J Mol Evol (1990) 31:3–9

Journal of
Molecular Evolution
© Springer-Verlag New York Inc. 1990

Functional Elements and Domains Inferred from Sequence Comparisons of a Heat Shock Gene in Two Nematodes

Mark F.P. Heschl* and David L. Baillie

Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

Summary. *Caenorhabditis elegans* and *Caenorhabditis briggsae* are two closely related nematode species that are nearly identical morphologically. Interspecific cross-hybridizing DNA appears to be restricted primarily to coding regions. We compared portions of the *hsp-3* homologs, two grp78-like genes, from *C. elegans* and *C. briggsae* and detected regions of DNA identity in the coding region, the 5′ flanking DNAs, and the introns. The *hsp-3* homologs share approximately 98% and 93% identity at the amino acid and nucleotide levels, respectively. Using the nucleotide substitution rate at the silent third position of the codons, we have estimated a lower limit for the date of divergence between *C. elegans* and *C. briggsae* to be approximately 23–32 million years ago. The 5′ flanking DNAs and one of the introns contain elements that are highly conserved between *C. elegans* and *C. briggsae.* Some of the regions of nucleotide identity in the 5′ flanking DNAs correspond to previously detected identities including viral enhancer sequences, a heat shock element, and an element present in the regulatory regions of mammalian grp78 and grp94 genes. We propose that a comparison of *C. elegans* and *C. briggsae* sequences will be useful in the detection of potential regulatory and structural elements.

Key words: *Caenorhabditis elegans* — *Caenorhabditis briggsae* — hsp70 — grp78 — Gene comparison — Evolution — Regulatory elements

* *Present address:* Department of Physiological Chemistry, 587 Medical Sciences Building, 1300 University Avenue, Madison, Wisconsin 53706, USA
Offprint requests to: M.F.P. Heschl

Introduction

The nematode *Caenorhabditis elegans* is well suited to genetic and developmental studies for several reasons. A detailed genetic map has been compiled (Edgley and Riddle 1987), a complete description of the cell lineage has been achieved (Sulston and Horvitz 1977; Sulston et al. 1983), and a library of ordered cosmid clones and yeast artificial chromosomes is nearing completion (Coulson et al. 1986, 1988). With the development of a *C. elegans* transformation system (Fire 1986) it has become feasible to study the regulation of genes on a cellular basis as well as to identify mutations by rescue with the wild-type copy of the gene. Recently, a cross-hybridization technique has been developed to identify coding regions, aiding the alignment of a genetic map with the physical map (Snutch 1984; Prasad and Baillie 1989). Most of the cross-hybridizing DNA sequences detected between *C. elegans* and *Caenorhabditis briggsae* represent functional transcribed elements (Prasad and Baillie 1989).

To identify potential regulatory elements, it is helpful to compare species that have been separated for evolutionary periods long enough for unconstrained sequences to diverge. Such comparisons have been made using several *Drosophila* species, and sequences representing genes and elements important for gene regulation have been identified. For example, a comparison of the 5′ flanking DNA sequences of the hsp82 genes from four different *Drosophila* species (Blackman and Meselson 1986) revealed that most of the conserved features detected were known to be involved in transcriptional control [for example, the TATA and heat shock element (HSE) motifs]. In the *Sgs-3* genes of several *Drosophila* species, a number of conserved elements in

the 5′ flanking DNAs were detected (Martin et al. 1988), some of which are known to be involved in gene regulation. In addition, conserved sequences at the 5′ end of the *Gart*-nested cuticle gene are believed to be involved in developmental expression of the cuticle gene (Henikoff and Eghtedarzadeh 1987).

Caenorhabditis elegans and *C. briggsae*, two closely related species, have been described as "twin species" (Nigon and Dougherty 1949). Although nearly identical morphologically, DNA annealing experiments have suggested that any related sequences in *C. elegans* and *C. briggsae* must differ by at least 20% (Emmons et al. 1979) and that the two species have been evolutionarily separated for tens of millions of years. The divergence observed appears to be restricted primarily to the noncoding regions (Snutch 1984; Heine and Blumenthal 1986; Prasad and Baillie 1989). Therefore, the relatedness of and the relatively high degree of nucleotide divergence between *C. elegans* and *C. briggsae* is ideal for a comparison of homologous DNA sequences to detect potential coding regions and regulatory elements.

