Functional Genomics in Caenorhabditis elegans: An Approach Involving Comparisons of Sequences from Related Nematodes

Colin Thacker, Marco A. Marra, Alana Jones, David L. Baillie and Ann M. Rose

Supplementary data	*" Supplemental Research Data"* **http://www.genome.org/cgi/content/full/9/4/348/DC1**
References	This article cites 41 articles, 20 of which can be accessed free at: **http://www.genome.org/cgi/content/full/9/4/348#References**
	Article cited in: **http://www.genome.org/cgi/content/full/9/4/348#otherarticles**
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here**

Notes

Letter

Functional Genomics in *Caenorhabditis elegans*: An Approach Involving Comparisons of Sequences from Related Nematodes

Colin Thacker,[1] Marco A. Marra,[2] Alana Jones,[1] David L. Baillie,[3] and Ann M. Rose[1,4]

[1]*Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, V6T 1Z3, Canada;* [2]*Genome Sequencing Center, Washington University, St. Louis, Missouri 63108 USA;* [3]*Institute of Molecular Biology and Biochemistry, Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada*

Comparative genomic analysis was used to investigate the gene structure of the *bli–4* locus from two related *Caenorhabditis* species, *C. elegans* and *C. briggsae*. In *C. elegans*, *bli–4* is a complex gene encoding a member of the kex2/subtilisin-like family of proprotein convertases. Genomic sequence comparisons coupled with RT–PCR analysis identified five additional coding exons that had not been identified previously using standard recombinant DNA techniques. The *C. briggsae* gene was able to rescue both viable blistered and developmentally arrested mutants of *C. elegans bli–4*, demonstrating functional conservation. In addition, deletion analysis of conserved sequences outside of coding regions, combined with phenotypic rescue experiments, identified regulatory elements that alter the expression of the *bli–4* gene. These results demonstrate the utility of genomic sequence comparisons of homologous genes in related species as an effective tool with which to dissect the functional information of complex genes.

[The sequence for cosmid K04I0 is available at GenBank (accession no. AFO 39719); fosmids G06P23 and G25K0I are available as online supplementary material at www.genome.org.]

The efforts of the *Caenorhabditis elegans* (*C. elegans*) Genome Sequencing Consortium are leading to a comprehensive description of the nematode genome (Berks et al. 1995; Hodgkin et al. 1995; the *C. elegans* Sequencing Consortium 1998) including the complete sequence of the *bli-4* locus. Furthermore, to facilitate studies of gene structure and regulation, the Washington University Genome Sequencing Center has begun a similar project to determine the genome sequence of the highly related nematode *Caenorhabditis briggsae* (*C. briggsae*). Although both nematodes are almost identical in morphology and development (Nigon and Dougherty 1949), their genomes have diverged. Several estimates suggest that separation of the two species occurred 23–40 million years ago (Emmons et al. 1979; Heschl and Baillie 1990; Lee et al. 1992; Kennedy et al. 1993). Conservation of DNA sequences is confined largely to protein-coding regions and short flanking sequences. Comparative studies have demonstrated that conserved 5′ flanking sequences often constitute *cis*-acting elements that are involved in the regulation of gene expression as first described by Snutch (1984) and subsequently by several other groups (Heine and Blumenthal 1986; Prasad and Baillie 1989; Zucker-

Aprison and Blumenthal 1989; Heschl and Baillie 1990; Lee et al. 1992; Xue et al. 1992; Kennedy et al. 1993; Kuwabara and Shah 1994; de Bono and Hodgkin 1996; Kuwabara 1996). In addition, functional conservation between the two species has also been demonstrated by rescue of mutant phenotypes via DNA-mediated transformation (Kennedy et al. 1993; Krause et al. 1994; de Bono and Hodgkin 1996; Kuwabara 1996; Maduro and Pilgrim 1997).

bli-4 is one of four loci that are members of the *C. elegans* kex2/subtilisin-like family of proprotein convertases (Thacker et al. 1995). Family members encode serine endoproteinases that are conserved throughout the metazoa. These proteins function to activate a diverse number of physiologically important polypeptides by limited proteolytic cleavage at single, di- or oligomeric basic amino acids (Seidah and Chrétien 1992; Seidah et al. 1994). Molecular and genetic characterization of *bli-4* has demonstrated the gene structure to be complex (Peters et al. 1991; Thacker et al. 1995). A large number of mutations within the gene have been recovered and all but one result in an arrest of nematode development in late embryogenesis. The exception is the allele *e937* in which animals homozygous for the mutation are viable but display blistering of the adult cuticle. The penetrance of the blistered phenotype is incomplete; only ~90% of homozygotes

[4]**Corresponding author.**
E-MAIL arose@genekit.medgen.ubc.ca; FAX 822-5348.

express the mutant phene. Molecular characterization of *bli-4* has revealed that the gene encodes at least four products that arise from alternative splicing of twelve common exons, encoding a catalytic domain, to different 3′ exons. Alternate splicing results in the generation of isoforms that share the same signal peptide, prodomain, protease, and middle domains yet diverge structurally because their unique carboxyl termini. The generation of isoforms that differ at their carboxyl termini is a property shared by other members of the kex2/subtilisin-like family (Kiefer et al. 1991; Seidah et al. 1992; Chan et al. 1992; Roebroek et al. 1993; de Bie et al. 1995, 1996). The different carboxyl termini are thought to direct the transport and localization of the individual isoforms to specific sites within the cell (Molloy et al. 1994; Bosshart et al. 1995; Schäfer et al. 1995; Creemers et al. 1996; De Bie et al. 1996).

The generation of more than one gene product may explain at least some of the genetic complexity, yet adds to the difficulty of establishing functional roles for *bli-4*. Phenotypic examination of the various *bli-4* mutants and analysis of a genetic suppressor of blistering suggest that *bli-4* convertase activity may be required for the cleavage of cuticle collagens (Peters et al. 1991; Thacker et al. 1995), although whether this putative function is shared between the individual gene products or whether it is isoform-specific remains to be determined. The lethal phenotypes exhibited by 12 independent *bli-4* mutants are the result of lesions within the first 12 common exons, suggesting that the embryonic arrest is caused by the lack of expression or catalytic activity of all *bli-4* isoforms (Thacker et al. 1995; C. Thacker, unpubl.). In contrast, the blistered mutant *e937* is the result of a 3325-bp deletion believed previously to affect the expression of only one isoform, blisterase A. Together, these results suggest that *bli-4* convertase activity is regulated during nematode development such that one or several products are required during late embryogenesis, whereas blisterase A activity is required functionally for the production or maintenance of the adult cuticle.

In this report we describe an effort to more fully delineate functionally important sequences by comparing the *C. elegans bli-4* locus to *C. briggsae bli-4*. We show that like *C. elegans*, the *C. briggsae* gene is organized such that a common 5′ group of exons is predicted to be alternately spliced at the 3′ end to exons that are conserved between the two species even though the intron–exon structures display several differences. The analysis also suggests that the *bli-4* locus may be more complex than at first thought, with the potential existence of an additional five isoforms. Finally, we describe a preliminary investigation of conserved sequences outside of the coding region that may comprise transcriptional elements that are important for regulating the expression of the *bli-4* gene.

