MOLECULAR ANALYSIS WITHIN THE DPY-20 TO UNC-22 REGION OF CHROMOSOME IV IN *Caenorhabditis elegans*

by

Shiv Shankar Prasad

B.Sc.(Biology), University Of Guyana, 1981

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in the Department

of

Biological Sciences

© Shiv S. Prasad, 1988.

Simon Fraser University

June, 1988

All rights reserved. This work may not be reproduced in whole or in part, by photocopy or other means, without permission of the author.

Approval

Name: Shiv Shankar Prasad

Degree: Doctor of Philosophy

Title of Thesis: Molecular analysis within the *dpy-20* to *unc-22* region of chromosome IV in *Caenorhabditis elegans*

Examining Committee:

Chairman: Dr. J.M. Webster, Professor

Dr. D.L. Baillie, Professor, Senior Supervisor

Dr. B.M. Honda, Associate Professor

Dr. M. Smith, Professor

Dr. K.C. Eastwell, Assistant Professor, Public Examiner

Dr. A.H. Burr, Associate Professor, Public Examiner

Dr. J. Sulston, Professor, Laboratory of Molecular Bio., MRC, Cambridge, England, External Examiner

June 29, 1988

Date approved

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

Molecular analysis within the dpy-20 to unc-22 region of chromosomes IV in Caenorhabditis elegans

Author:

(signature)

Shiv Shankar Prasad

(name)

June 16, 1988

(date)

ABSTRACT

An excellent opportunity exists in *Caenorhabditis elegans* to study the organization of a complex genome. The alignment of the genetic and molecular maps over a large stretch of the genome is an essential part of this study. The objective of this thesis was the identification and characterization of coding regions in four cosmids containing DNA from the interval between *dpy20* and *unc-22* on linkage group *IV*. These cosmids were characterized with regard to the map position and the developmental patterns of expression of coding sequences. Since an extensive genetic map already exists for this region, this detailed description of the coding sequences in the *dpy-20* - *unc-22* region will make possible alignment of the molecular and genetic maps for this portion of the *C. elegans* genome.

In this thesis, I have used interspecies cross-hybridization in order to localize and identify potential coding elements (a method suggested by T. Snutch). I have investigated four cosmids containing approximately 150 kilobases of *C. elegans* genome adjacent to the well characterized muscle gene, *unc-22(IV)*. In total cosmids (kindly provided by A. Coulson and J. Sulston, MRC, Cambridge) in this region represent approximately 300 kilobases of DNA from the 0.2 map unit interval between *dpy-20* and *unc-22*. Fragments subcloned from the four cosmids were hybridized at moderate stringency to the genome of the related species, *Caenorhabditis briggsae*. In this way nine potential coding regions were identified. Seven of these nine fragments also hybridized to mRNA transcripts on

Northern blots. Interestingly five of the seven showed maximal hybridization to RNA from L2 stage animals, a pattern that resembles that of actin transcription. It is speculated that the functions of these five may in some way be related to one another, and perhaps also to that of unc-22, which is itself a muscle gene

Two of the fragments hybridized to clones in a cDNA library. These cDNA clones were sequenced and the corresponding genomic DNA were located for their sequence analyses. The corresponding DNA sequences in *C. briggsae* were also determined; the nature and extent of DNA sequence divergence between the two species was consistent with strong conservation of amino acid residues in the exons, and with an elapsed time of at least 35 million years since the last shared common ancestor.

DEDICATION

To my mom.

ACKNOWLEDGEMENTS

This work would not have been possible without the support, patience and encouragement of my senior supervisor, Dr. David Baillie. I also thank my supervisory committee, Dr. B.M. Honda and Dr. M.J. Smith, for their suggestions and guidance; Raja Rosenbluth for patiently describing genetic analyses; my colleagues Dr. Ann Rose, Karen Beckenbach, Denise Clark, Dr. Donald Nelson, Brian Kuchinka, Mark Heschl, Marcia Wakarchuk, John Boom, David Kowbel and Eric Cabot for their helpful discussions and encouragement; Denise Clark for constructing the deficiency strains; Brian Kuchinka for his help in screening of a charon 4 *C. briggsae* genomic library and sequencing; Raja Rosenbluth and Dr. Ann Rose for critically reading this thesis; Eric Cabot for help with his ESEE sequence editor program; Karen Beckenbach and Margaret Rogalski for their excellent technical assistance; John Boom for help with the figures; and finally to my father, Dr. Shree K. Prasad, for suggesting this laboratory for my study.

This work was in part supported by the predoctoral fellowship from the Muscular Dystrophy Association of Canada.