During analysis of the hsp70 gene family from *C. elegans*, we characterized a gene, *hsp-3* (Heschl and Baillie 1989), that shared identity with the mammalian grp78 genes. The grp78 gene is a member of the highly conserved hsp70 multigene family (Lindquist and Craig 1988) and encodes a protein located in the endoplasmic reticulum (Lee 1987). The *hsp-3* gene, encoding the hsp70C protein, is constitutively expressed, developmentally regulated, and not heat inducible (Snutch et al. 1988). We previously detected several identities to known regulatory elements including viral enhancers and an HSE (Heschl and Baillie 1989) as well as an element that shares considerable identity with one that is important for the high-level expression of the rat grp78 gene (Resendez et al. 1988; Heschl and Baillie 1989). Our present objective is to determine if a comparison of the *C. elegans* and *C. briggsae* homologs can detect, by DNA sequence conservation and divergence, potential regulatory elements. This paper characterizes part of the *hsp-3* homolog from *C. briggsae*, including the first two-thirds of the coding region and the 5′ flanking DNA, and presents a comparison of the *C. briggsae hsp-3* gene with the *C. elegans hsp-3* gene.

Materials and Methods

Construction of Plasmids. The isolation and characterization of the *C. elegans hsp-3* gene [*hsp-3*(Ce)] has been described elsewhere (Snutch et al. 1988; Heschl and Baillie 1989). The phage CB10 contains the *C. briggsae hsp-3* homolog [*hsp-3*(Cb)] (Snutch 1984). The *hsp-3*(Cb) coding region was delineated by hybridization of the restriction fragments back to the *hsp-3*(Ce) gene. CB10 was digested with HindIII and cloned into the plasmid vector Bluescript⁺ (Stratagene). The plasmid pCbs1 contains the *hsp-3* coding region and 5′ flanking DNAs. pCbs1 was further subcloned by a partial SalI digestion and a complete HindIII digestion to retrieve 5′ flanking DNA. The 3.4-kb fragment encompassing the SalI site nearest the coding region to the HindIII 5′ of the coding region was electroeluted as described in Maniatis et al. (1982) and ligated into Bluescript⁺ cut with SalI and HindIII to give rise to pCbs2. Overlapping plasmid deletions were made using either exonuclease III (Henikoff 1987) or restriction enzymes.

Plasmid DNA Preparation. Plasmid DNA for deletion and sequencing reactions was prepared by the mini alkali lysis method (Maniatis et al. 1982) except that two phenol:chloroform extractions were done, and, after the first ethanol precipitation, the air-dried pellet was resuspended in 0.25 M sodium acetate and reprecipitated with ethanol. After digestion with RNase, the plasmid DNA samples were precipitated with polyethylene glycol (Hattori and Sakaki 1986).

DNA Sequencing and Sequence Analysis. Dideoxy sequencing was performed on denatured plasmid DNA as described by Mierendorf and Pfeffer (1987). DNA sequence analysis was done using the computer program Microgenie (Beckman). Visual inspection and preparation of the DNA sequences for publication were done with the aid of the computer program ESEE (Cabot and Beckenbach 1989). Nucleotide divergence was calculated by subtracting the percent identity from 100. Correction for multiple mutational events was calculated on a Poisson distribution using the formula $-\frac{3}{4}\ln(1 - \frac{4}{3}f)$, where f is equal to the uncorrected percent nucleotide divergence divided by 100.

Results and Discussion

Characterization of the hsp-3 Homolog from a Closely Related Species, C. briggsae

Two phage of one type containing the *C. briggsae hsp-3* homolog [*hsp-3*(Cb)] were isolated from a partial EcoRI genomic library in Charon 4 (Snutch 1984). Only the first two-thirds of the *hsp-3*(Cb) homolog and 5′ flanking DNA were represented in the phage isolated. These data are summarized in Fig. 1. The sequence of the *hsp-3*(Cb) homolog and 5′ flanking DNA is in Fig. 2 and is aligned against the *C. elegans hsp-3* [*hsp-3*(Ce)] gene (Heschl and Baillie 1989) for maximum identity.

Comparison of the hsp-3 Homologs from C. elegans and C. briggsae

The *hsp-3* coding regions are highly conserved between the two species sharing 92.6% identity at the nucleotide level and 98% identity at the amino acid level (Fig. 2). A summary of the nucleotide substitutions is presented in Table 1. The *C. briggsae* hsp70C protein has a hydrophobic leader sequence from residue 1 to 17 similar to the *C. elegans* hsp70C protein (Fig. 2). Although the amino acid sequence in this region is not completely conserved, the hy-

Table 1. Summary of changes in the nucleotide sequence between the *hsp-3* homologs

