such a copy of *pes-1* was unsuccessful (Messom, 1996). With the essentially complete *C. elegans* genome sequence now available a close homologue of *pes-1* is still not apparent. In addition to *pes-1* and the four forkhead transcription factor genes detected previously by forward genetic approaches (*daf-16*, Riddle et al., 1981, Ogg et al., 1997; *lin-31*, Miller et al., 1993; *pha-4*, Mango et al., 1994, Kalb et al., 1998 and *unc-130*, B. Nash, personal communication), the *C. elegans* genome sequence has revealed eleven other forkhead transcription factor genes. Comparison

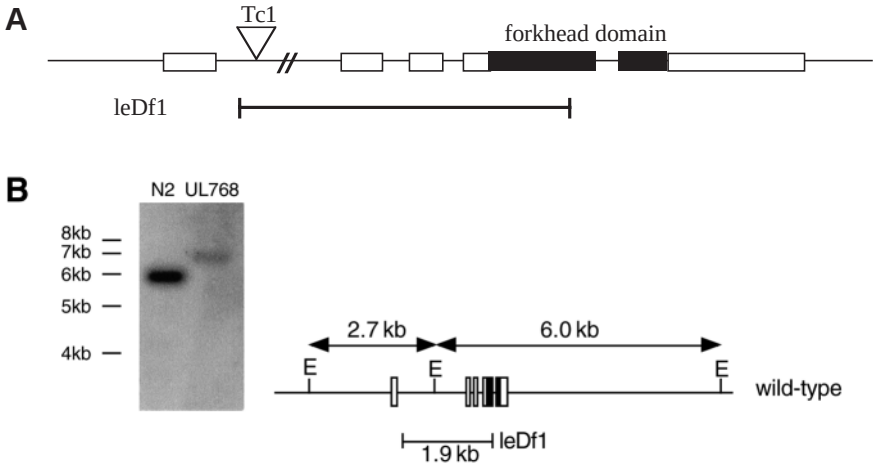

Fig. 1. Characterization of a Tc1-dependent deletion within *pes-1*. (A) The *pes-1* gene is represented by a horizontal line with exons as boxes and the region encoding the forkhead domain filled in. The Tc1 insertion is in the first intron of *pes-1*. The first intron is large and not drawn to scale, as indicated by the diagonal cross lines. The extent of the characterized deletion, *leDf1*, in strain UL768, obtained by imprecise excision of the Tc1 element, is indicated beneath. (B) Southern blot analysis of UL768. N2 (wild type) and UL768 genomic DNA, digested with *Eco*RI, was probed with a DNA fragment encoding the forkhead domain of PES-1. This probe detects a 6 kb fragment in N2. The *Eco*RI site (E), within the gene, is deleted in *pes-1(leDf1)* such that a larger 6.8 kb fragment is detected in UL768.

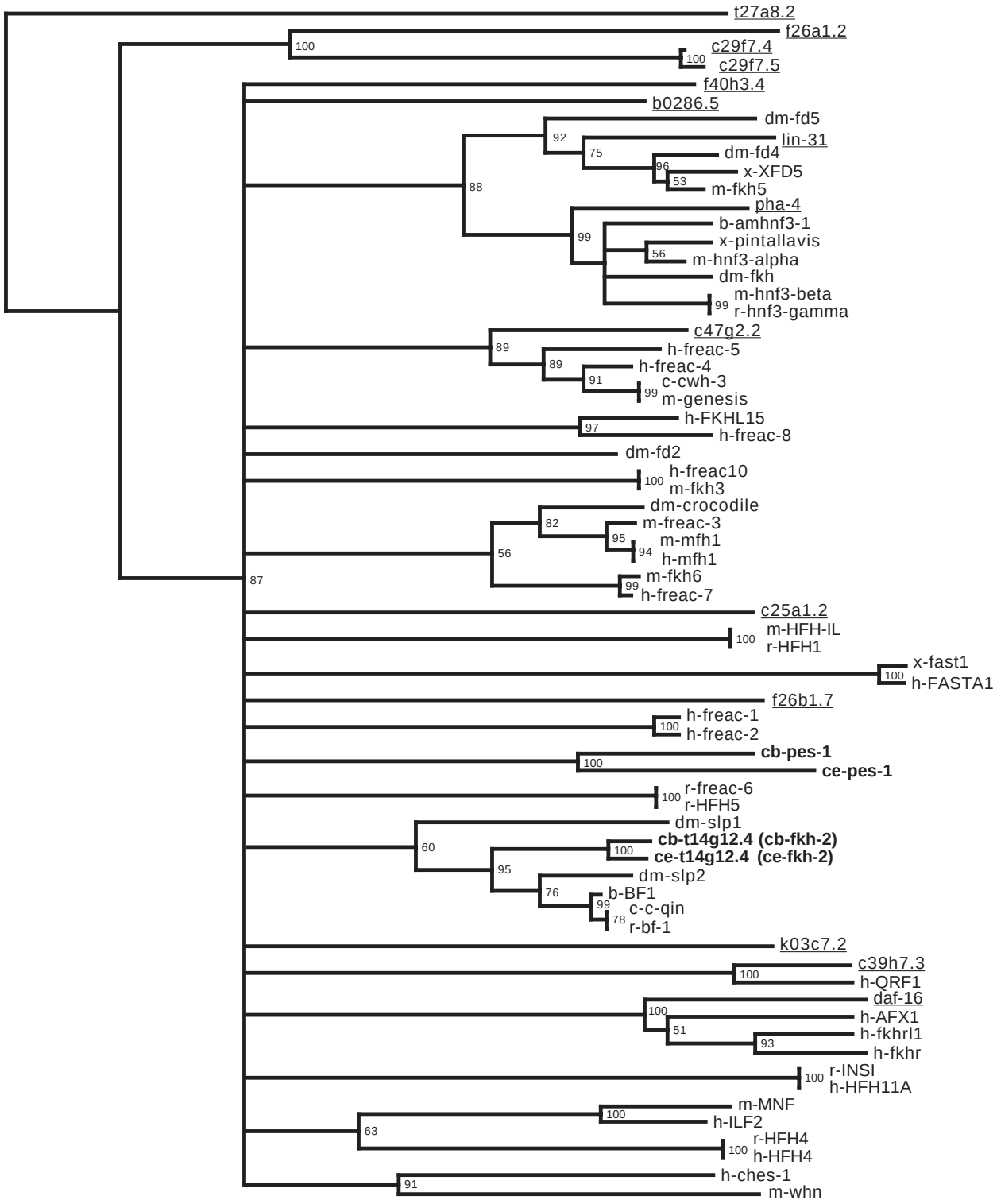

Fig. 2. Results of a phylogenetic analysis of forkhead genes from *C. elegans* and other species. The genes analysed are as in Ruvkun and Hobert (1998) but with added members of missing groups as defined by Kaestner et al. (2000). The groups obtained here are similar to those described by Kaestner et al. (2000). All *C. elegans* forkhead genes are underlined. T14G12.4 (*fkh-2*) and *pes-1* from *C. elegans* and *C. briggsae* are in bold. The forkhead domains (100 amino acids) of each protein were aligned with ClustalX version 1.4b (Thompson et al., 1997) using default parameters. Phylogenetic analyses were performed using a neighbour joining program implemented in the PHYLO_WIN software version 1.2 (Galtier et al., 1996). Bootstrap analysis was performed with 500 replicates to determine confidence values on the clades within the tree. Branches with bootstrap values less than 50 were collapsed using the TreeView software version 1.5 (Page, 1996).

of the protein sequences by phylogenetic analysis did not identify any as particularly closely related to *pes-1*, as would be expected for a recent gene duplication event (Fig. 2).