TABLE OF CONTENTS

Page#

Approval....................(ii)

Abstract....................(iii)

Dedication....................(v)

Acknowledgement....................(vi)

List of Tables....................(ix)

List of Figures....................(x)

A. Introduction....................1

B. Materials and Methods....................17

Strains and Culture Methods....................17

Construction of Strains with Deficiencies Balanced over nT1 lethal....................17

Preparation of Genomic DNA....................18

Restriction Enzyme Digestions....................18

Agarose Gel Electrophoresis....................19

Transfer of DNA to Nitrocellulose....................19

Isolation of DNA Fragments From Agarose Gels....................20

Labelling of DNA Probes....................20

Hybridization of Probes to DNA Filters and the Hybridization Criteria....................21

Subcloning of DNA from Cosmids....................22

Isolation of cDNA Clones....................22

Isolation of Nematode RNA....................23

Northern Blot Hybridizations....................24

Screening of a Charon 4 C. briggsae Genomic Library....................26

DNA Sequence Analysis....................27

Comparative Sequence Analysis....................28

C. Results..29

Section I..30

Mapping of *sP4* To Genetic Map and Positioning of Cosmids in the *unc-22* Region..30

Restriction Mapping of Cosmids..35

Detection of Transcribed Regions..37

Section II..40

Northern Blot Analysis..40

Screening of a *C. elegans* Mixed Stage Lambda *gt*10 cDNA Library..40

Northern Blot Analysis For The Developmental Stage Specific Expression..48

Screening of 24Hour Stage Specific cDNA Library.............54

Positioning of The C2 Gene on Genetic Map..................55

Section III..59

Sequence Analysis of cDNA Clones...............................59

Isolation and Characterization of The C2 and C5 Related Sequence From *C. briggsae*.........................62

Sequence Comparison of C2 Homologous Regions Between *C. elegans* and *C. briggsae*.......................64

D. Discussion..75

Correlation of The Physical and The Genetic Maps...........76

Identification Of Coding Regions In Cosmids................79

Isolation Of cDNA Clones Representing The *unc-22* Region.....81

Characterization Of The Identified Coding Regions..........83

The C2 Gene Lies In Zone 5 Of The Genetic Map...............84

Sequence Analyses..86

E. Summary and Conclusions..91

F. Proposals For Further Research.................................96

G. References..98

LIST OF TABLES

Page#

Table 1: Southern blot analysis of deficiency DNAs with the sP4 fragment....33

Table 2: Hybridization test of fragments derived from the restriction map in Figure 4....39

Table 3: Relative expression of *C. elegans* probes in total RNAs of different stages of development....53

Table 4: Southern blot analysis of deficiency DNAs with the C2 gene....57

Table 5: Percent divergence between *C. elegans* and *C. briggsae*....70

Table 6: Percent divergence in the replacement codons....71

Table 7: Intron boundaries in *C. elegans* and *C. briggsae*....72

Table 8: Codon usage comparison between the C2 genes of *C. elegans* and *C. briggsae* and the *lin-12* and *vit-5* genes of *C. elegans*....74

LIST Of FIGURES:

Page#

Figure 1: A detailed genetic map of the unc-22 region on the linkage group IV....10

Figure 2: Hybridization of ^{32}P-labelled inserts from sP4 and pS18 fragments to deficiency strain DNAs. 32

Figure 3: The alignment of the cosmid contigs to the genetic map in the dpy-20 - unc-22 interval on linkage group IV....34

Figure 4: Restriction maps of cosmids....36

Figure 5: Hybridization of ^{32}P-labelled subcloned fragments to C. elegans and C. briggsae genomic DNAs....38

Figure 6: Northern blot hybridization with total RNA from mixed developmental stages of C. elegans populations....41

Figure 7: Ethidium bromide stained pattern of the isolated cDNA clones on EcoRI digestions....43

Figure 8: Hybridization of the subcloned cDNAs C1 and C2 to the cosmid C11F2 DNA....44

Figure 9: Hybridization of the cDNA C4 and C5 to the 9.2kb. PstI subclone "g"....46

Figure 10: Hybridization of C2, C4 and C5 cDNAs to C. elegans and C. briggsae genomic DNAs....47

Figure 11: Hybridization of C5 cDNA to the cosmid DNAs....49

Figure 12: Developmental expression of identified coding regions on cosmids....51

Figure 13: Relative developmental stage specific expression of each C. briggsae homologous region....53

Figure 14: The genetic map of the region around unc-22(IV) with the position of the C2 gene....58

Figure 15: Sequence of the C2 cDNA from C. elegans....60

Figure 16: Sequence of the C5 cDNA from C. elegans....61

Figure 17: A detailed restriction map of the fragments FH6 and "g" with the directions of transcription for the C2, C4 and C5 genes....63