DNA region	Nucleotide matches	Nucleotide mismatches	Total number of shared nucleotides	Percent identity[a]
Coding				
Exon 1	180	17	197	91.4
Exon 2	441	33	474	93.0
Exon 3 (partial)	614	48	662	92.7
Exon 1 + 2 + 3	1235	98	1333	92.6
Third codon position[b]	160	44	204	78.4
Noncoding				
5′ Flanking	270	109	379	71.2
Intron 1[c]	18	14	32	56.3
Intron 2[c]	141	40	181	77.9

[a] Not corrected for multiple mutation events
[b] Includes only the codons CTN, GTN, TCN, CCN, ACN, GCN, CGN, and GGN
[c] Intron calculations exclude the 5′ and 3′ conserved splice sites; because of uncertainties in the alignments, the percent identity values shown must be regarded as approximations

drophobicity is. Amino acid replacements appear to be randomly distributed throughout the first and second exons (5 out of 62 and 7 out of 158, respectively; Fig. 2). The partial sequence of the third exon from *hsp-3*(Cb) representing 221 amino acids does not show any changes in amino acid content when compared to the *hsp-3*(Ce) gene. However, nucleotide substitutions appear to be as prevalent in the third exon (7.3%) as in the first and second exons (8.6% and 7.0%, respectively). Functionally, this domain may be highly important and thereby less able to tolerate amino acid replacements than the domains encompassed by the first and second exons. In fact, the domain encoded by the third exon diverges quite highly from the analogous domains in other hsp70-like proteins but not the grp78-like proteins. It may be that this domain is highly specific for and unique to the grp78 subfamilies and that the analogous regions in the other hsp70 subfamilies would, by the same argument, encode domains highly specific and unique to each hsp70 subfamily.

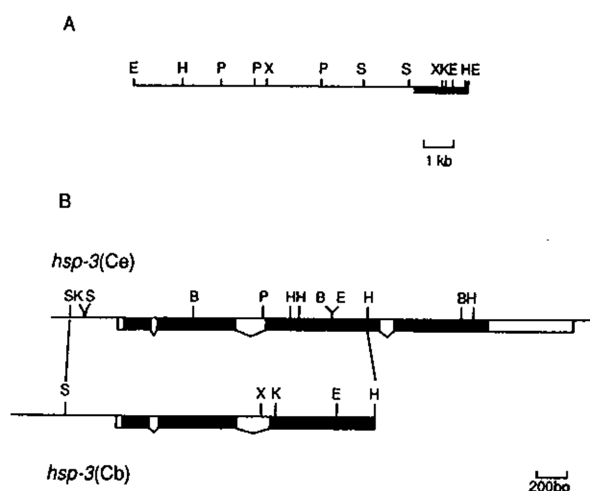

Fig. 1. **A** Restriction map of the CB10 phage containing the *hsp-3* gene from *C. briggsae*. **B** Restriction maps of the *hsp-3* genes from *C. elegans* [*hsp-3*(Ce)] and *C. briggsae* [*hsp-3*(Cb)]. The genes are aligned to show regions of homology. The coding regions are shaded and the transcribed, untranslated regions are unshaded. Introns are shown by breaks in the coding region. B, BamHI; E, EcoRI; H, HindIII; K, KpnI; P, PstI; S, SalI; X, XbaI.