Expression patterns have been determined for all of the *C. elegans* forkhead transcription factor genes (data not shown). Amongst these, the expression pattern directed by a T14G12.4*::gfp* reporter gene fusion presents similarities with that directed by a *pes-1::gfp* reporter gene fusion (Fig. 3). Identical expression was observed in the descendants of the D founder cell. T14G12.4*::gfp* expression was also detected in many other cells more anteriorly that probably overlap with, but are not identical to, those of the AB founder cell lineage that express the *pes-1::gfp* fusion gene. Lineage analysis will be necessary to determine precisely which embryonic cells express the T14G12.4*::gfp* fusion gene. The similarity of expression pattern between *pes-1* and T14G12.4 prompted us to determine whether these two forkhead genes are, at least partially, redundant in function. T14G12.4 will be referred to as *fkh-2* (forkhead) from now on.

RNA interference was used to inactivate *fkh-2* either alone or in combination with *pes-1* (Table 1, rows 1-5). When dsRNA specific to *fkh-2* was injected into wild-type *C. elegans* adults their progeny were completely viable. The only noticeable phenotype is the slow and spatially restricted movement of about 30% of the L1 larvae, which nonetheless grow normally and do not display any obvious phenotype as late larvae or adults. *fkh-2*, like *pes-1*, does not appear to be essential for embryogenesis.

In contrast, co-injection of dsRNAs specific to both *fkh-2* and *pes-1* into wild-type adults had a pronounced effect. Twelve percent of eggs produced from the injected hermaphrodites during the selected period, arrested at various late stages of embryogenesis and 81% arrested after hatching as first stage larvae. The rest of the progeny (7%) escaped arrest and developed into sexually mature adults. Injection of dsRNA specific to *fkh-2* into hermaphrodites with the *pes-1* genetic deletion gave a very similar result (Table 1, row 5) suggesting that the *pes-1* RNAi is equivalent to complete inactivation of the gene.

These results reveal that the embryonic functions of *fkh-2* and *pes-1* overlap in *C. elegans* such that either is sufficient for embryogenesis. Genetic redundancy does indeed explain the lack of an altered phenotype upon inactivation of *pes-1*. The functional overlap between these two genes, for which the sequence does not suggest a particularly close relationship, raises questions about the evolutionary stability of this organization.

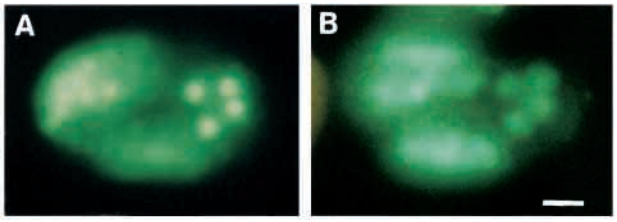

Fig. 3. Comparison of expression patterns for *pes-1* and T14G12.4 (*fkh-2*) reporter gene fusions. The *C. elegans* embryos have approximately 185 cells and are orientated with anterior to the left. *pes-1::gfp* (A) and T14G12.4*::gfp* (B) reporter gene fusions are both expressed in the 4 descendants of the D blastomere, toward the posterior. Expression of T14G12.4::GFP is weaker and not as restricted to the nucleus as PES-1::GFP. Expression is also observed more anteriorly but these expressing cells do not seem to be identical for the two genes. Scale bar in B, 10 μm.

pes-1 and *fkh-2* have homologues in *Caenorhabditis briggsae*

The two, morphologically very similar species, *Caenorhabditis elegans* and *Caenorhabditis briggsae*, diverged about 40 million years ago (Emmons et al., 1979; Kennedy et al., 1993). This evolutionary distance is considered sufficient for only functional genetic elements to have been retained by both species (Prasad and Baillie, 1989). To explore the evolutionary conservation of this redundant gene pair, the homologues of *pes-1* and *fkh-2* were cloned from *C. briggsae*.

The *C. briggsae* homologue of *fkh-2* (*cb-fkh-2*) was identified by searching the *C. briggsae* DNA database using the *C. elegans* gene (*ce-fkh-2*) sequence as the query. The only difference between the *C. elegans* and *C. briggsae* genes is the absence of the second intron of *ce-fkh-2* in *cb-fkh-2* (Fig. 4A). The two proteins are 76% identical in total with 96% identity within the forkhead domain (Fig. 4B). These figures are typical for comparisons of *C. elegans*/*C. briggsae* homologues (De Bono and Hodgkin, 1996) and contrast with those for *pes-1*.

The *C. briggsae* homologue of *pes-1* (*cb-pes-1*) was

Table 1. Phenotypic analysis of inactivation of *pes-1*, *fkh-2* or both *pes-1* and *fkh-2*, in *C. elegans* and *C. briggsae*

Row	Strain	dsRNA injected	Number of individuals examined‡	% embryonic arrest	% L1 arrest	% individuals that reach adulthood
1	Wild type *C. elegans*	none*	479	0.6	0	99.4
2	Wild type *C. elegans*	*ce-pes-1*	225	0.4	0	99.6
3	Wild type *C. elegans*	*ce-fkh-2*	1195	1.2	0	98.8
4	Wild type *C. elegans*	*ce-pes-1* + *ce-fkh-2*	242	12.0	81.4	6.6
5	UL768 *(pes-1(leDf1))*	*ce-fkh-2*	464	13.8	74.3	11.9
6	Wild type *C. briggsae*	none*	303	0.7	0	99.3
7	Wild type *C. briggsae*	*cb-pes-1*	332	0.6	0	99.4
8	Wild type *C. briggsae*	*cb-fkh-2*	830	1.9	46.3	51.8
9	Wild type *C. briggsae*	*cb-pes-1* + *cb-fkh-2*	461	8.7	77.6	13.7

*Animals were injected with water instead of a dsRNA solution.