RESULTS

Genomic Organization of the *C. elegans* and *C. briggsae* *bli–4* Locus

A sequence comparison of the *bli-4* region from both *C. elegans* and *C. briggsae* was performed using the program DOTTER (Sonnhammer and Durbin 1996) which provides both a graphical dot-matrix plot and a sequence alignment display. These tools offer a powerful means of finding homology between two sets of sequences. Comparing the complete sequence of cosmid K04F10 from *C. elegans* and the *C. briggsae* fosmid clone G06P23 we were able to determine genomic conservation throughout the *bli-4* region (Fig. 1). The program GeneFinder (P. Green and L. Hillier, unpubl.) predicted five genes in the cosmid K04F10, located 3′ of *bli-4*. These genes are in the order K04F10.6, K04F10.3, K04F10.2, K04F10.1, and K04F10.7. All six genes are conserved and positioned in the same orientation in *C. briggsae* (labeled regions 1–6 in Fig. 1), demonstrating that syntenic conservation exists across ~40 kb of genomic sequence at the *bli-4* locus of *C. elegans* and *C. briggsae*. Syntenic conservation between the two species has been observed for a number of loci and has been exploited as a useful means of cloning homolo-
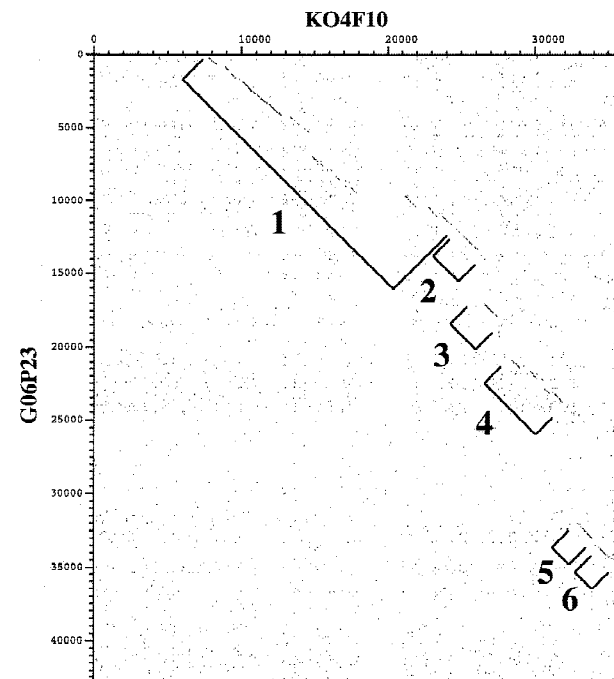

Figure 1 Alignment of *C. elegans* (K04F10) and *C. briggsae* (G06P23) sequences showing syntenic conservation within the *bli-4* locus. K04F10 sequences (GenBank accession no. AF039719) are plotted horizontal against G06P23 (vertical axis). The relative positions of genes predicted by the program GeneFinder (P. Green and L. Hillier, unpubl.) are indicated by brackets as (*1*) *bli-4* K04F10.4, (*2*) K04F10.6, (*3*) K04F10.3, (*4*) K04F10.2, (*5*) K04F10.1, and (*6*) K04F10.7.

gous genes that have been difficult to identify by hybridization (Kuwabara and Shah 1994).

Using DOTTER we determined the degree of conservation of the *bli-4* gene between the two *Caenorhabditis* species. Exons that we have shown previously to comprise the various *C. elegans bli-4* (*Ce–bli-4*) transcripts (Thacker et al. 1995) have been numbered in a 5′ → 3′ orientation that identifies exons 1–21. In addition, the homologous coding regions that comprise the predicted *C. briggsae* gene (*Cb–bli-4*), have been identified using corresponding numbers. Unique exons that are alternately spliced to the common exons are labeled alphabetically for both genes. As can be seen (Fig. 2), the *C. elegans* gene spans ~20 kb and is composed of at least 21 exons (the common region includes exons 1–12) as determined previously by the characterization of several cDNA clones (Thacker et al. 1995) In comparison, the *C. briggsae bli-4* gene is smaller, spanning ~12 kb, including 15 exons that show similarity to *C. elegans* sequences. One reason for the larger size of the *C. elegans* locus is the insertion of a 2339-bp retrotransposon-like element (labeled RTE in Fig. 2) between exons 13 and 14. The predicted orientation of transcription for the retrotransposon-like element is opposite that of *bli-4*, and presumably does not interfere with the expression of the *bli-4* gene. Ad-

ditional factors contributing to the compact nature of the *Cb–bli-4* gene include short or absent intervening sequences. In the common region of the *C. briggsae* gene, and partially encoding the protease domain, exons 6–9 are merged into a single exon of 1242 bp. In addition, a single exon of 337 bp is found at the end of the common region that corresponds to exons 11 and 12 in the *C. elegans* gene. Additional examples of major differences in the intron–exon structure of the *bli-4* locus can be found at the 3′ end of the gene. The coding sequences for the carboxyl terminus of the BLI-4B isoform are encoded by a single exon in *Ce–bli-4*. In contrast, the corresponding region of the *C. briggsae bli-4B* sequence is composed of two exons interrupted by a short intron of 46 bp. Curiously, sequences homologous to *C. elegans* exon 18 and unique to the blisterase C isoform(s) are missing completely in the *C. briggsae* gene. This finding suggests strongly that blisterase C function [and a possible variant called blisterase C′ (Thacker et al. 1995)] is not conserved in *C. briggsae*. The final difference in intron–exon structure is the prediction of a single exon of at least 283 bp that has sequences homologous for *C. elegans* exons 19 and 20. The disparity in intron size and number has also been observed in previous interspecies comparisons of other genes including *mec-3* (Xue et al. 1992), *ges-1*
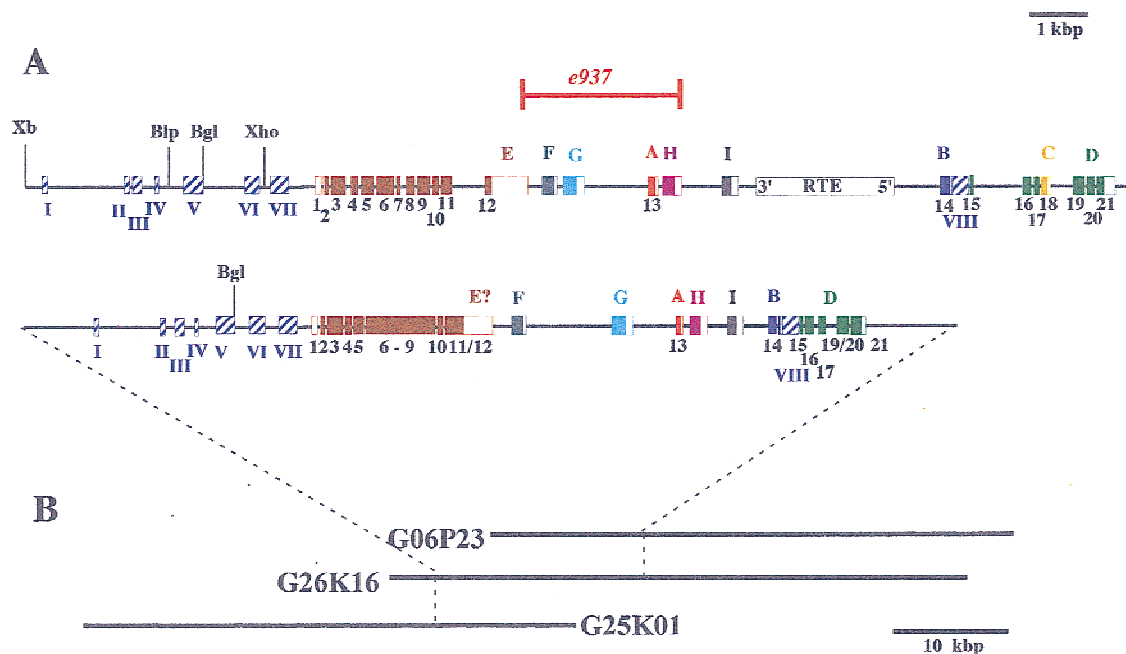

Figure 2 Schematic representation of the *bli-4* gene (*A*) from *C. elegans* (*top*) and *C. briggsae* (*bottom*). All exons are numbered according to the gene structure presented previously (Thacker et al. 1995). The common region(s) encoding the signal peptide, prodomain, protease domain, and middle domain are shown in brown. Alternatively spliced exons that encode carboxyl termini unique to the individual isoforms are labeled alphabetically and color coded. The position and extent of the *e937* 3325-bp deletion is indicated. Open boxes represent noncoding exons or untranslated regions; shaded boxes represent coding exons. Hatched boxes (labeled I–VIII) represent regions of nucleotide homology that may constitute regulatory elements, particularly those at the 5′ end. (Xb) *Xba*I; (Blp) *Blp*I; (Bgl) *Bgl*II; (Xho) *Xho*I. (*B*) The approximate location of the *C. briggsae bli-4* gene is indicated by dashed lines on the fosmid clones used in this report. The entire sequence of clones G25K01 and G06P23 has been determined, whereas G26K16 was used solely for transformation rescue experiments.