Comparison of the hsp-3 5′ Flanking DNAs

The 5′ flanking DNAs of the *hsp-3* homologs were aligned for maximum identity from the translation start site using Microgenie (Beckman) and visually with ESEE (Cabot and Beckenbach 1989). Both methods gave comparable alignments and results. Overall, the percent identity of the 5′ untranslated region was determined to be approximately 71%. A summary of the nucleotide changes is presented in Table 1. Several conserved elements were observed and are outlined in Fig. 2. For this study, only those regions that exhibited 75% identity or more over their lengths were considered to be conserved elements. Block 5 contains an identity to the E1A enhancer and blocks 3 and 8 contain identities to the SV40 and E1A enhancers in the reverse orientation, respectively (Hearing and Shenk 1983; Weiher et al. 1983). Block 3A is a highly conserved repeat of block 3 [from −140 to −122 in *hsp-3*(Ce)] and also contains a copy of the SV40 enhancer in the reverse orientation. (The conservation of the viral enhancers raises a couple of intriguing possibilities. First, *Caenorhabditis* viruses that are related to SV40 and adenovirus may exist, or, second, the viral enhancer motifs may be functionally quite ancient and may have been recruited as enhancers for use by the mammalian viruses.) Block 5 also contains an identity to the HSE (CNNGAANNTTCNNG; Pelham

```
Cb  -665  AACTGTGTCTATAGACTTGCGTCCTCGTAATGTTTAAGTTCCTTCTCTCATTAAATGTCTCTAATTCATCCTAACTTCCCATTTTATGTATTCCAATTTT

Ce  -441                                                                            GGACACAGGGCCACACGGCCGTCGC
Cb  -565  AGGAGCGTTTATACGCATAGCGACCGCGACGCACCACTCTCCTGACGTGTATTAGGAGTGTCTTCCATCCTCGATCTCGG.CA...A..GTT..G.CTCA

                      9                           ****** **     8                                  SalI 7
Ce  -416  CGATTGGCCGAATGACTCTGCGTCTCTGCGCGCTGCACACGGTGAGCCTTCGCTGTGCTGACGTTGCCTTGTCCTATCGTCCTAGGCCACGTCGACGATT
Cb  -465  ACGA...T..GC..............T...T...TTTACGC.A........T...............C...CGG.TCTT..............ATG.
                                                    ****** *                                    SalI

                                                                                                  6
Ce  -316  CGG-CAGTTCGTTCCTTCG-CTCTCTCCCACTCGATGCGCTCGTCGAT--------------------------------CCGTCAGTTTGCTCTCCCTT
Cb  -365  ...G.......GA.G.GTTT....TC...-...T-...C...T..TC.CGCTCTCCATTTCTGTTACCTTCACAATCCCA.................
                                                         XX XXXX 3A
                                KpnI              SalI          5           *** ** **                4
Ce  -250  CACCACTCCC-AT--CGGTTGACGGTACCATTTCGGCCTACAGTCGACCTTGAGCATTCGGGCGGTCT--ATCGGGAGAGACGACCTACAAACAGAAGCA
Cb  -269  .TTTT.AT.GG..GA....-..-.C.TC....----.-....T.....................GT.........C.TT.ATG-----......GC
                                                                              *** **  *

                      XXXXXXX        3                      2A
Ce  -155  GTCCTA--------GGTTTTCCT---GCATTCCATTTCTCTCACCGACTGGCCTTGTTTCGGTCTCTTTCT---TTATC--------------------
Cb  -183  .A....CAAACCAA..AAA...GCAC........--.G....--.AG.....A.....GA.-..---......CCG.G...CCTTTTTGATTTAATATCCCA
                      XXXXX

                  2                                    1
Ce   -90  --------TCTTTCTTCTCAGCAATTCAACAAGTCGTTTCATATTTTAGGCCTAA-TAATAATTTTTA--TTTTTACAGGAAAATAAATCAAACACAAAG
Cb   -91  TTCTCTCA..........---TTT..T-.---.T.-.AA..T..........A.TAA--.....CTC.....C.....GG............C..

           M  K  T  L  F  L  L  G  M  I  A  I  T  A  V  S  I  Y  C  K  E  E  E  K  T  E  K  K  E  T  K  Y  E
Ce    -1  TATGAAGACCTTATTCTTATTGGGCATGATCGCCATCACCGCCGTCAGTATCTACTGCAAGGAAGAGGAAAAAACCGAGAAGAAGGAGACCAAGTATGAA
Cb    -1  .........................C.C......C..............G.A......G..............G.......................C.G.
                                   L        L              V        E                                        G

           T  I  I  G  I  D  L  G  T  T  Y  S  C  V  G  V  Y  K  N  G  R  V  E  I  I  A  N  D  Q
Ce   100  ACCATTATTGGTATCGATCTCGGAACCACCTACTCGTGTGTCGGAGTTTACAAGAACGGACGTGTTGAAATCATTGCCAACGACCAAGgtatgtgaacga
Cb   100  .....C..C.......................T..C.....T.....A.............................C..T...........a.ca...tt

                                                                       G  N  R  I  T  P  S  Y  V  A  F  S  G
Ce   200  aaaataaacgtaaatta-----------------taaccc-----------atcattttcagGAAACCGTATCACCCCATCCTACGTTGCTTTCTCTGG
Cb   200  tct..ttta.g-.....gcaggatcatcctctcac....atcattttgacga..t...........................G........C........

           D  Q  G  D  R  L  I  G  D  A  A  K  N  Q  L  T  I  N  P  E  N  T  I  F  D  A  K  R  L  I  G  R  D
Ce   271  AGATCAAGGAGATCGTCTGATCGGAGATGCTGCTAAGAATCAGCTCACCATCAACCCAGAAAACACAATCTTTGATGCCAAGCGTCTTATCGGAAGAGAT
Cb   299  ...G........C.....T...........C..C.....C.........................C.....C........A.....C............
           E