‡The number of individuals examined was determined by counting the number of eggs and L1 larvae present when the injected parent was removed after 24 hours on the assay plate. The proportion that arrested as embryos was determined by counting the unhatched eggs remaining after a further 24 hour incubation. The percentage that failed to arrest was determined from the number of L4 larvae removed each day over a four day period. The arrested L1 larvae die at different times and so were difficult to count. Therefore the number of arrested L1 larvae was calculated as the brood size minus the number of dead eggs and the total number of animals that escaped arrest

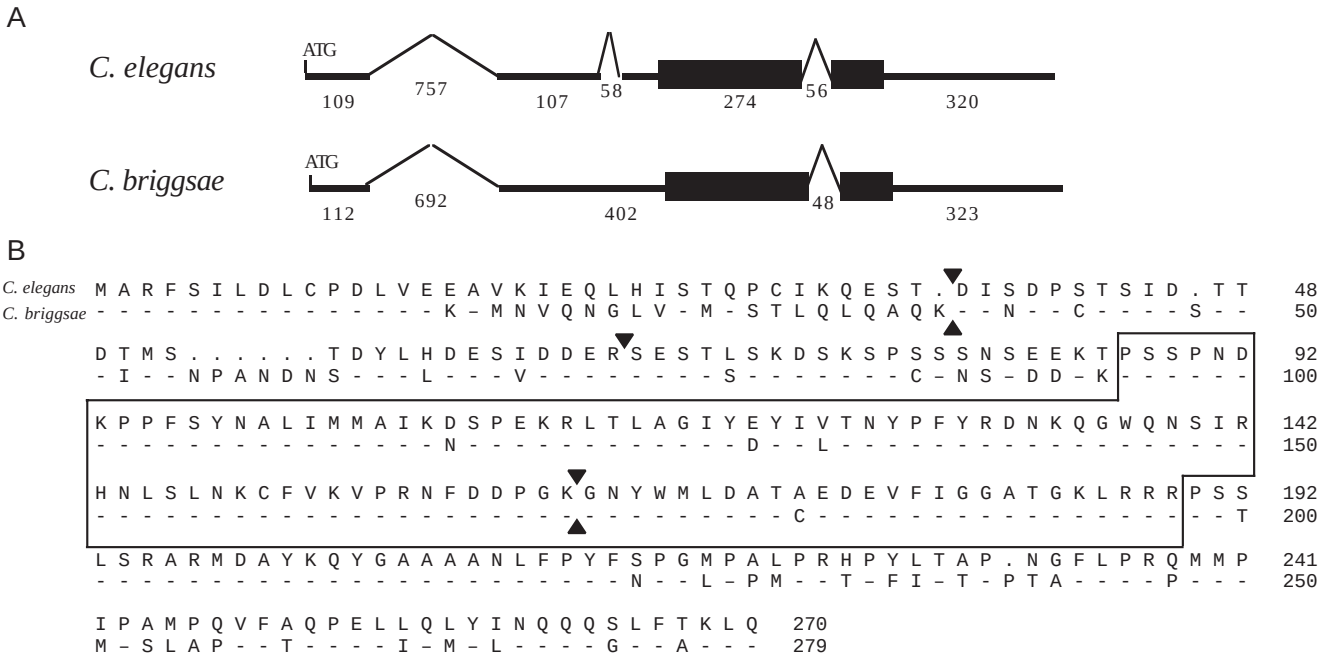

Fig. 4. (A) *fkh-2* gene structure in *C. elegans* and *C. briggsae* based on available ESTs in ACeDB, comparison between the two species and FGENESH splicing pattern predictions. Exons are represented by thick lines. The forkhead domain-encoding region is indicated by black boxes in each gene. The potential sites for translation initiation are indicated by ATG. Figures below the exons and introns indicate the size in nucleotides. (B) Sequence alignment of FKH-2 in *C. elegans* and *C. briggsae*. The forkhead domain is boxed. Dashes represent conserved amino acids in the two species, dots indicate gaps in aligned sequences. Black triangles show the locations of introns.

obtained by screening a gridded *C. briggsae* genomic DNA library with a *C. elegans pes-1* (*ce-pes-1*) cDNA as a probe and transcripts from the gene have been analysed.

ce-pes-1 and *cb-pes-1* are more diverged than expected. The gene structures are similar (Fig. 5) except that in *cb-pes-1* there are two small extra exons and two of the introns are substantially larger. Curiously, three *cb-pes-1* mRNAs were identified, *trans*-spliced to SL1, in precisely the same arrangement as for *ce-pes-1*: the larger transcripts begin with the first and second exons respectively and the smallest transcript starts with the first exon encoding the forkhead domain. In both species, the smallest transcript would lack an appropriate initiation codon, suggesting that this transcript, even though conserved, is not functional. Both the two larger transcripts have the first initiation codon in the appropriate reading frame and are presumed functional, encoding proteins with or without an N-terminal extension, in both species, although the sequence of that extension does not appear to be conserved.

C. elegans and *C. briggsae* PES-1 proteins, however, share only 44% identity in total and 69% identity within the forkhead domain (Fig. 6). The level of identity is sufficient to be confident that these two genes are orthologous (further evidence below) but sequence homology outside the forkhead domain is hard to detect. The preponderence of serine residues N-terminal to the forkhead domain and of proline, glutamine and asparagine residues C-terminal to the forkhead domain is conserved. These residues could have a role in the regulation of transcription of target genes (Mitchell and Tjian, 1989).

The expression pattern of *cb-pes-1* was investigated. *C. briggsae* transformed with a *cb-pes-1::gfp* gene fusion produced a GFP expression pattern that is very similar, if not identical, to the one directed by a *ce-pes-1::gfp* gene fusion in *C. elegans* (Fig. 7). All the components appear to have been conserved. Thus, although the gene sequence has diverged more than expected, the pattern of gene expression has been retained by the two species.

Interestingly, although the *pes-1* expression pattern may be conserved, the mechanism by which the expression pattern is generated may have been modified slightly through evolution. Expression patterns obtained in reciprocal reporter gene fusion experiments (*cb-pes-1::gfp* into *C. elegans* and *ce-pes-1::gfp* into *C. briggsae*) were significantly weaker than those obtained in the non-reciprocal experiments and included a few additional components (Fig. 7). For *C. briggsae* transformed with twelve other *C. elegans* gene*::gfp* fusions the expression patterns were all conserved both in strength and distribution (data not shown) and no weakening of expression was detected.

A more limited redundancy between *pes-1* and *fkh-2* is observed in *C. briggsae*

The possibility that *cb-pes-1* and *cb-fkh-2* are redundant, as in *C. elegans*, was addressed. Double-stranded RNA specific to *cb-pes-1* was injected into wild-type *C. briggsae* adults (Table 1, row 7). Embryos laid by injected animals developed normally and gave rise to animals with general morphology and behaviour that could not be distinguished from wild-type animals. *cb-pes-1*, like *ce-pes-1*, appears not to be essential for embryogenesis.