(Kennedy et al. 1993), *tra-1* (de Bono and Hodgkin 1996), and *unc-119* (Maduro and Pilgrim 1997).

Of interest was the striking sequence conservation of the noncoding first exon (Fig. 3A). Although the *C. elegans* exon is 128 bp compared to 131 bp for *Cb–bli-4*, the overall nucleotide sequence identity is 82%. A single AUG codon is contained within the exon but is followed six nucleotides downstream by an in-frame stop codon that predicts that translation of these sequences does not occur.

Although there is general divergence in intronic nucleotide sequence content, an exception to this rule was found within the intervening sequence that separates the *bli-4B* exon(s) from the downstream 3′ exon (labeled region VIII in Fig. 2). Overall nucleotide homology within this intron was 64% over 415 bp, with short stretches of high conservation (Fig. 3B). The nucleotide similarity shared between the introns raises the possibility that this region may represent a *cis* element that functions in transcriptional regulation or in controlling alternate splicing of the downstream exons.

Conservation of the *bli–4* Gene Products at the Amino Acid Level

The alignment of the regions common to *Ce–bli-4* and *Cb–bli-4* reveals a high degree of amino acid homology within the two protein segments that encode the signal peptide, prodomain, protease domain, and middle domain (Fig. 4). The two regions are colinear (657 amino acids) and are 96% identical with most divergence occurring within the signal peptide sequences, prodomains, and the middle domain. The extensive conservation in sequence suggests that this portion of the *bli-4* gene product is essential for convertase activity.

In contrast, alignment of the amino acid sequences of each carboxyl terminus shows divergence between different isoforms (Fig. 4B). A comparison of the blisterase A carboxyl termini show a difference in length and sequence similarity such that Ce–BLI-4A is 5 amino acids shorter than the *C. briggsae* counterpart although 8 of 13 residues are conserved. Similarly, the predicted Ce–BLI-4B (73 amino acids) and Cb–BLI-4B (66 amino acids) carboxyl termini have diverged in sequence, sharing only 55% identity. The carboxyl termini of the BLI-4D isoforms are well conserved sharing 79% identity over almost 300 residues (Fig. 4B). The Ce–BLI-4D (289 amino acids) and Cb–BLI-4D (292 amino acids) carboxyl termini include two cysteine-rich regions, a transmembrane domain, and two motifs that are shared with human furin (Thacker et al. 1995). These amino acid motifs are located in the cytoplasmic tail of human furin and are required for trafficking of the protein within the constitutive secretory pathway (Molloy et al. 1994; Bosshart et al. 1995; Jones et al. 1995; Schäfer et al. 1995).

Identification of Additional BLI-4 Isoforms

Analysis of the sequence of *Ce–bli-4* using GeneFinder had predicted the existence of an additional exon just distal to the common region (labeled E in Fig. 2). That
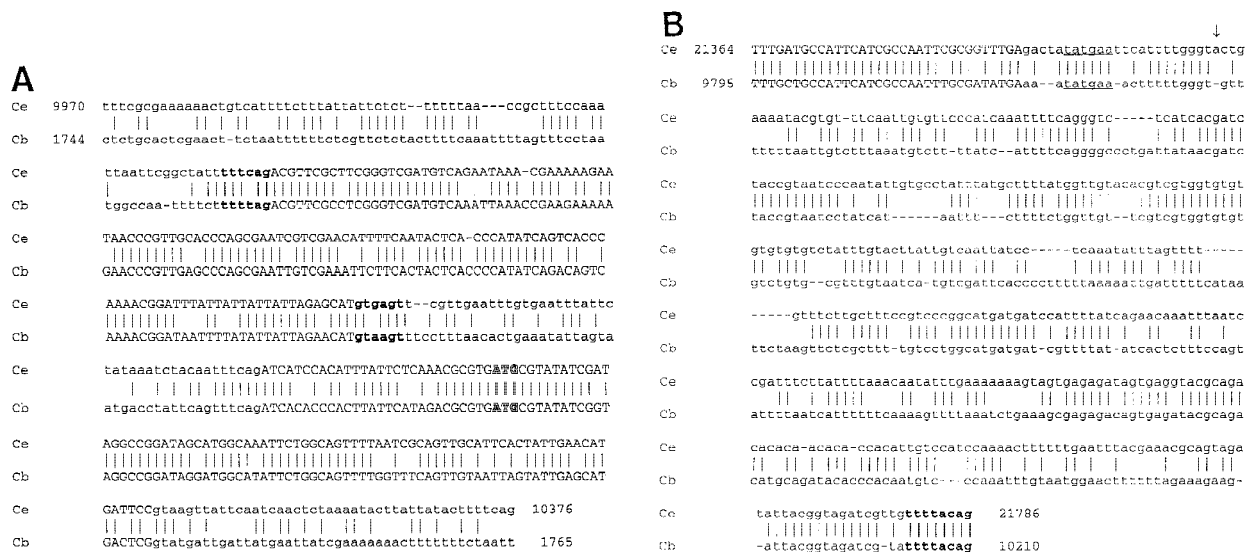
Figure 3 Alignment of nucleotide sequences that comprise untranslated exon one and exon two from *Ce–bli-4* (Ce) and *Cb–bli-4* (Cb). The splice acceptor for the outron/exon one boundary and splice donor are in boldface type. The *bli-4* initiator methionine is outlined within exon two. Exonic sequences are in uppercase, introns in lowercase. (*B*) Alignment of region VIII sequences from *Ce–bli-4* with the corresponding region from *Cb–bli-4. bli-4B* predicted coding sequences are in upper case. The *bli-4B* polyadenylation motif has been underlined, and the site of addition for the poly(A)⁺ tail is indicated by the vertical arrow. The splice acceptor within the intron has been highlighted in bold.
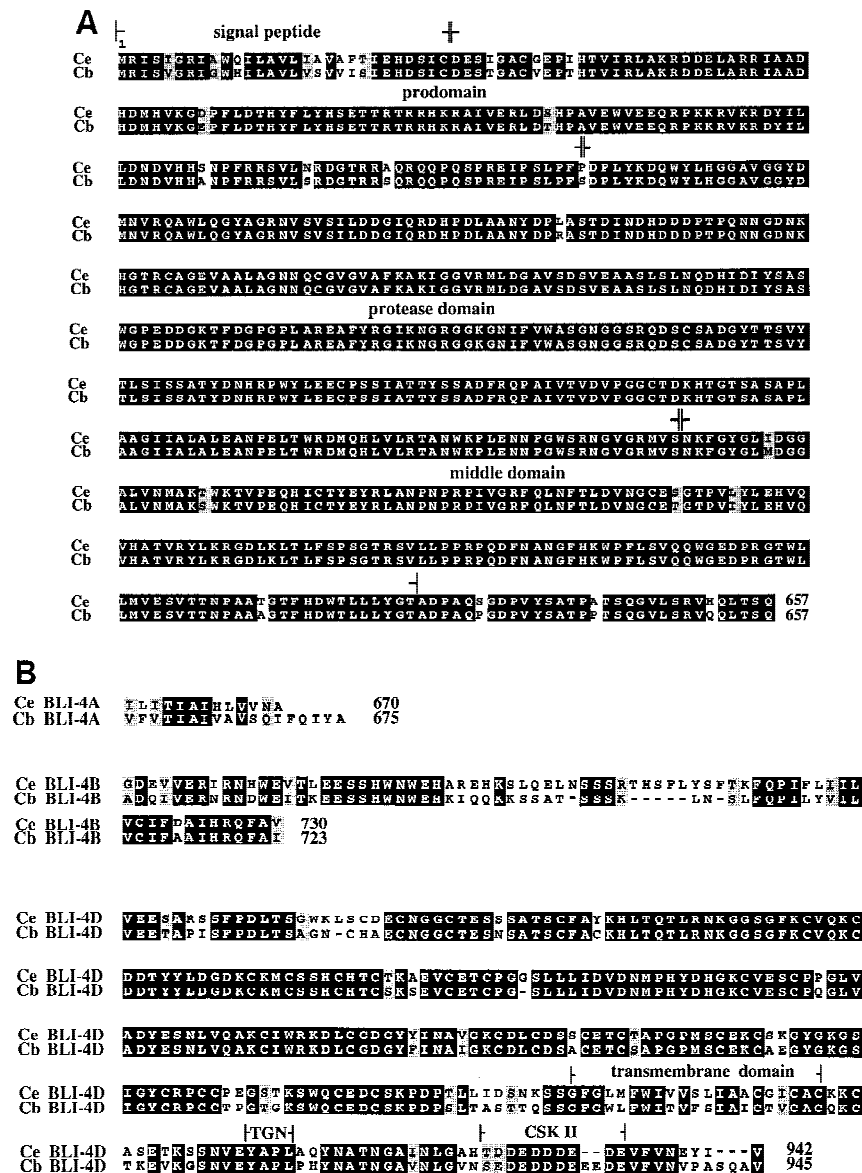
Figure 4 (*A*) Alignment of the predicted BLI-4 protein sequence encoded by the common region of *C. elegans* (Ce) and *C. briggsae* (Cb). Identical residues have been boxed; conserved amino acids are shaded. Residues that are predicted to compose the signal peptide, prodomain, protease domain, and middle domains are shown within the corresponding brackets. (*B*) Alignment of the predicted carboxyl termini of the conserved BLI-4 isoforms from *C. elegans* (Ce) and *C. briggsae* (Cb). Identical residues are boxed; conserved residues are shaded. Conserved amino acid motifs within the carboxyl termini of the BLI-4D isoform are identified within brackets and include a putative transmembrane domain, a *trans*-Golgi network localization signal (TGN) and a casein kinase II phosphorylation site (CSK II).