           Y  N  D  K  T  V  Q  A  D  I  K  H  W  P  F  K  V  I  D  K  S  N  K  P  S  V  E  V  K  V  G  S  D
Ce   371  TACAACGACAAGACTGTTCAAGCTGACATCAAGCACTGGCCATTCAAGGTTATTGACAAGAGCAACAAGCCATCCGTCGAAGTCAAGGTTGGATCCGACA
Cb   399  .....T.................C..T.......................T.C...........................................T....
                                                              F

           N  K  Q  F  T  P  E  E  V  S  A  M  V  L  V  K  M  K  E  I  A  E  S  Y  L  G  K  E  V  K  N  A  V  V
Ce   471  ACAAGCAATTCACCCCAGAAGAAGTTTCCGCTATGGTTCTCGTCAAGATGAAGGAGATCGCCGAGTCCTACCTTGGAAAGGAAGTCAAGAACGCCGTCGT
Cb   499  ........G...........................................................A..............C...........
                                                                                               H

           T  V  P  A  Y  F  N  D  A  Q  R  Q  A  T  K  D  A  G  T  I  A  G  L  N  V  V  R  I  I  N  E  P  T
Ce   571  CACTGTCCCAGCTTATTTCAACGACGCCCAACGTCAAGCTACCAAGGATGCCGGAACCATCGCTGGATTGAACGTTGTTCGTATCATCAACGAGCCAACC
Cb   599  ...C...................TT.....G.........ATTA...T..................T............C............T..........
                                 V                 L     Y                V

           A  A  A  I  A  Y  G  L  D  K  K  D
Ce   671  GCCGCCGCCATCGCCTACGGACTTGACAAGAAGGACGgtgagtt--tat-gagaaa-----gtgctctctaatatttgtctctggactacccttttgacc
Cb   699  ......................................ag..ga......acagaaa.ac.............a.ga........-...--...

Ce   763  a-------ttttgtgtaaacaatagatttttgggtcagtgactggtacaggttc-ctctctcg-tt-agg-aatgaggaatagggaatgtttgctcaggtcc
Cb   796  .atgaccg.......c-..........c...c.......................g...at...a..g...c.....cagtgct.--.......a..ac..t
                                                                                                          Xb
                                                                            PstI
Ce   852  gaagctgtaccaaatcacag----------------------attaa---gatatagaggt-t-gac--tgcagatttgaaacaaaataattcttcca
Cb   893  --..a.---......t-..tgtttcgaaaactagatttgctc.....tag.ga..t.....ac.a...aa..a.c.a.aatg-----------------
           al

                                   G  E  R  N  I  L  V  F  D  L  G  G  G  T  F  D  V  S  N  L  T  I  D  N  G  V
Ce   922  atcatgaatgtctttcatttacagGAGAACGCAACATCCTCGTCTTCGATCTTGGAGGTGGTACTTTCGATGTATCCATGCTCACCATTGACAACGGAGT
Cb   970  --------------tt...t.......................................C...........G............................
                                                                     KpnI
                                                                                        HindIII
           F  E  V  L  A  T  N  G  D  T  H  L  G  G  E  D  F  D  Q  R  V  M  E  Y  F  I  K  L  Y  K  K  K  S
Ce  1022  CTTCGAAGTTTTGGCCACCAACGGAGACACTCACTTGGGAGGAGAAGACTTTGACCAACGTGTCATGGAATACTTCATCAAGCTTTACAAGAAGAAGTCT
Cb  1055  ......G..C............................T..C..T..............................C..............C

                                   HindIII
           G  K  D  L  R  K  D  K  R  A  V  Q  K  L  R  R  E  V  E  K  A  K  R  A  L  S  T  Q  H  Q  T  K  V
Ce  1122  GGAAAGGATCTCCGCAAAGACAAGCGTGCCGTTCAAAAGCTTCGTCGTGAGGTCGAGAAGGCAAAGAGAGCTCTCTCCACTCAACATCAAACCAAGGTTG
Cb  1155  ...............G.........C..T.....G.........A..........C.....................C......C..........

           E  I  E  S  L  F  D  G  E  D  F  S  E  T  L  T  R  A  K  F  E  E  L  N  M  D  L  F  R  A  T  L  K  P
Ce  1222  AGATTGAATCTCTTTTCGACGGAGAAGACTTCTCTGAGACCCTTACTCGTGCCAAGTTCGAGGAGCTCAACATGGATCTCTTCCGTGCCACCCTTAAGCC
Cb  1255  ....C..G..C..C...........................A.....C..C.......................T........C......C..T.....C.....

                                                                  BamHI      EcoRI
           V  Q  K  V  L  E  D  S  D  L  K  K  D  D  V  H  E  I  V  L  V  G  G  S  T  R  I  P  K  V  Q  Q  L
Ce  1322  AGTCCAGAAGGTTCTTGAAGATTCTGATCTTAAGAAGGATGATGTTCACGAGATTGTTCTCGTCGGAGGATCCACTAGAATTCCAAAGGTCCAACAGCTC
Cb  1355  ........................C.....C...............C.................T............T..C...............T.....A...
                                                                                          EcoRI

           I  K  E  F  F  N  G  K  E  P  S  R  G  I  N  P  D  E  A  V  A  Y  G  A  A  V  Q  G  G  V  I  S  G
Ce  1422  ATCAAGGAGTTCTTCAACGGAAAGGAGCCATCCCGCGGAATCAACCCTGACGAGGCCGTCGCCTACGGAGCCGCCGTCCAAGGAGGAGTTATCTCTGGAG
Cb  1455  ..............................................T.........................T..........................

                                                                                        HindIII
           E  E  D  T  G  E  I  V  L  L  D  V  N  P  L  T  M  G  I  E  T  V  G  G  V  M  T  K  L
Ce  1522  AGGAAGACACTGGAGAGATTGTTCTTCTTGATGTCAATCCGCTTACCATGGGTATTGAGACTGTCGGAGGAGTTATGACCAAGCTT
Cb  1555  ..................C.........C..C.....C..A..............C...................C...........
                                                                                        HindIII
```