Whereas the progeny of *C. elegans* adults injected with *ce-fkh-2* dsRNA displayed a limited and transitory phenotype, injection of dsRNA specific to *cb-fkh-2* into wild-type *C.*

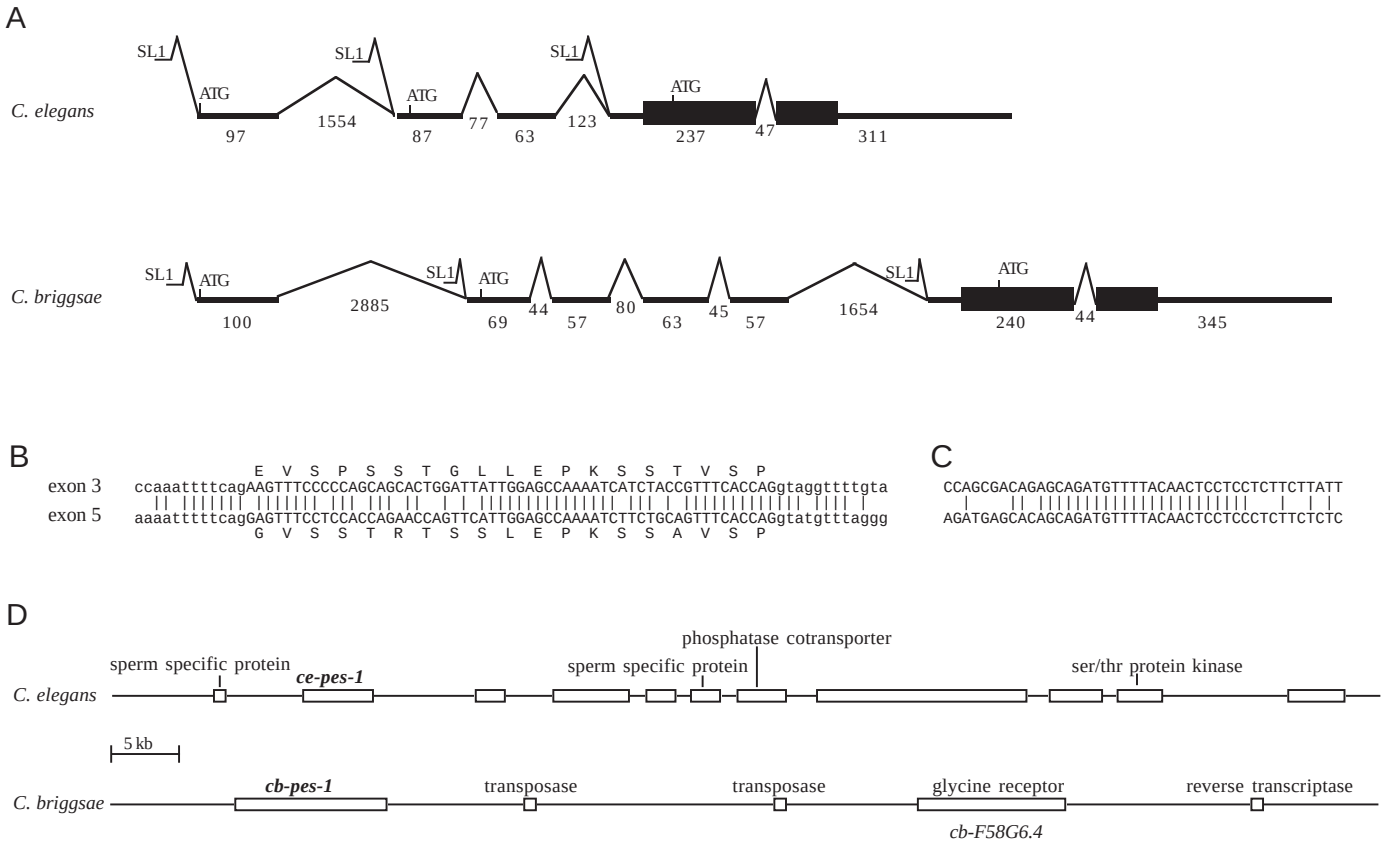

Fig. 5. The *pes-1* gene structure and environment in *C. elegans* and *C. briggsae.* (A) The *pes-1* gene structure in *C. elegans* and *C. briggsae* is based on cDNA clone analysis apart from the 3′ end of the last exon of *cb-pes-1*, which is predicted by FGENESH. Exons are represented by thick lines. The forkhead domain-encoding region is indicated by black boxes in each gene. Figures below the exons and introns indicate the size in nucleotides. SL1 is the *trans*-spliced leader. The potential sites for translation initiation are indicated by ATG. (B) Sequence similarity between the third and the fifth exons of *cb-pes-1*. The similarity extends to the splicing sites in the flanking introns. (C) A block of conserved sequence between *ce-pes-1* (top) and *cb-pes-1* (bottom) genes, located 690 and 810 nucleotides upstream of the first ATG in *C. elegans* and in *C. briggsae*, respectively, which could identify a transcription regulating element. However, the lack of perfect conservation in the regulation of *pes-1* between the two species, as revealed in the reciprocal reporter gene fusion experiments, might suggest promoter elements would not be perfectly conserved. (D) The gene environment around *pes-1* in *C. elegans* and *C. briggsae*, drawn approximately to scale. Genes are represented by open boxes. Homologies are indicated where known. The gene environment around *pes-1* in *C. elegans* is from ACeDB. In *C. briggsae*, the region represented is the 45 kb, in clone CB038P19, which has been sequenced. Genes around *cb-pes-1* were identified using BLASTX against the *C. elegans* protein database Wormpep. cb-F58G6.4 has 86% identity with ce-F58G6.4 at the amino acid level, but no similarity is observed between the gene environment of F58G6.4 in *C. elegans* and *C. briggsae* (data not shown). In *C. elegans*, F58G6.4 and *pes-1* are both on chromosome IV, but 2.9 map units apart (about 6 Mb).

briggsae adults had a much more pronounced effect (Table 1, row 8). Although the slight increase in embryonic arrest (to 1.9%) was not statistically significant, almost half of the progeny, produced during the selected period, terminally arrested at the first larval stage. The rest of the progeny developed into fertile adults. *cb-fkh-2* appears to be more crucial than *ce-fkh-2*.

Simultaneous injection of dsRNAs specific to *cb-fkh-2* and *cb-pes-1* into wild-type *C. briggsae* adults, however, has an even stronger effect than injection of *cb-fkh-2* alone (Table 1, row 9). Now a statistically significant increase in embryonic lethality (to 8%) was observed and the proportion of the progeny arresting as L1s increased to nearly 80%. A minority of the progeny still escaped arrest and developed into fertile adults. The rates of embryonic and larval lethality were similar to the ones observed upon co-injection of dsRNAs specific to *ce-fkh-2* and *ce-pes-1* into wild-type *C. elegans* adults. The functional overlap, between *pes-1* and *fkh-2*, appears to have been retained from *C. elegans* to *C. briggsae*, although the relative contribution to that overlap appears to be greater for *fkh-2* in *C. briggsae*.