evidence for additional splicing. The presence of a stop codon 30 bp downstream of the last codon within exon 12 suggests that yk85f9 does have the capacity to encode an isoform that adds an additional nine amino acids to the polypeptide encoded by the common region (Fig. 5). To further examine whether this exon is transcribed, we performed RT–PCR using a primer that anneals to sequences within exon 12 (KRp70) in combination with a primer designed from sequences toward the 3′ end of yk85f9 (KRp170). A PCR product of the expected size of 617 bp was amplified from cDNA synthesized from RNA isolated from wild-type worms (Fig. 6A). Furthermore, RT–PCR using template synthesized from *bli-4(e937)* RNA failed to amplify a product. This result was expected because the *e937* deletion removes sequences that include the annealing site for KRp170. Taken together, these data suggest that an additional *bli-4* isoform is expressed that encodes sequences consisting solely of amino acids encoded by the first 11 coding exons (herein designated as BLI-4E). In addition, the expression of BLI-4E maybe affected by the *e937* deletion which would remove 381 bp within the 3′ UTR of the transcript, including a putative polyadenylation site (Fig. 5). The corresponding nucleotide sequence of *Cb–bli-4* in the region depicted as exon E is not well conserved which is typical of most 3′ UTR sequences, so it is not known whether the common region expresses a similar gene product in this species.

this region of the gene encoded an exon was further supported by the isolation of an expressed sequenced tag (EST) called yk85f9, which although only the 3′ sequence was available, matched the complementary strand of this region. Additional analysis of yk85f9 (obtained from Yuji Kohara, National Institute of Genetics, Mishima, Japan) showed it to contain an additional 615 bp extending 3′ of exon 12, including the splice donor and continuing to nucleotide 13988 with no

Alignment of sequences between the common region and the exon specific to BLI-4B showed four blocks of conserved sequences ranging in length from ~350 to 500 bp (horizontal boxes labeled F–I in Fig. 2). Upon closer examination we found that each of these four blocks contained signatures that would suggest that they encode individual coding exons. Specifically, our criteria for considering these blocks as putative exons was such that each region of homology contained
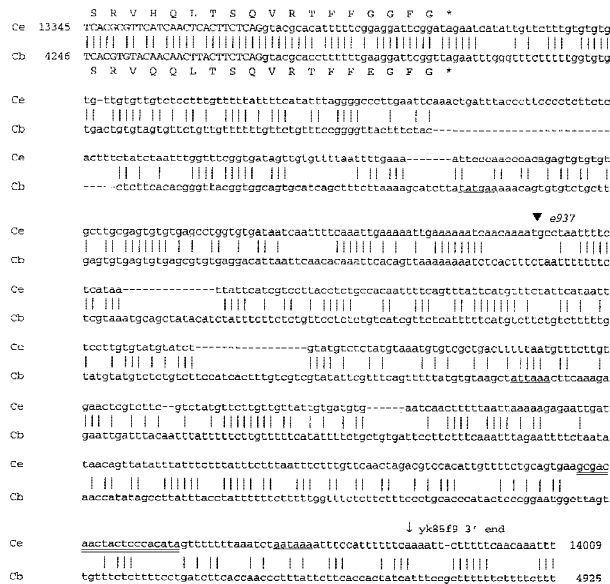
Figure 5 Alignment of sequences at the 3′ end of the common region that may encode the carboxyl terminus of an alternate *bli-4* isoform. The coding nucleotides within exon 12 are shown in uppercase, with the conceptual translation of the predicted carboxyl terminus shown in single-letter code above the middle base of each codon. (▼) The left breakpoint of the 3325-bp *e937* deletion. The site of annealing for primer KRp170 used in the RT–PCR experiments is double underlined. The 3′ end of the EST yk85f9 is indicated (↓). Putative polyadenylation sites have been underlined in both the *Ce–bli-4* (Ce) and *Cb–bli-4* (Cb) sequences.

a conserved splice acceptor that agrees with known *C. elegans* consensus sequences (Blumenthal and Steward 1997) followed by an open reading frame (ORF) which, if spliced to the common region exons, would be in-frame with the first 657 amino acids (Fig. 7). The first region did not adhere strictly to our criteria, but as we show later appears to constitute an expressed exon. This region, designated exon F, is conserved poorly between the two species and is predicted to encode 67 amino acids in *C. elegans* compared to 71 residues in *C. briggsae* with an overall identity of 44%. The predicted exon G of *Ce–bli-4* can encode 78 amino acids as compared to 80 residues for the corresponding *Cb–bli-4* region with ~84% overall identity. Region H is predicted to encode 89 amino acids in both species and shares ~80% identity. Region I can encode 59 amino acids in both species and shares 81% identity. Within the conserved ORF encoded by region H are two putative motifs that are also found in the the BLI-4D isoform. The first 26 residues of region H are mainly hydrophobic and may constitute a transmembrane domain. In addition, a putative casein kinase II phosphorylation site is also found within the conserved H sequences. Experiments on the membrane-bound form of human furin have shown that phosphorylation of a carboxy-terminal casein kinase II site regulates trafficking of the

convertase between the cell surface and the *trans*-Golgi network (Molloy et al. 1994; Bosshart et al. 1995; Jones et al. 1995; Schäfer et al. 1995). Conservation of the phosphorylation site within the carboxyl termini of the predicted BLI-4D and BLI-4H gene products suggests that the localization of these two isoforms may be controlled in a similar fashion to human furin.

We used RT–PCR to test whether these regions are transcribed and spliced to the common region. Our approach was to perform PCR using a primer that anneals to exon 12 in the common region in combination with one primer specific for each putative exon. Because the predicted F, G, and H exons are within the region that is deleted in *e937*, RT–PCR was performed using template synthesized from both *C. elegans* wild-type and *e937* poly(A)⁺ RNA. The results showed that all of the newly identified exons, including exon F, are expressed at the RNA level in wild-type *C. elegans*. As predicted, expression of transcripts containing exons E, F, G, and H was missing in RNA prepared from *e937* animals (Fig. 6A), although products specific for *bli-4D* and *bli-4I* which reside outside of the deletion were obtained. These results suggest that the *e937* mutation is more complicated than we first reported (Thacker et al. 1995), and the blistered phenotype arises from the
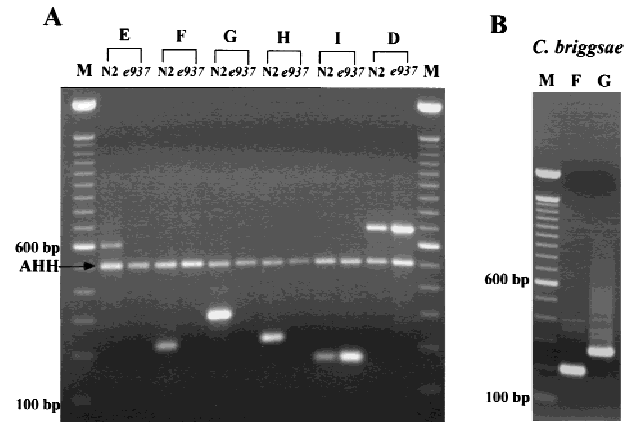