1982, 1985), although the *C. elegans hsp-3* gene is not heat inducible under the conditions tested (Snutch et al. 1988). None of these identities detected were perfect matches. An element in block 7 contains an identity detected in a comparison of the *hsp-3*(Ce) regulatory region with a portion of the rat grp78 regulatory region (Heschl and Baillie 1989; Resendez et al. 1988). This element, which is known to be important for the high-level expression of the rat grp78 gene, is also conserved in the human grp78 and the human and chicken grp94 genes (Chang et al. 1987; Resendez et al. 1988). Block 6 is an imperfect repeat of block 7, sharing identity with nucleotides −326 to −313 of block 7 in *hsp-3*(Ce). The other blocks contain structural features including inverted repeats (block 9) and pyrimidine-rich sequences (blocks 2 and 2A). Block 1 probably represents the transcribed, untranslated region. Dot matrix analyses did not detect any identities in the reverse complement of one strand that may have been missed by our alignment analyses. On the basis of the identities to known functional elements and structural features, we believe that we have detected most, if not all, of the conserved elements at the 5′ end of the *hsp-3* homologs that may be involved in regulation and/or structure.

Blumenthal and Zucker-Aprison (1987) have stated that Box 1 and Box 2 elements, two heptameric sequences potentially involved in transcriptional regulation, are conserved between the *C. elegans* and *C. briggsae* vitellogenin genes. Box 2 homologies appear to be required for estrogen stimulation of *Xenopus* vitellogenin gene transcription in transfected human mammary carcinomas (Klein-Hitpass et al. 1986). Potential stem-forming regions at the 5′ end of the vitellogenin genes are also conserved between *C. elegans* and *C. briggsae.* Even among multigene families such as the major sperm protein (MSP) genes from *C. elegans,* conserved putative 5′ regulatory elements are observed scattered amid highly diverged sequences (Klass et al. 1988).

Comparison of Intron Sequences between C. elegans and C. briggsae

The two introns represented in the *hsp-3*(Cb) sequence were found at corresponding positions when compared to the first two introns of the *hsp-3*(Ce) sequence (Figs. 1 and 2). The second introns of *hsp-3*(Cb) and *hsp-3*(Ce) are approximately the same length (243 vs 238 nucleotides, respectively), whereas the first introns are not (74 vs 46 nucleotides, respectively). A summary of nucleotide substitutions is presented in Table 1. The percent identity of intron 2 is much higher than that of intron 1 (Table 1), suggesting that there may be conserved elements in intron 2. Therefore, we compared the intron 2 sequences further to detect any possible conserved elements (Fig. 2). Two sets of conserved elements were detected in intron 2 including the intron boundaries (5′ AG/GTAAGT ... TTTTCAG/G 3′; Blumenthal and Thomas 1988) and a noticeable region of nucleotide identity extending over most of intron 2.