Figure 18: Isolation of C2 and C5 related clones from a C. briggsae genomic library....................65

Figure 19: The EcoRI restriction maps of three C. briggsae overlapping phage (Cb8, Cb12 and Cb15) isolated using C2 as a probe....................66

Figure 20: Sequence alignment of C2 homologous regions from C. elegans and C. briggsae genome....................67

Figure 23: The region between the unc-43 and dpy-20 genes on the linkage group IV....................78

A. Introduction:

One of the fundamental problems in biology is the question how a single cell, the fertilized egg, develops into a complex multicellular organism. A complex set of biochemical and morphological changes occurs during the development of a fertilized egg to the adult stage. Since the work of T. H. Morgan (1934), it has been clear that the information for development resides in the genome of the organism. Complex multicellular eucaryotes in comparison to procaryotes, are composed of many different cell types. Differences among these cell types arise as a result of their individual abilities to synthesize and accumulate different sets of proteins even though all cells contain the same DNA sequences (Gurdon 1962; Garrels 1979). Thus, the development of a multicellular eucaryote from a single cell is thought to be brought about by regulating (either positively or negatively) the expression of different genes or sets of genes in particular cells at specific times during development. The molecular mechanisms of the regulation of gene expression in eukaryotes is still poorly understood.

Genetic and biochemical analyses of procaryotes have demonstrated that gene expression can be regulated at any point in transcription from DNA to RNA, and/or in translation from RNA to protein. Furthermore, a gene or an operon in procaryotes is regulated by a protein or proteins that recognize 5′ region of DNA sequences, thereby controlling the outcome of transcription (gene activation or repression) (Miller and Reznikoff, 1978). A similar mechanism is believed

to occur in the case of higher eukaryotes. The puffing patterns of *Diptera* suggested that different genes were turned on or off during development (Beerman, 1956). Several different studies in *Drosophila* and other eukaryotes have demonstrated differential gene expression as a major control phenomenon for the maintenance of developmental processes, e.g. actins (Karlik et al., 1984), globins (reviewed by Proudfoot et al., 1980), mouse liver and salivary gland specific alpha amylases (Young et al., 1981), etc. Regulation of transcription of these genes occurs both temporally and spatially. In order to examine the mechanisms of tissue specific and stage specific expression of genes, the molecular genetical approach is ideal. This approach uses a combination of classical genetic analysis together with molecular cloning and sequencing (Murray and Murray, 1974; Sanger et al., 1977; Maxam and Gilbert, 1977; Blattner et al., 1977; Maniatis et al., 1978).

In order to achieve an understanding of the mechanisms by which the expression of eukaryotic genes are governed, it is important to first know how genes are organized in the genome. *Drosophila melanogaster* has been intensively studied by both genetic and molecular approaches (Spradling and Rubin, 1981; reviewed by Scott, 1987). Molecular tools used have included chromosomal "walking" by the isolation of overlapping cloned fragments. This method has provided interesting molecular information regarding many loci in a given region of *Drosophila melanogaster* (Bender, 1983). The P-element method, which

involves transforming modified genes into the germ line has been another important tool in Drosophila since it permits a gene to be tested for its function during normal development as a single copy insert into a chromosome (Spradling and Rubin, 1982).

Approximately 70% of Drosophila DNA is single copy DNA while 30% is composed of repetitive sequences. The latter class includes the genes for rRNA, 5SRNA, tRNA, histones and multigene protein families. The single copy DNA includes most of the genes coding for proteins and probably their regulatory sequences. So far the best estimate of the total number of genes in D. melanogaster has been made on the basis of genetic analyses of defined chromosomal regions. Even when non-translated or non-transcribed sequences are included, Drosophila has much more single copy DNA than necessary to code for the estimated number of genes (5000-6000) (Young and Judd, 1978; Hochman, 1978; Woodruff and Ashburner, 1979; Wright et al., 1981; and Zhimulev et al., 1981; Lefevre, 1986). A higher estimate of about 10,000 - 15,000 different mRNA transcripts has been derived by measurements of mRNA complexity (Izquierdo et al., 1979; Hough-Evans et al., 1979). Both estimates may appear quite low for Drosophila, but the studies on complex loci have suggested some explanations. There is no necessity for Drosophila to have a very high number of genes if each gene has multiple products and complex regulation. Several genes like actin, myosin heavy chain and tropomyosin may be genetically more complex than presently classified. Molecular

analyses in this group of genes have determined multiple RNA and protein products with multiple _cis_ acting regulatory elements (Karlik et al., 1984).