Figure 6 RT–PCR analysis. (*A*) Examination of *bli-4* E, F, G, H, and I expression in *C. elegans* N2 and *e937* animals. Amplification was performed using a primer that annealed to exon 12 in combination with a primer specific for each newly predicted exon and *bli-4D* which resides outside of the *e937* deletion. Products were amplified for each predicted exon from N2 template RNA, but were missing for exons E (617 bp), F (219 bp), G (300 bp), and H (224 bp) using *e937* RNA. The 517-bp product seen in all lanes is the result of amplification using primers designed from the gene encoding *S*-adenosyl homocysteine hydrolase, used as a positive control. (*B*) Examination of *Cb–bli-4F* and *bli-4G* expression. RNA from *C. briggsae* animals was used for RT–PCR analysis to determine whether poorly conserved putative F exon was expressed in this species. As can be seen in lane F, amplification using a common region primer in combination with the F specific primer yielded a product of the expected size (199 bp). As a control, we were also able to demonstrate expression of the highly conserved G exon by amplifying the expected 272-bp product (lane G).

Figure 7 Alignment of conserved sequences that are predicted to encode the carboxyl termini of additional *bli-4* isoforms. The predicted unique exons have been designated *bli-4F*, *bli-4G*, *bli-4H*, and *bli-4I*. Exons are shown in uppercase; the predicted translation is shown in one-letter code above the middle base of each codon for the *C. elegans* sequence. Only amino acids that are not conserved in the predicted *C. briggsae* translation are shown below the corresponding nucleotide sequence. Sequences that may comprise a casein kinase II phosphorylation site in exon H have been underlined.

loss of expression of the BLI-4 A, E, F, G, and H isoforms.

Although the RT–PCR result had shown that the F exon was transcribed in *C. elegans*, the discrepancy of this region between the two species, such as the apparent lack of a conserved splice acceptor and poor conservation within the individual ORFs prompted us to test whether this isoform was also expressed in *C. briggsae*. Using an approach identical to that described above, we were able to demonstrate that the proposed F exon is indeed transcribed in *C. briggsae* as well as the highly conserved G exon as shown by RT–PCR ampli-

fication of the expected 199-bp (Fig. 6B, lane F) and 272-bp (lane G) products.

Conservation of *bli-4* Functional Activity

Phenotypic rescue of three *C. elegans bli-4* mutants was performed to determine interspecies conservation of gene function and regulation. Transgenic animals were established that contained transforming DNA in the form of extrachromosomal arrays containing the fosmid G06P23 or an overlapping clone, G26K16. The latter fosmid includes all *Cb–bli-4* coding sequences

and extends ~8 kb farther 5′ than G06P23 (Fig. 2B). Transfer of the array carrying G26K16 by genetic crosses resulted in complete phenotypic rescue of two different classes of *bli-4* lethal mutants, *h791* and *s90*, as well as rescue of the blistered phenotype of *e937* animals. In contrast, G06P23 arrays rescued the lethal mutants only partially (embryos hatched but arrested as late L_1 larvae), and partially reduced blistering of *e937* animals from 90% to 52%. The explanation for the discrepancy in rescuing ability of these two fosmids can be attributed to the additional 5′ sequences contained in G26K16, evidence for which is presented below. The rescue results led us to conclude that *Ce–bli-4* and *Cb–bli-4* are functionally equivalent, and that the *cis*-regulatory elements of the *C. briggsae* gene are conserved sufficiently to direct expression in the appropriate temporal and spatial fashion.

Identification of Putative Transcriptional Elements

Significant nucleotide homology between *Ce–bli-4* and *Cb–bli-4* was found at the 5′ end of the genes (depicted as diagonally hatched boxes in Fig. 2A). Like other genes that have been studied between the two species, this conservation of sequences within the 5′ flanking region may constitute promoter elements that are required for transcriptional regulation. Seven regions of homology (labeled I–VII in Fig. 2) can be aligned. It is not likely they encode a separate gene because several gaps were needed to achieve the alignment and no obvious ORFs could be recognized. Interestingly, regions I–IV were identified in the sequence of fosmid G25K01 (which overlaps the first 6187 nucleotides of G06P23 and extends further 5′; see Fig. 2B), but are not contained within G06P23. This finding may explain why G06P23 was unable to rescue the *Ce–bli-4* mutants fully. To test whether some of these regions of conservation serve as transcriptional elements we took advantage of a plasmid construct containing a *bli-4–lacZ* fusion (pCeh320) which had been shown previously to direct the expression of β-galactosidase in hypodermal and neural tissues (Thacker et al. 1995). pCeh320 contains a 5219 bp *Xba*I–*Cla*I fragment (including regions I–VII) with the first eight codons of *Ce–bli-4* cloned in-frame with β-galactosidase (Fig. 8). A series of deletions were generated that removed sequences within the 5′ region of *Ce–bli-4* (Figure 8) and these constructs were used to establish transgenic strains that were examined for β-galactosidase activity (Fire et al. 1990). Two constructs in

which sequences upstream of the *Xho*I site (pCeh254), or upstream of a conserved *Bgl*II site (pCeh288) were deleted failed to show any expression of the *lacZ* gene. Deletion of sequences between a *Blp*I site and *Xho*I which removed regions V and VI (pCeh291) gave low levels of β-galactosidase activity in hypodermal cells, predominantly at late embryogenesis and during each of the four larval stages with no observable expression in adults (data not shown). Finally, transgenic animals that carry pCeh289, in which only region VI had been deleted, showed markedly decreased *lacZ* expression in all stages except for adults, similar to that observed for pCeh291. Region VI is deleted in both pCeh289 and pCeh291 suggesting that sequences within this domain may be required for expression during the adult stage. Although the deletions generated in this series of constructs are relatively large (1–4 kb), the expression results clearly show that the 5′ region of *Ce–bli-4* contains sequences necessary for controlling the expression of the *bli-4–lacZ* fusion. The preliminary promoter analysis described here, together with the rescue data suggests that the proposed elements (I–VII) are necessary and essential for expression of the *bli-4* gene.

DISCUSSION

A comprehensive study of complex genes is limited by the sensitivity of techniques available to the investigator and often relies on extensive screening of cDNA libraries or exon-trapping experiments to identify cod-
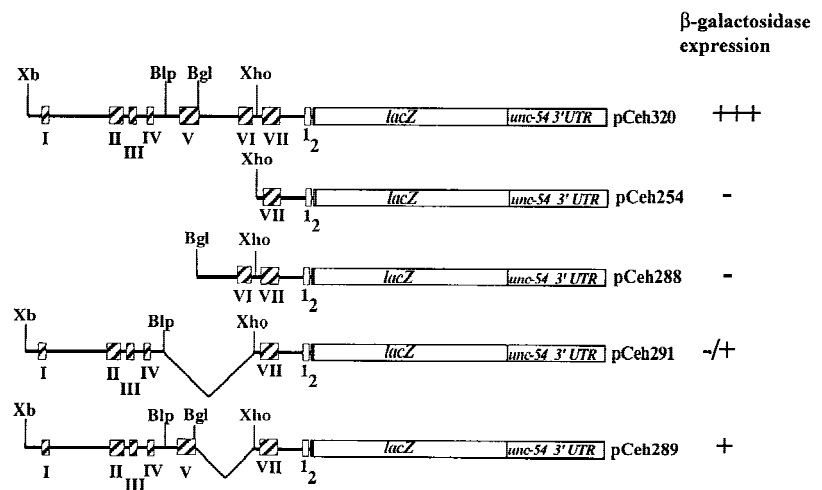