In the relatively short introns of *C. elegans,* there are apparently few conserved elements other than the 5′ and 3′ splice sites. The need for specific sequences involved in splicing, such as the splice branch point, have apparently disappeared in *C. elegans* (Blumenthal and Thomas 1988). Therefore, it came as a surprise that a conserved element was present in the second intron of the *hsp-3* homologs. If, as suggested above, conserved elements in *Caenorhabditis* are functionally important, it may be that this intronic element is important in the regulation of *hsp-3* either as an enhancer of transcription or by affecting the splicing or stability of the *hsp-3* mRNA. In fact, it has been suggested that an enhancer element in the third intron of the *unc-54* gene (encoding the major myosin heavy chain) plays a major role in directing tissue-specific expression (Fire and Harrison, personal communication).

An Estimate of Divergence for C. elegans and C. briggsae

An estimate of the time of divergence between *C. elegans* and *C. briggsae* can be made by determining the nucleotide divergence of the silent third position of fourfold degenerate codons. The nucleotide divergence calculated for these sites in the *hsp-3* coding region is 25.5% when corrected for multiple mutational events. Nucleotide substitution rates at the silent third codon position have been determined

←

Fig. 2. Nucleotide sequences of the *hsp-3*(Ce) and *hsp-3*(Cb) genes. The sequences are aligned for maximum identity. Only the sequence up to and including a shared HindIII site is shown. The *hsp-3*(Ce) gene sequence is presented elsewhere in its entirety (Heschl and Baillie 1989). Introns are shown in lowercase letters. Numbering is with respect to the start of translation. Dots (.) indicate a nucleotide match and gaps (-) in the sequence are shown to maintain maximum homology of the sequences. The amino acid sequences are shown above and below their respective genes with only the amino acid replacements shown for the *hsp-3*(Cb) gene. Sequences homologous to the E1A core enhancer sequence [(A/C)GGAAGTG(A/C); Hearing and Shenk 1983] are indicated with an *, and the SV40 core enhancer sequences [GTGG(A/T)(A/T)(A/T)G; Weiher et al. 1983] are indicated with an x. Blocks 1–9 represent regions conserved in the 5′ flanking DNA between *C. elegans* and *C. briggsae.* Inverted repeats in block 9 are underlined and the identity to the heat shock element (CNNGAANNTTCNNG; Pelham 1982, 1985) in block 5 is underlined.

for bacteria [0.8%/million years (Myr) (Ochman and Wilson 1987)] and the nuclear genes of mammals [0.9%/Myr (Li et al. 1985a,b)], insects [1.1%/Myr (Bodmer and Ashburner 1984; Blackman and Meselson 1986)], plants [1%/Myr (Chang and Meyerowitz 1986)], and sea urchins [1.1%/Myr (Busslinger et al. 1982)]. Because all of these organisms have similar percent nucleotide substitutions/Myr, it is not unreasonable to assume that *Caenorhabditis* will also fall within the range of 0.8–1.1% nucleotide substitutions/Myr. Therefore, we estimate that *C. elegans* and *C. briggsae* last shared a common ancestor no less than 23–32 Myr ago.

Acknowledgments. We thank S. Prasad, A. Fire, and the reviewers for their helpful comments on this manuscript and A.C. Wilson for his help with the analysis of our results. We also thank A. Fire and S. Harrison for communicating their results to us prior to publication. This work was supported by a Medical Research Council of Canada Studentship and Simon Fraser University Graduate Fellowships to M.F.P.H. and a Natural Sciences and Engineering Research Council of Canada grant to D.L.B. The *hsp-3*(Cb) and *hsp-3*(Ce) gene sequences can be accessed from Genbank under the accession numbers M26906 and M26604, respectively.

References

Blackman RK, Meselson M (1986) Interspecific nucleotide sequence comparisons used to identify regulatory and structural feature of the *Drosophila hsp82* gene. J Mol Biol 188:499–515

Blumenthal T, Thomas J (1988) *Cis* and *trans* mRNA splicing in *C. elegans*. Trends Genet 4:305–308

Blumenthal T, Zucker-Aprison E (1987) Evolution and regulation of vitellogenin genes. In: O'Connor JD (ed) Molecular biology of invertebrate development. Liss, New York, pp 3–19

Bodmer M, Ashburner M (1984) Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of *Drosophila*. Nature 309:425–430.