DISCUSSION

The observations that *pes-1* encodes a forkhead transcription factor and is expressed in specific cells during embryogenesis suggested *pes-1* would have a role in controlling *C. elegans* development, through regulation of gene expression. Inactivation of the gene in otherwise wild-type animals, however, does not cause any obvious altered phenotype, placing *pes-1* into the vast group of *C. elegans* genes that are apparently unnecessary (Johnsen and Baillie, 1997). One explanation for the presence of non-essential genes is genetic

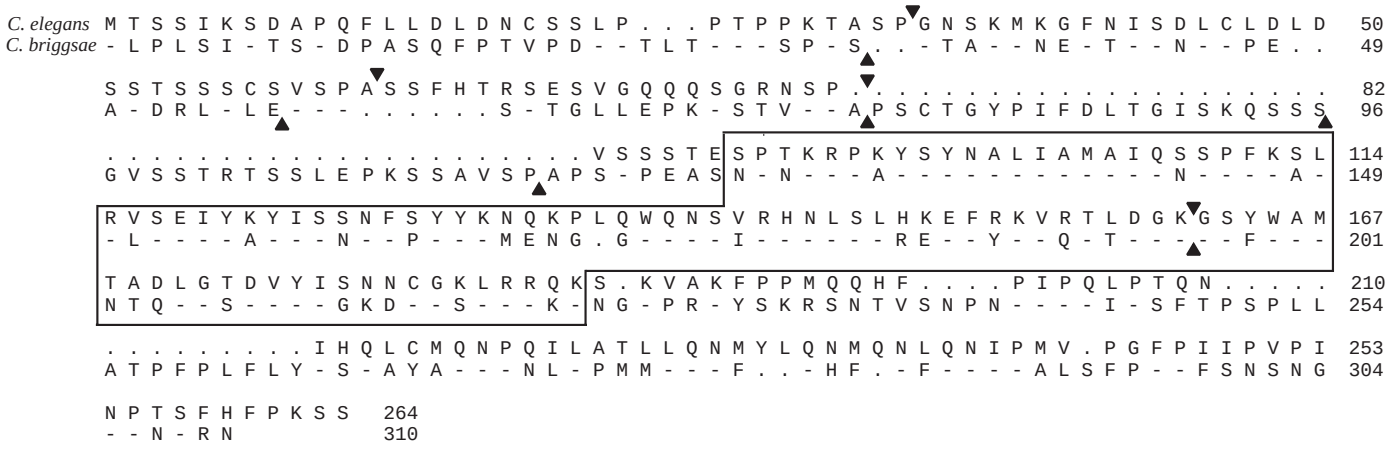

Fig. 6. Sequence alignment of the *C. elegans* and *C. briggsae* PES-1 proteins. Dashes represent conserved amino acid in the two species, dots indicate gaps in aligned sequences. The forkhead domain is boxed. The preponderance of serine residues immediately upstream of the forkhead domain is similar in ce-PES-1 (16 of 36 residues; 44%) and cb-PES-1 (14 of 33 residues; 42%). Similarly, high proportions of proline and of glutamine plus asparagine residues are also found downstream of the forkhead domain of ce-PES-1 (13 of 76 residues; 17% and 18 of 76 residues; 24%, respectively) and cb-PES-1 (15 of 88 residues; 17% and 20 of 88 residues; 23%, respectively). Several QN motifs are conserved between the two proteins. Black triangles show the location of introns.

redundancy and this has now been demonstrated to apply to *pes-1*; *pes-1* does have a function, but at least part of this function is also performed by another forkhead gene, T14G12.4 (renamed *fkh-2*).

Amongst the *C. elegans* forkhead family, *fkh-2* is not particularly closely related to *pes-1* in sequence. Similarity in the expression patterns of *pes-1* and *fkh-2* was the first suggestion that their functions might overlap. This functional overlap has now been experimentally demonstrated, with disruption of both genes being needed to arrest development. Elucidation of the nature of this function will depend on further characterization of the developmental arrest. A genetic deletion for *fkh-2* will be needed as the variation in developmental arrest, as described here, could be a consequence of RNAi not completely inactivating *fkh-2* function. However, the expression patterns of these two genes, as revealed using reporter gene fusions, do not appear to be identical, suggesting that either *pes-1* does not function everywhere it is expressed, and/or there is yet another gene with overlapping function. Candidates might be identified on the basis of expression pattern data, as for *fkh-2*.

Genetic redundancy could arise from simple gene duplication. Such redundancy might be expected to be evolutionarily unstable because there would be no selection to retain both copies. After the duplication, one of the copies would acquire inactivating mutations through genetic drift and this is the origin of numerous pseudogenes scattered through genomes. Although *fkh-2* and *pes-1* both encode forkhead proteins, the degree of sequence divergence and the absence of homology amongst neighbouring genes suggest this is not the consequence of a recent gene duplication event. Furthermore, both have function, so neither is a pseudogene. Alternatively, if a gene with multiple functions is duplicated, genetic drift might result in loss of distinct functions from each copy. In such a situation, while some functional overlap may be retained, full genetic redundancy is no longer present and both copies can be maintained by selection (e.g. *glp-1* and *lin-12*, Lambie and Kimble, 1991; *apx*-1 and *lag-2*, Gao and Kimble, 1995). Even the residual functional overlap might be lost eventually, but this could require very specific mutations and therefore be relatively stable. *fkh-2* and *pes-1* do not appear to fit this scenario as neither is essential for *C. elegans* development and a non-redundant function for *pes-1* is not apparent.

Theoretical models have been described, however, that explain how full genetic redundancy could be maintained, or even generated, by selection (Thomas, 1993; Nowak et al., 1997). These models could explain redundancy between genes where either only one of the genes (e.g. *hop-1* and *sel-12*, Westlund et al., 1999; *egl-27* and *egr-1*, Solari et al., 1999), or neither (e.g. *lin-15*A/B and *lin-8*/*lin-9*, Ferguson and Horvitz, 1989), is apparently essential. This might also apply to *pes-1* and *fkh-2* and characterization of these genes in *C. briggsae* was initiated to explore the evolutionary conservation of genetic redundancy.

Functional redundancy was observed between the *pes-1* and *fkh-2* homologues in *C. briggsae* confirming that, as suggested in the theoretical models, such redundancy can be evolutionarily stable (Nowak et al., 1997). This redundancy is, however, not completely conserved between *C. elegans* and *C. briggsae*. *fkh-2* appears to have a much more important role than *pes-1* in *C. briggsae* embryogenesis and this may be linked to the differences in sequence between *cb-pes-1* and *ce-pes-1*.

Only 44% amino acid identity is observed between ce-PES-1 and cb-PES-1, 69% identity for the forkhead DNA binding domain. This represents a high divergence rate within the range observed for other genes for which *C. briggsae* and *C. elegans* homologues have been sequenced (43-100%; De Bono and Hodgkin, 1996; Kuwabara, 1996). However, evolutionary selection on protein sequence varies with the nature of protein encoded and comparison within a gene family would be more informative. The amino acid identity between all four other *C. briggsae* forkhead transcription factors, for which sequence is available, and their *C. elegans* orthologues is, on average, 73% in total and at least 96% over the forkhead domain (Table 2).