Figure 8 Interspecies sequence homology at the 5′ end of the *bli-4* gene. Sequence alignment of regions I–VII within the respective 5′ ends of the *Ce–bli-4* (Ce) and *Cb–bli-4* (Cb) genes are available online at www.genome.org. For clarity, a conserved *Bgl*II site is shown in boldface lettering. A schematic representation of constructs is used to examine the transcriptional regulation of a *bli-4–lacZ* reporter gene. The plasmid pCeh320 contains 5219 bp of the *Ce–bli-4* 5′ end, in which a series of deletion clones have been constructed. The relative level of β-galactosidase expression exhibited by transgenic strains that carry each plasmid is shown as high (+++), low (+), variable (+/−), or not observed (−). (Xb) *Xba*I; (Blp) *Blp*I; (Bgl) *Bgl*II; (Xho) *Xho*I.

ing elements. Here we demonstrate an approach that involves the use of comparative genomic analysis to delineate the structural organization of the proprotein convertase gene, *bli-4*, using sequence information generated from related nematode species. Our results show that the *bli-4* gene is highly conserved between *C. elegans* and *C. briggsae* both in structure and sequence content, such that a common set of exons including those that encode the catalytic domain, are alternately spliced to 3′ exons. Conservation of protein structure is also highly conserved, particularly within the catalytic domain, which is reflected in the ability of *Cb–bli-4* to rescue the mutant phenotype of several *C. elegans bli-4* mutants. An interesting difference between the two species is the complete absence of the exon unique to the *bli-4C* isoform in the *C. briggsae* gene. As *Cb–bli-4* was able to rescue the lethal and blistered phenotypes fully, the absence of a Cb–BLI-4C isoform suggests that this gene product is not essential during late embryogenesis or for the production of the adult cuticle. Several regions of homology exist outside of coding exons. At the 5′ end of the gene, at least seven conserved regions (I–VII) were shown to be essential for controlling the expression of a *lacZ* reporter gene. These results suggest that the regulation of expression of *bli-4* may be complex. One possible model to explain how the processing of substrate proproteins by the kex2/subtilsin-like convertases may be controlled is by their concomitant temporal expression (Seidah et al. 1994). However, a comparison of the *bli-4* conserved 5′ sequences with genes that may encode candidate substrates for processing by the BLI-4 convertases, such as the cuticle collagen genes *bli-1*, *dpy-5*, and others predicted by the *C. elegans* Genome Consortium, did not reveal any extensive nucleotide homologies although small regions of homology were noticed.

Conservation of the noncoding first exon was also observed and to date, this feature is unique to *bli-4* among those genes in which an interspecies comparison has been made, further implicating functional significance for this sequence. Although for the most part intronic sequences were not conserved and mostly missing or shorter in *Cb–bli-4*, the intervening sequence designated region VIII did show stretches of nucleotide homology suggesting a functional role either in the regulation of transcription or possibly an involvement in the control of splicing events at the 3′ end of the gene. Recent results have demonstrated that the *mec-8* gene product encodes a *trans*-acting factor that may be involved in alternate splice site choice by binding specific elements in pre-mRNAs (Lundquist et al. 1996), suggesting that such mechanisms exist in *C. elegans*.

The use of comparative genomic analysis has proved to be a powerful tool for the elucidation of the *bli-4* gene structure. Specifically, this approach has led to the identification of five additional *bli-4* isoforms. To date, *bli-4* represents one of the more complex of *C. elegans* genes, encoding at least nine products as determined from this analysis. These results emphasize the efficacy of the method for retrieving structural and functional information of complex genes in which some transcripts are clearly underrepresented in cDNA libraries. cDNAs and ESTs have only been identified for the *bli-4A*, *bli-4B*, *bli-4C*, *bli-4D*, and *bli-4E* products, possibly reflecting temporal expression patterns for the nine isoforms. Preliminary evidence suggests that the *bli-4* gene products are expressed at different times and abundance during development (C. Thacker, unpubl.) which may explain why the newly identified exons have not been discovered previously as ESTs. This analysis also has an impact on the interpretation of the blistered *e937* mutation. Initially our data were consistent with the lesion affecting the expression of only the BLI-4A isoform. This work has now shown that the deletion also abrogates expression of the F, G, H, and perhaps E isoforms.

The reason for the conservation of many variant forms of the BLI-4 convertase may be a consequence of their apparent role in the processing of cuticle collagens. The *C. elegans* genome sequencing consortium has identified 170 collagen genes of which the majority may be required for the production of the cuticle. A growing amount of evidence has shown that there is a complex differential expression of individual cuticle collagen genes at each developmental stage (Cox and Hirsh 1985; Kramer et al. 1985; Johnstone and Barry 1996). The result of this differential expression is the generation of stage-specific cuticles that differ in collagen composition reflecting possibly a need for functional specialization (Cox et al. 1981; for review, see Kramer 1997). Therefore, it is conceivable that several forms of BLI-4 are required to provide the necessary specificity to process the many procollagens correctly. The production of variant isoforms that differ at their carboxyl termini suggests a compelling role for these domains in colocalizing the convertases with their appropriate substrate(s). Understanding the function of these enzymes within the nematode may uncover conserved regulatory mechanisms that influence the activity of proprotein convertase family members in all metazoa. This may be of particular value with regard to the apparent important clinical relevance of the convertase family in the control of several physiological processes, the disruption of which may lead to certain pathologies in vertebrates.

METHODS

Isolation of the *C. briggsae bli–4* Gene

Fosmids that contain *C. briggsae* genomic DNA from the *bli-4*

locus were isolated and sequenced using the following approach. Fifty nanograms of a 1.0-kb *Hin*DIII fragment corresponding to a region distal of the protease domain of the *Ce–bli-4* gene was labeled radioisotopically by random priming using commercially available reagents and protocols (Prime-It kit, Stratagene). Probe DNA was denatured and added to a 22 × 22-cm nylon membrane (Hybond N⁺, Amersham) carrying 36,000 individual *C. briggsae* fosmid DNA samples deposited on the membrane as small spots (the membrane is hereafter referred to as the fosmid grid). The 36,000 DNA samples represent 18,000 independently derived *C. briggsae* fosmid clones (M. Marra, unpubl.). Identical filters can be obtained from Genome Systems (St. Louis). Prehybridization of the fosmid grid was achieved by a 4-hr incubation with agitation at 65°C, in 50 ml of prehybrization solution (6× SSC, 0.5% SDS, 5× Denhardt's reagent). Hybridization of the *Ce–bli-4* probe to the *C. briggsae* fosmid grid was allowed to proceed in 50 ml of prehybridization buffer for 16 hr at 65°C in a Robbins Scientific rotary hybridization oven. Filters were removed to 150 ml of a solution consisting of 2× SSC, 0.5% SDS (wash buffer). The filter was incubated, with agitation, in room temperature wash buffer for 5 min. A 15-min room temperature wash step was then performed with a fresh aliquot of wash buffer. This wash was followed by two consecutive 30-min washes, with agitation, in 65°C wash buffer. *C. briggsae* fosmid DNAs hybridizing to the *Ce–bli-4* probe were identified by exposure of the washed filter to Kodak Biomax MS film at room temperature for 45 min. Intensifying screens were not required. More than 98% of the fosmid clones present on the *C. briggsae* fosmid grid have been fingerprinted by agarose-gel-mediated restriction analysis (Marra et al. 1997) and the fingerprints entered into a computer database (M. Marra, unpubl.). Contigs of overlapping clones can be assembled readily using as a nucleating point any clone or clones scored as positive in hybridization experiments which employ the fosmid grids. Thus, the fingerprints of the clones identified in the hybridization experiment described above were consulted and used to construct a contig, from which a minimal tiling path of clones were selected for sequencing at Washington University Genome Sequencing Center. Fosmid DNA isolation, construction of M13 sequencing libraries and DNA sequencing were as described (Chissoe et al. 1997).

Computer Analysis

DNA sequences were analysed using Dotter (Sonnhammer and Durbin 1996), the GCG Sequence Analysis Software Package (Devereux et al. 1984), and GeneFinder (P. Green and L. Hillier, unpubl.). Sequence data for the *C. briggsae bli-4* fosmids can be obtained from the Washington University Genome Sequencing Center web site: ftp://genome.wustl.edu/pub/gsc1/sequence/st.louis/briggsae/finish/lin31/G06P23.seq; ftp://genome.wustl.edu/pub/gsc1/sequence/st.louis/briggsae/preliminary/lin31/G25K01.seq. The sequence for G25K01 was reverse complemented for the alignments, as it was found that the *bli-4* gene was encoded in the opposite orientation to that presented in the ftp site.