Busslinger M, Rusconi S, Birnsteil ML (1982) An unusual evolutionary behaviour of a sea urchin histone gene cluster. EMBO J 1:27–33

Cabot EL, Beckenbach AT (1989) Simultaneous editing of multiple nucleic acids and protein sequences with ESEE. Comput Applic Biosci 5:233–234

Chang C, Meyerowitz EM (1986) Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. Proc Natl Acad Sci USA 83:1408–1412

Chang SC, Wooden SK, Nakaki T, Kim YK, Lin AY, Kung L, Attenello J, Lee AS (1987) Rat gene encoding the 78 kDa glucose regulated protein GRP78: its regulatory sequences and the effect of protein glycosylation on its expression. Proc Natl Acad Sci USA 84:680–684

Coulson A, Sulston J, Brenner S, Karn J (1986) Towards a physical map of the genome of the nematode *Caenorhabditis elegans*. Proc Natl Acad Sci USA 83:7821–7825

Coulson A, Waterston R, Kiff J, Sulston J, Kohara Y (1988) Genome linking with yeast artificial chromosomes. Nature 335:184–186

Edgley ML, Riddle DL (1987) *Caenorhabditis elegans*. In: O'Brien SJ (ed) Genetic maps 1987, vol 4. Cold Spring Harbor Laboratory, Cold Spring Harbor NY, pp 351–365

Emmons SW, Klass MR, 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

Fire A (1986) Integrative transformation of *Caenorhabditis elegans*. EMBO J 5:2673–2680

Hattori M, Sakaki Y (1986) Dideoxy sequencing method using denatured templates. Anal Biochem 152:232–238

Hearing P, Shenk T (1983) The adenovirus type 5 E1A transcriptional control region contains a duplicated enhancer element. Cell 33:695–703

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

Henikoff S (1987) Unidirectional digestion with exonuclease III in DNA sequence analysis. Methods Enzymol 155:156–165

Henikoff S, Eghtedarzadeh MK (1987) Conserved arrangement of nested genes at the *Drosophila Gart* locus. Genetics 117: 711–725

Heschl MFP, Baillie DL (1989) Characterization of the hsp70 multigene family of *Caenorhabditis elegans*. DNA 8:233–243

Klass M, Ammons D, Ward S (1988) Conservation in the 5′ flanking sequences of transcribed members of the *Caenorhabditis elegans* major sperm protein gene family. J Mol Biol 199: 15–22

Klein-Hitpass L, Schorpp M, Wagner U, Ryffel GU (1986) An estrogen-responsive element derived from the 5′ flanking region of the *Xenopus* vitellogenin A2 gene functions in transfected human cells. Cell 46:1053–1061

Lee AS (1987) Coordinated regulation of a set of genes by glucose and calcium ionophores in mammalian cells. Trends Biochem Sci 12:20–23

Li W-H, Luo C-C, Wu C-I (1985a) Evolution of DNA sequences. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum, New York, pp 1–94

Li, W-H, Wu C-I, Luo C-C (1985b) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174

Lindquist S, Craig EA (1988) The heat-shock proteins. Annu Rev Genet 22:631–677

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor NY

Martin CH, Mayeda CA, Meyerowitz EM (1988) Evolution and expression of the *Sgs-3* glue gene of *Drosophila*. J Mol Biol 201:273–287

Mierendorf RC, Pfeffer D (1987) Direct sequencing of denatured plasmid DNA. Methods Enzymol 152:556–562

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

Ochman H, Wilson AC (1987) Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J Mol Evol 26:74–86

Pelham HRB (1982) A regulatory upstream promoter element in the *Drosophila* hsp70 heat-shock gene. Cell 30:517–528

Pelham HRB (1985) Activation of heat shock genes in eukaryotes. Trends Genet 1:31–35

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

Resendez E Jr, Wooden SK, Lee AS (1988) Identification of highly conserved regulatory domains and protein binding sites

in the promoters of the rat and human gene encoding the stress-inducible 78-kDa glucose-regulated protein. Mol Cell Biol 8:4579–4584

Snutch TP (1984) Molecular and genetic analysis of the heat shock response of *Caenorhabditis elegans*. PhD thesis, Simon Fraser University, Burnaby, British Columbia

Snutch TP, Heschl MFP, Baillie DL (1988) The *Caenorhabditis elegans* hsp70 gene family: a molecular genetic characterization. Gene 64:241–255

Sulston JE, Horvitz HR (1977) Post-embryonic cell lineages of the nematode *Caenorhabditis elegans*. Dev Biol 56:110–156

Sulston JE, Schierenberg E, White JG, Thomson JN (1983) The embryonic cell lineage of the nematode *Caenorhabditis elegans*. Dev Biol 100:64–119

Weiher H, Konig M, Gruss P (1983) Multiple point mutations affecting the simian virus 40 enhancer. Science 217:626–631

Received July 14, 1989/Revised November 30, 1989