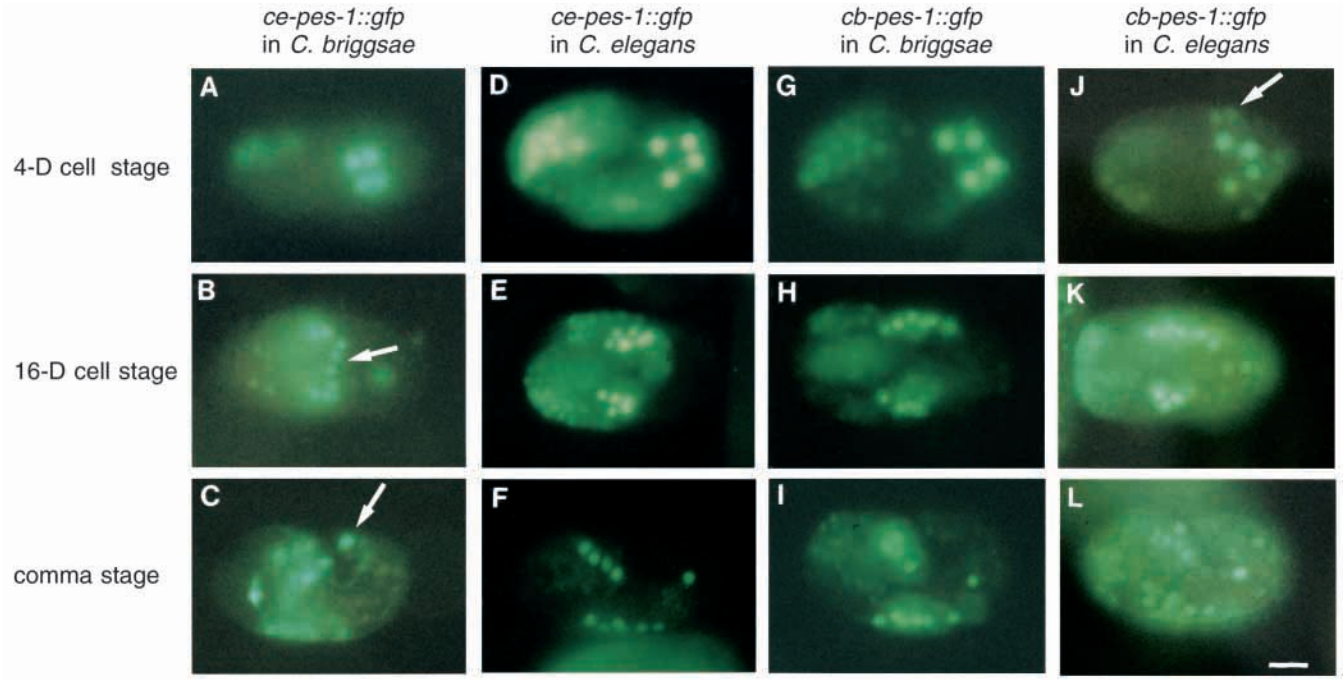

Fig. 7. Interspecies comparison of *pes-1::gfp* expression. Expression of *ce-pes-1::gfp* (A-F) and *cb-pes-1::gfp* (G-L) reporter gene fusions were observed in *C. elegans* (D-F and J-L) and in *C. briggsae* (A-C and G-I). Anterior is to the left in all panels. The expression pattern directed by *ce-pes-1::gfp* in *C. elegans* at (D) the 185-cell stage (expression is observed in the four D cell descendants and in several AB descendants more anteriorly), (E) the 385-cell stage (expression is visible in the two clusters of 8 D cell descendants and is fading in the AB component) and (F) the comma stage (expression is visible in the two clusters of D cell descendants and in Z1 and Z4, only one of which is visible in this focal plane) is apparently identical to that directed by *cb-pes-1::gfp* in *C. briggsae* at the same stages (G,H and I). The level of expression obtained in reciprocal reporter gene fusion experiments (*cb-pes-1::gfp* into *C. elegans* and *ce-pes-1::gfp* into *C. briggsae*) is significantly lower than that obtained in the non-reciprocal experiments for all components of the patterns. To allow visualization of the weaker components, expression in the D component was adjusted to a similar intensity by increasing the exposure time. Two weak extra components are directed by *ce-pes-1::gfp* in *C. briggsae*: four cells in a row between the two clusters of D descendants at the 16 D-cell stage (B, arrow) and one cell at the posterior tip of the embryo at the comma stage (C, arrow). Expression of *cb-pes-1::gfp* in *C. elegans* is also observed in cells around the D descendants at the 4 D-cell stage (J, arrow). Scale bar in L, 10 μm.

Clearly, *ce-pes-1* and *cb-pes-1* are more divergent than might be expected.

The data suggest the following possible evolutionary history for *pes-1*. Functional redundancy between *fkh-2* and *pes-1* would have been present in the last common ancestor to *C. briggsae* and *C. elegans*. The greater importance for *fkh-2* in *C. briggsae* might suggest that the need to maintain, and the selection for, *cb-pes-1* has been diminished in the *C. briggsae* ancestry, but not in the *C. elegans* ancestry. Consistent with this, *ce-pes-1* is located in a typical gene environment, whereas, from the 45 kb of sequence available, *cb-pes-1* is located in a markedly gene poor region, surrounded by transposase genes (Fig. 5D). Perhaps, the genetic redundancy actually permitted the original mutation that reduced the role of the ancestral *cb-pes-1*, and this may even have been the rearrangement that placed the gene in its current environment. No *C. briggsae* orthologue of any gene within 100 kb of *ce-pes-1* has been found as yet, although this may be because only 10% of the *C. briggsae* genome has been sequenced so far. A reduction in importance of the ancestral *cb-pes-1* would have allowed further genetic drift, leading to the observed divergence in sequence between the *C. elegans* and *C. briggsae pes-1* genes, but some selection must have remained on *cb-pes-1* to keep the gene functional. Indeed the apparently recent, internal tandem duplication in *cb-pes-1* (83% nucleotide sequence identity between exons 3 and 5) (Fig. 5B) could reflect selection to increase the protein's transcriptional

Table 2. Comparison of sequenced *C. briggsae* forkhead transcription factors with their *C. elegans* orthologues

	% Amino acid sequence identity between *C. elegans* and *C. briggsae* homologues		
Gene	Forkhead domain	Outside of the forkhead domain	Total protein
pes-1	69	28	44
T14G12.4	96	65	76
K03C7.2*	99	67	78
F26B1.7‡	97	54	65
pha-4§	96	65	71

Comparisons are based on FGENESH gene structure predictions.

*In comparison to that presented in ACeDB, FGENESH does not predict the 5′ exon and predicts an extra 3′ exon for K03C7.2. The same gene structure is predicted for the *C. briggsae* homologue.

‡FGENESH predicts three extra exons at the 3′ end of F26B1.7, as compared to that presented in ACeDB. The same gene structure is predicted for the *C. briggsae* homologue.