Expression Studies

RT–PCR was performed using a 3′ RACE system (Life Technologies) using an approach similar to that described previously (Thacker et al. 1995). First-strand cDNA was synthesized from 1 µg of *C. elegans* total RNA or poly(A)⁺ RNA using the universal primer supplied and conditions described by the

manufacturer. Amplification of cDNA was performed using the primer KRp6 or KRp70, which anneal to sequences toward the end of the common region in exon 12, in combination with one antisense primer specific for each putative exon labeled D, E, F, G, H, and I in Figure 1. The sequences of the oligonucleotides are defined below. Reaction mixtures contained a 1-µl aliquot of the cDNA template, 50 pmoles of each primer, 200 mM dNTPs, and 2.5 units of *Taq* DNA polymerase (Promega) in 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 0.1% Triton X-100, and 1.5 mM MgCl$_2$, in a final volume of 25 µl. Reactions were performed in a Perkin-Elmer Cetus thermal cycler for 32 cycles of denaturation (94°C, 45 sec), annealing (54–58°C, 30 sec), and extension (72°C, 1 min) followed by extension at 72°C for 7 min. The sizes of the expected amplification products were 723 bp (*bli-4D*), 617 bp (*bli-4E*), 219 bp (*bli-4F*), 300 bp (*bli-4G*), 224 bp (*bli-4H*), 169 bp (*bli-4I*), and 5/7 bp [control gene, *S-adenosyl homocysteine hydrolase* (AHH)]. The sequence of oligonucleotides used in the experiments is as follows: Common region primer KRp6, 5′-CTACTCGGCTACTCCTGC-3′; common region primer KRp70, 5′-AGTTCTCCACGCGTTCATC-3′; *blisterase E* (*bli-4E*) KRp170, 5′-TATGTGGGAGTAGTTGTCGC-3′; *blisterase F* (*bli-4F*) KRp248, 5′-GTAAGTTGGACGATGGTTAC-3′; *blisterase G* (*bli-4G*) KRp235, 5′-CTAATGCTTGTATCGTCGTC-3′; *blisterase H* (*bli-4H*) KRp236, 5′-GATAGTGAAGTCATTGCAAC-3′, *blisterase I* (*bli-4I*) KRp171, 5′-TGGCTATGCCAGTCTGATAC-3′; *blisterase D* (*bli-4D*) KRp72, 5′-GAACATCAATCCAAATC-CAGA-3′; *ahh* gene KRp12, 5′-CGTCCGTTCTTGAGGGTG-3′; *ahh* gene KRp14, 5′-CTAAGATGCTCGCCAAGG-3′.

Expression of *Cb–bli-4F* and *Cb–bli-4G* was examined using RNA extracted from the *C. briggsae* strain G16 using the procedure described previously (Thacker et al. 1995). RT–PCR analysis was performed using a primer specific for the *C. briggsae* common region, KRp311, 5′-CACGTGTACAACAACT-TACT-3′ in combination with the *Cb–bli-4F*-specific primer KRp310, 5′-TGGAAGATAAATCGATGTGT-3′, or KRp235 which can anneal and amplify the *Cb–bli-4G* unique exon. The expected amplification products were 199 bp for *bli-4F* and 272 bp for *bli-4G*.

Mutant Rescue Experiments

Fosmid DNA was prepared using a procedure described previously (Chissoe et al. 1997). DNA injection solution contained 80 µg/ml of pCes1943 mixed with 10 µg/ml of G06P23. pCes1943 carries a dominant marker, *rol-6(su1006)* which imparts a Roller phenotype to transgenic animals. Injections into wild-type animals was performed by T. Ha (Simon Fraser University) using the procedure of Janke et al. (1997), modified from Mello et al. (1991). The extrachromosomal arrays from two stable lines transformed with G06P23 and pCes1943, and a single line transformed with fosmid G26K16, an overlapping clone that starts ~8 kb upstream of the *Cb–bli-4* first exon and includes all of the *bli-4* coding region, were used to examine rescue of the *bli-4* lethal mutant phenotype of alleles *h791* and *s90* and the blistered allele *e937*.

Identification of *bli–4* Regulatory Elements

Preliminary investigation of putative promoter elements that may regulate *bli-4* expression identified at the 5′ end of the gene were examined by deletion analysis of the corresponding region of a *Ce–bli-4–lacZ* reporter construct, pCeh320. This plasmid contains a 5219-bp *Xba*I–*Cla*I fragment from *Ce–bli-4* including the first four codons cloned in-frame with

lacZ. Transgenic animals transformed with pCeh320 and pRF4 showed expression of β-galactosidase in nuclei of hypodermal and neuronal origin (Thacker et al. 1995). Deletion derivatives of pCeh320 were constructed by digestion of the plasmid using a combination of two restriction enzymes both of which cut only once in the 5′-untranslated region of the *bli-4* sequence followed by recircularization. Four constructs were prepared as follows (see Fig. 8B): pCeh254 contains a 4096-bp deletion *Xba*I–*Xho*I; pCeh288 contains a 3024-bp deletion *Xba*I–*Bgl*II; pCeh291 contains a 1643-bp deletion *Blp*I–*Xho*I; pCeh289 contains a 1072-bp deletion *Bgl*II–*Xho*I. Each of these plasmids was injected by D. Janke (Simon Fraser University) into wild-type worms at 50 µg/ml along with 80 µg/ml of plasmid pRF4 (Mello et al. 1991). Only transgenic animals that transmitted extrachromosomal arrays heritably were maintained for additional analysis. Extrachromosomal arrays were integrated by exposing transgenic animals to UV irradiation for 20 sec at 257 nm using a germicidal UV lamp. Integrated lines were outcrossed with wild-type males at least twice and then examined for β-galactosidase activity by staining with X-Gal (Fire et al. 1990).

ACKNOWLEDGMENTS

REFERENCES

Berks, M. and the *C. elegans* Genome Mapping and Sequencing Consortium. 1995. The *C. elegans* genome sequencing project. *Genome Res.* **5:** 99–104.

Blumenthal, T. and K. Steward. 1997. RNA processing and gene structure. In Caenorhabditis elegans *II* (ed. D.L. Riddle, T. Blumenthal, B.J. Meyer, and J.R. Priess), pp. 117–145. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Bosshart, H., J. Humphrey, E. Designan, J. Davidson, J. Drazba, L. Yuan, V. Oorschot, P.J. Peters, and J. Bonifacino. 1995. The cytoplasmic domain mediates localization of furin to the *trans*-Golgi network en route to the endosomal/lysosomal system. *J. Cell Biol.* **126:** 1157–1172.

C. elegans Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: A platform for investing biology. *Science* **282:** 2012–2018.

Chan, S.J., A.A. Oliva, Jr., J. LaMendola, A. Grens, H. Bode, and D.F. Steiner. 1992. Conservation of the prohormone convertase gene family in metazoa: Analysis of cDNAs encoding a PC3-like protein from hydra. *Proc. Natl. Acad. Sci.* **89:** 6678–6682.

Chissoe, S.L., M.A. Marra, L. Hillier, R. Brinkman, R.K. Wilson, and R.H. Waterston. 1997. Representation of cloned genomic sequences in two sequencing vectors: Correlation of DNA sequence and subclone distribution. *Nucleic Acids Res.* **25:** 2960–2966.

Cox, G.N. and D. Hirsh. 1985. Stage-specific patterns of collagen gene expression during development of *Caenorhabditis elegans*. *Mol. Cell. Biol.* **5:** 363–372.

Cox, G.N., S. Straprans, and R.S. Edgar. 1981. The cuticle of *Caenorhabditis elegans*. II. Stage-specific changes in ultrastructure and protein composition during postembryonic development. *Dev. Biol.* **86:** 456–470.

Creemers, J.W.M., E.F. Usac, N.A. Bright, J.-W. Van de Loo, E. Jansen, W.J.M. Van de Ven, and J.C. Hutton. 1996. Identification of a transferable sorting domain for the regulated pathway in the prohormone convertase PC2. *J. Biol. Chem.* **271:** 25284–25291.

De Bie, I., D. Savaria, A.J.M. Roebroek, R. Day, C. Lazure, W.J.M. Van de Ven, and N.G. Seidah. 1995. Processing specificity and biosynthesis of the *Drosophila melanogaster* convertases dfurin1, dfurin-CRR, dfurin1-X, and dfurin2. *J. Biol. Chem.* **270:** 1020–1028.

De Bie, I., M. Marcinkiewicz, D. Malide, C. Lazure, K. Nakayama, M. Bendayan, and N.G. Seidah. 1996. The isoforms of proprotein convertase PC5 are sorted to different subcellular compartments. *J. Cell Biol.* **135:** 1261–1275.

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

Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12:** 387–395.

Emmons, S.W., M.R. Klass, and D. Hirsh. 1979. Analysis of the constancy of DNA sequences during development and evolution of the nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* **76:** 1333–1337.

Fire, A., S. White Harrison, and D. Dixon. 1990. A modular set of *lacZ* fusion vectors for studying gene expression in *Caenorhabditis elegans*. *Gene* **93:** 189–198.

Heine, U. and T. Blumenthal. 1986. Characterization of regions of the *Caenorhabditis elegans* X chromosome containing vitellogenin genes. *J. Mol. Biol.* **188:** 301–312.

Heschl, M.F.P. and D.L. Baillie. 1990. Functional elements and domains inferred from sequence comparisons of a heat shock gene in two nematodes. *J. Mol. Evol.* **31:** 3–9.

Hodgkin, J., R.H.A. Plasterk, and R.H. Waterston. 1995. The nematode *Caenorhabditis elegans* and its genome. *Science* **270:** 410–414.

Janke, D.L., J.E. Schein, T. Ha, N.W. Franz, N.J. O'Neil, G.P. Vatcher, H.I. Steward, L.M. Kuervers, D.L. Baillie, and A.M. Rose. 1997. Interpreting a sequenced genome: Toward a cosmid transgenic library of *Caenorhabditis elegans*. *Genome Res.* **7:** 974–985.

Johnstone, I.L. and J.D. Barry. 1996. Temporal reiteration of a precise gene expression pattern during nematode development. *EMBO J.* **15:** 3633–3639.

Jones, B.G., L. Thomas, S.S. Molloy, C.D. Thulin, M.D. Fry, K.A. Walsh, and G. Thomas. 1995. Intracellular trafficking of furin is modulated by the phosphorylation state of a casein kinase II site in its cytoplasmic tail. *EMBO J.* **14:** 5869–5883.

Kiefer, M.C., J.E. Tucker, R. Joh, K.E. Landsberg, D. Saltman, and P.J. Barr. 1991. Identification of a second human subtilisin-like protease gene in the *fes/fps* region of chromosome 15. *DNA Cell Biol.* **10:** 757–769.

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

Kramer, J. 1997. Extracellular matrix. In Caenorhabditis elegans *II* (ed. D.L. Riddle, T. Blumenthal, B.J. Meyer, and J.R. Priess), pp. 117–145. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Kramer, J.M., G.N. Cox, and D. Hirsh. 1985. Expression of the *Caenorhabditis elegans* collagen genes *col-1* and *col-2* is developmentally regulated. *J. Biol. Chem.* **260:** 1945–1951.

Krause, M., S.W. Harrison, S.-Q. Xu, L. Chen, and A. Fire. 1994. Elements regulating cell- and stage-specific expression of the *C. elegans* MyoD family homolog *hlh-1*. *Dev. Biol.* **166:** 133–148.

Kuwabara, P.E. and S. Shah. 1994. Cloning by synteny: Identifying *C. briggsae* homologues of *C. elegans* genes. *Nucleic Acids Res.* **22:** 4414–4418.

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

Lee, Y.H., X.Y. Huang, D. Hirsh, G.E. Fox, and R.M. Hecht. 1992. Conservation of gene organization and trans-splicing in the glyceraldehyde-3-phosphate dehydrogenase-encoding genes of *Caenorhabditis elegans*. *Gene* **121:** 227–235.

Lundquist, E.A., R.K. Herman, T.M. Rolgalski, G.P. Mullen, D.G. Moerman, and J.E. Shaw. 1996. The *mec-8* gene of *C. elegans* encodes a protein with two RNA recognition motifs and regulates alternative splicing of *unc-52* transcripts. *Development* **122:** 1601–1610.

Maduro, M. and D. Pilgrim. 1996. Conservation of function and expression of *unc-119* from two *Caenorhabditis* species despite divergence of non-coding DNA. *Gene* **183:** 77–85.

Marra, M.A., T.A. Kucaba, N.L. Dietrich, E.D. Green, B. Brownstein, R.K. Wilson, K.M. McDonald, L.W. Hillier, J.D. McPherson, and R.H. Waterston. 1997. High throughput fingerprint analysis of large-insert clones. *Genome Res.* **7:** 1072–1084.

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

Molloy, S.S., L. Thomas, J.K. VanSlyke, P.E. Stenberg, and G. Thomas. 1994. Intracellular trafficking and activation of the furin proprotein convertase: Localization to the TGN and recycling from the cell surface. *EMBO J.* **13:** 18–33.

Nigon, V. and E.C. Dougherty. 1949. Reproductive patterns and attempts at reciprocal crossing of *Rhabditis elegans* Maupas 1900, and *Rhabditis briggsae*. *J. Exp. Zool.* **112:** 485–503.

Peters, K., J. McDowall, and A.M. Rose. 1991. Mutations in the *bli-4 (I)* locus of *Caenorhabditis elegans* disrupt both adult cuticle and early larval development. *Genetics* **129:** 95–102.

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

Roebroek, A.J.M., J.W.M. Creemers, I.G.L. Pauli, T. Bogaert, and W.J.M. Van de Ven. 1993. Generation of structural and functional diversity in furin-like proteins in *Drosophila melanogaster* by alternative splicing of the *Dfur-1* gene. *EMBO J.* **12:** 1853–1870.

Schäfer, W., A. Stroh, S. Berghofer, J. Seiler, M. Vey, M.L. Kruse, H.F. Kern, H.D. Klenk, and W. Garten. 1995. Two independent targeting signals in the cytoplasmic domain determine *trans*-Golgi network localization and endosomal trafficking of the proprotein convertase furin. *EMBO J.* **14:** 2424–2435.

Seidah, N. and M. Chrétien. 1992. Proprotein and prohormone convertases of the subtilisin family. *Trends Endocrinol. Metab.* **3:** 133–140.

Seidah, N.G., R. Day, J. Hamelin, A. Gaspar, M.W. Collard, and M. Chrétien. 1992. Testicular expression of PC4 in the rat: Molecular diversity of a novel germ cell-specific kex2/subtilisin-like proprotein convertase. *Mol. Endocrinol.* **6:** 1559–1570.

Seidah, N., M. Chrétien, and R. Day. 1994. The family of subtilisin/kexin like pro-protein and pro-hormone convertases: Divergent or shared functions. *Biochimie* **76:** 197–209.

Snutch, T.P. 1984. "Molecular and genetic analysis of the heat shock response of *Caenorhabditis elegans*." Ph.D. thesis, Simon Fraser University, Burnaby, B.C.

Sonnhammer, E.L.L. and R. Durbin. 1996. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **167:** 1–10.

Thacker, C., K. Peters, M. Srayko, and A.M. Rose. 1995 The *bli-4* locus of *Caenorhabditis elegans* encodes structurally distinct kex2/subtilisin-like endoproteinases essential for early development and adult morphology. *Genes & Dev.* **9:** 956–971.

Xue, D., M. Finney, G. Ruvkin, and M. Chalfie. 1992. Regulation of the *mec-3* gene by the *C. elegans* homeoproteins UNC-86 and MEC-3. *EMBO J.* **11:** 4969–4979.

Zucker-Aprison, E. and T. Blumenthal. 1989. Potential regulatory elements of nematode vitellogenin genes revealed by interspecies sequence comparison. *J. Mol. Evol.* **28:** 487–496.