§The comparison does not include the first exon of *pha-4* because the full sequence is not yet available for the *C. briggsae* homologue (comparison is for 491 residues out of 505).

activation activity and restore the gene's full function. Conservation of the *pes-1* expression pattern between *C. elegans* and *C. briggsae* is also suggestive of retention of selection pressure. In fact, the reduction in expression levels in the reciprocal reporter gene fusion experiments, associated with the appearance of new components of expression, might suggest that there has been some co-evolution of the *cis*-acting elements and the *trans*-acting factors responsible for *pes-1* expression. A similar co-evolution has been described in *Drosophila* within the promoter of the *even-skipped* gene (Ludwig et al., 2000). Such an evolutionary scenario, testable from the prediction that *pes-1* orthologues from other nematode species such as *Caenorhabditis remanei* should be more similar to *ce-pes-1* than *cb-pes-1*, may be revealing the potential of genetic redundancy as a substrate for evolutionary change.

While molecular phylogenetic analysis has revealed likely orthologues for many of the *C. elegans* forkhead genes in species outside the Nematoda (Fig. 2), no orthologue of *pes-1* has yet been identified beyond *Caenorhabditis*, not even in the substantially complete *Drosophila melanogaster* genome sequence (Adams et al., 2000). In contrast, *fkh-2* is quite a close homologue of the *Drosophila* segmentation gene *sloppy-paired* (*slp*) and of the chordate gene *Brain factor 1* (*BF-1*) (Fig. 2). These observations might suggest that *pes-1* is a phylum-specific gene and/or that a low level of selection on *pes-1* in the evolutionary history of the common ancestor of *C. briggsae* and *C. elegans*, because of genetic redundancy, might have contributed to the divergence of the *pes-1* gene. Indeed, *pes-1* and *fkh-2* may be related by a gene duplication event which, although predating the *C. elegans*/*C. briggsae* divergence, may be much more recent than the molecular phylogenetic analysis might suggest. In addition, the anterior expression of BF-1 in chordates (Tao and Lai, 1992) and of *slp* in *Drosophila* (Grossniklaus et al., 1992) could correlate with the expression of *pes-1* and *fkh-2* in the AB cell lineage in *C. elegans*. Curiously *slp* also demonstrates functional redundancy (Cadigan et al., 1994), in this case as an adjacent pair of more similar genes that therefore probably originated with a relatively recent, gene duplication event.

If genetic redundancy, as observed between *pes-1* and *fkh-2*, is evolutionarily stable then genetic redundancy may be common. There is one other gene pair, *ace-1* and *ace-2*, which are known to be redundant in *C. elegans* (Johnson et al., 1981) and well conserved by sequence in *C. briggsae* (Grauso et al., 1998), although conservation of functional redundancy, in this case, has not been demonstrated. Whatever the evolutionary history, *pes-1* and *fkh-2* do function in *C. elegans* embryonic development and could not have been detected easily through conventional genetic analysis. Expression pattern data could be important for identification of such redundant genes and for full comprehension of developmental processes.

We are grateful to R. Plasterk for providing the MT3126 strain, the Washington University Genome Sequencing Center for providing the *C. briggsae* clones and sequencing of the clone CB038P19, A. Coulson and R. Shownkeen for the T14G12 cosmid, and A. Fire for the GFP reporter plasmids. This research was supported by grants from the European Communities Human Capital and Mobility Programme, from the Biotechnology and Biological Sciences Research Council, and from the Medical Research Council.

REFERENCES

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F. et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185-2195.

Ashburner, M., Misra, S., Roote, J., Lewis, S. E., Blaze, R., Davis, T., Doyle, C., Galle, R., George, R., Harris, N. et al. (1999). An exploration of the sequence of a 2.9-Mb region of the genome of *Drosophila melanogaster*: the Adh region. *Genetics* **153**, 179-219.

Ausubel, F. M., Brent R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G. and Struhl, K. (1993). *Current Protocols in Molecular Biology.* New York, N.Y: John Wiley & Sons, Inc.

Cadigan, K., Grossniklaus, U. and Gehring, W. J. (1994). Functional redundancy: the respective roles of the two sloppy paired genes in *Drosophila* segmentation. *Proc. Natl Acad. Sci. USA* **91**, 6324-6328.

***C. elegans* sequencing consortium** (1998). Genome sequence of the nematode *C. elegans*. A platform for investigating biology. *Science* **282**, 2012-2018.

Clevidence, D., Overdier, D. G., Tao, W., Qian, X., Pani, L., Lai, E. and Costa, R. H. (1993). Identification of nine tissue-specific transcription factors of the hepatocyte nuclear factor 3/forkhead DNA-binding-domain family. *Proc. Natl Acad. Sci. USA* **90**, 3948-3952.

Cooke, J., Nowak, M. A., Boerlijst, M. and Maynard-Smith, J. (1997). Evolutionary origins and maintenance of redundant gene expression during metazoan development. *Trends Genet.* **13**, 360-364.

De Bono, M. and Hodgkin, J. (1996). Evolution of sex determination in *Caenorhabditis*: unusually high divergence of *tra-1* and its functional consequences. *Genetics* **144**, 587-595.

Emmons, S., Klass, M. R. and Hirsh, D. (1979). Analysis of the constancy of DNA sequences during development and evolution of the nematode C. elegans. *Proc. Natl. Acad. Sci. USA* **76**, 1333-1337.

Ferguson, E. L. and Horvitz, H. R. (1989). The multivulva phenotype of certain *C. elegans* mutants results from defects in two functionally redundant pathways. *Genetics* **123**, 109-121.

Fire, A., Harrison, S. W. and Dixon, D. (1990). A modular set of *lacZ* fusion vectors for studying gene expression in *Caenorhabditis elegans*. *Gene* **93**, 189-98.

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

Galtier, N., Gouy, M. and Gautier, C. (1996). SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.* **12**, 543-548.

Gao, D. and Kimble, J. (1995). APX-1 can substitute for its homolg LAG-2 to direct cell interactions throughout *Caenorhabditis elegans* development. *Proc. Natl Acad. Sci. USA* **92**, 9839-9842.

Grauso, M., Culetto, E., Combes, D., Fedon, Y., Toutant, J. P. and Arpagaus, M. (1998). Existence of four acetylcholinesterase genes in the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *FEBS Lett.* **424**, 279-284.

Grossniklaus, U., Pearson, R. K. and Gehring, W. J. (1992). The *Drosophila sloppy paired* locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors. *Genes Dev.* **6**, 1030-1051.

Hacker, U., Kaufmann, E., Hartmann, C., Jurgens, G., Knochel, W. and Jackle, H. (1995). The *Drosophila* fork head domain protein crocodile is required for the establishment of head structures. *EMBO J.* **14**, 5306-5317.

Hobert, O., Moerman, D. G., Clark, K. A., Beckerle, M. C. and Ruvkun, G. (1999). A conserved LIM protein that affects muscular adherens junction integrity and mechanosensory function in *Caenorhabditis elegans*. *J. Cell Biol.* **144**, 45-57.

Hope, I. A. (1991). Promoter trapping in *Caenorhabditis elegans*. *Development* **113**, 399-408.

Hope, I. A. (1994). PES-1 is expressed during early embryogenesis in *Caenorhabditis elegans* and has homology to the fork head family of transcription factors. *Development* **120**, 505-514.

Johnsen, R. and Baillie, D. L. (1997). Mutation. In *C. elegans II* (ed. D. L. Riddle, T. Blumenthal, B. M. Meyer and J. R. Priess), pp. 79-95. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Johnson, C., Duckett, J. G., Culotti, J. G., Herman, R. K., Meneely, P. M. and Russell, R. L. (1981). An acetylcholinesterase-deficient mutant of the nematode *C. elegans*. *Genetics* **97**, 261-279.

Kaestner, K. H., Knochel, W. and Marinez, D. E. (2000). Unified

nomenclature for the winged helix/forkhead transcription factors. *Genes Dev.* **14**, 142-146.

Kalb, J. M., Lau, K. K., Goszczynski, B., Fukushige, T., Moons, D., Okkema, P. G. and McGhee, J. D. (1998). *pha-4* is *Ce-fkh-1*, a *Forkhead*/HNF-3α,β,γ homolog that functions in organogenesis of the *C. elegans* pharynx. *Development* **125**, 2171-80.

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

Krause, M., Wild, M., Rosenzweig, B. and Hirsh, D. (1989). Wild type and mutant actin genes in *Caenorhabditis elegans*. *J. Mol. Biol.* **208**, 381-392.

Kuwabara, P. (1996). Interspecies comparison reveals evolution of control regions in the nematode sex-determining gene *tra-2*. *Genetics* **144**, 597-607.

Lambie, E. J. and Kimble J. (1991). Two homologous regulatory genes, *lin-12* and *glp-1*, have overlapping functions. *Development* **112**, 231-240.

Ludwig, M., Bergman, C., Patel, N. H. and Kreitman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* **403**, 564-567.

Mango, S. E., Lambie E. J. and Kimble, J. (1994). The *pha-4* gene is required to generate the pharyngeal primordium of *Caenorhabditis elegans*. *Development* **120**, 3019-3031.

Martienssen, R. (1998). Functional genomics: probing plant gene function and expression with transposons. *Proc. Natl Acad. Sci. USA* **95**, 2021-2026.

Mello, C. C., Kramer, J. M., Stinchcomb, D. and Ambros, V. (1991). Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959-70.

Messom, A. L. (1996). Characterisation of Forkhead Genes in *Caenorhabditis elegans*. PhD thesis, University of Leeds, Leeds, UK.

Miller, L. M., Gallegos, M. E., Morisseau, B. A. and Kim, S. K. (1993). *lin-31*, a *C. elegans* HNF-3/*fork head* transcription factor homolog, specifies three alternative cell fates in vulval development. *Genes Dev.* **7**, 933-47.

Mitchell, P. J. and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**, 371-378.

Molin, L., Schnabel, H., Kaletta, T., Feichtinger, R., Hope, I. A. and Schnabel, R. (1999). Complexity of developmental control: Analysis of embryonic cell lineage specification in *Caenorhabditis elegans* using *pes-1* as an early marker. *Genetics* **151**, 131-141.

Nowak, M. A., Boerlijst, M. C., Cooke, J. and Maynard-Smith, J. (1997). Evolution of genetic redundancy. *Nature* **388**, 167-71.

Ogg, S., Paradis, S., Gottlieb, S., Patterson, G. I., Lee, L., Tissenbaum, H. A. and Ruvkun, G. (1997). The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. *Nature* **389**, 994-999.

Oliver, S., van der Aart, Q. J., Agostoni-Carbone, M. L., Aigle, M., Alberghina, L., Alexandraki, D., Antoine, G., Anwar R., Ballesta, J. P., Benit, P. et al. (1992). The complete DNA sequence of yeast chromosome III. *Nature* **357**, 38-46.

Page, R. (1996). TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12**, 357-358.

Prasad, S. and Baillie, D. L. (1989). Evolutionarily conserved coding sequences in the *dpy-20–unc-22* region of *Caenorhabditis elegans*. *Genomics* **5**, 185-198.

Riddle, D. L., Swanson, M. M. and Albert, P. S. (1981). Interacting genes in nematode dauer larva formation. *Nature* **290**, 668-671.

Ruvkun, G. and Hobert, O. (1998). The taxonomy of developmental control in *Caenorhabditis elegans*. *Science* **282**, 2033-41.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular cloning: A laboratory manual*. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press.

Solari, F., Bateman, A. and Ahringer, J. (1999). The *Caenorhabditis elegans* genes *egl-27* and *egr-1* are similar to MTA1, a member of a chromatin regulatory complex, and are redundantly required for embryonic patterning. *Development* **126**, 2483-2494.

Sulston, J. and Hodgkin, J. (1988). In *The nematode Caenorhabditis elegans.* (ed. W. B. Wood), pp. 587-606. Cold Spring Harbor, N.Y: Cold Spring harbor Laboratory Press.

Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierry-Mieg, J., Qiu, L. et al. (1992). The *C. elegans* genome sequencing project: a beginning. *Nature* **356**, 37-41.

Tabara H., Grishok A. and Mello C. C. (1998). RNAi in *C. elegans* – soaking in the genome sequence. *Science* **282**, 430-431.

Tao, W. and Lai, E. (1992). Telencephalon-restricted expression of BF-1, a new member of the HNF-3/*fork head* gene family, in the developing rat brain. *Neuron* **8**, 957-966.

Tavernarakis, N., Wang, S. L., Dorovkov, M., Ryazanov, A. and Driscoll, M. (2000). Heritable and inducible genetic interference by double-stranded RNA encoded by transgenes. *Nat. Genet.* **24**, 180-183.

Thomas, J. H. (1993). Thinking about genetic redundancy. *Trends Genet.* **9**, 395-399.

Thompson, J., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876-4882.

Waterston, R. and Sulston, J. (1995). The genome of *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **92**, 10836-10840.

Westlund, B., Parry, D., Clover, R., Basson, M. and Johnson, C. D. (1999). Reverse genetic analysis of *Caenorhabditis elegans* presenilins reveals redundant but unequal roles for *sel-12* and *hop-1* in Notch-pathway signaling. *Proc. Natl Acad. Sci. USA* **96**, 2497-2502.

Williams, B. D., Schrank B., Huynh C., Shownkeen R. and Waterston R. H. (1992). A genetic mapping system in *C. elegans* based on polymorphic sequence tagged sites. *Genetics* **131**, 609-624.

Zwaal, R. R., Broeks, A., van Meurs, J., Groenin, J. T. M. and Plasterk, R. H. A. (1993). Target-selected gene inactivation in *Caenorhabditis elegans* by using a frozen transposon insertion mutant bank. *Proc. Natl. Acad. Sci. USA* **90**, 7431-7435.