With a detailed and well ordered genetic map of a given region and the identification of corresponding genomic DNA, one can address the question of how different genes, which are physically linked in the genome of an organism, are regulated during eucaryotic development. It is of interest to know whether: 1) any genes which are coregulated are found adjacent to one another; 2) genes with similar functions are located next to each other in the genome; and 3) whether all genes are separated by large spacers or some are tightly clustered. In _Drosophila_ _melanogaster_ large regions of the genome have been analysed through this approach. Gene clusters in _Drosophila_ have been classified into four categories: a) reiterated genes e.g. rRNA genes (Long et al., 1980) and the histone gene clusters (Lifton et al., 1977); b) gene complexes with _cis_ acting transcription units e.g. Bithorax complex (Lewis, 1978); c) functionally but not structurally related clusters e.g. the 88F region where the indirect flight muscle specific actin gene (_act88F_) is located within 140 kilobases of two transcription units which encode tropomyosin isoforms and tropomyosin-like proteins (Karlik et al., 1984); and d) functionally and structurally related clusters e.g. two chorion gene clusters at 7F1-2 and 66D11-15 (Spradling et al., 1980 and 1981) and three salivary gland glue protein genes within 5 kb of DNA at 68C (Meyerowitz et al., 1982; Crowley et al., 1983). Additionally,

the DOPA decarboxylase (Ddc) gene cluster (reviewed by T.R.F. Wright, 1986) in Drosophila is not comparable to any of the gene clusters mentioned above. Most of the genes in this cluster are functionally and structurally related but unlike other mentioned genes these genes have diverged far enough from each other to acquire separate enzymatic functions in the same pathway. Ddc is one of the largest clusters with at least 18 genes in approximately 160 kb of cloned and analysed DNA. The presence of several deficiencies has facilitated a coordinated genetic and molecular analysis. Through this analysis it was demonstrated that 16 of 18 genes have separate functions, although sequence homologies are present among those genes. Mutations in any of these genes are lethal. Other examples of elegant combined molecular and genetic approaches include the rosy-ace region (Clark et al., 1986), the Shaker genes (Papajian et al., 1987) and the Gooseberry - Zipper region (Cote et al.,1987).

The homeotic genes in Drosophila melanogaster have been intensively examined because they are among the best candidates for master regulatory genes controlling development. Several mutations in the homeotic genes that transform one part of the fly into another have been reported. The spatial and temporal patterns of expression by homeotic genes demonstrate their complexity in regulation. This complexity is not as much due to multiple gene products as it is due to the intricacy of the gene structure and to the effect of DNA insertions (Lewis, 1978; Scott, 1986; reviewed by Scott, 1987).

Although D. melanogaster is an organism of choice for studying genome organization and gene expression, it is important that this type of study be done in several complex eucaryotes. Caenorhabditis elegans is an excellent organism for this kind of study since it can be easily analyzed at both the genetic and the molecular levels. Its short generation time (3.5 days at 20° C) (Brenner, 1974; Riddle, 1978), and small number of cells (about 908 somatic cells) (Sulston and Horvitz, 1977) make it an ideal model system for this kind of investigation. Furthermore, the unique ability to freeze mutant strains of C. elegans provides the potential to isolate and maintain mutant strains representing all the genes in the genome. The complete cell lineage of this nematode from zygote to adult has been determined (Sulston and Horvitz, 1977). The haploid genome of this simple hermaphroditic organism contains approximately 8×10^7 base pairs of DNA, 17% and 83% of which are repetitive and single copy respectively (Sulston and Brenner, 1974). This organism is particularly suited for the study of genes that are involved in muscle cell development and maintenance. Several genes altering muscle structure and organization have been identified and studied (Waterston, 1980; Waterston et al., 1977; MacLeod et al., 1977; Moerman, 1982).

Investigation of the genome of Caenorhabditis elegans at the genetic level has resulted in the identification of over 500 genes (Brenner,1974; Swanson, 1984). The recombination map of each of the five autosomes suggests that most genes are in clusters (pointed out by Brenner, 1974). This striking feature

might be due to uneven recombination frequencies along the autosomes or alternatively be due to a real nonrandom gene distribution on the physical map. The calculated average amount of DNA/map unit on the basis of genome size (8 X 10^7 bp.) and the total number of map units (approximately 300; Swanson, 1984) is 267 kb/map unit. If there is uneven recombination frequency along the autosomes, then for clusters the amount of DNA/map unit should be greater than average. A correlation of the genetic and physical maps around lin-12(III) region has supported the proposal that the apparent clustering of genes observed in C. elegans autosomes is due to the relative decrease in recombination frequency within clusters rather than resulting from nonrandom gene distributions (Greenwald et al., 1987). Several other small regions in the genome of C. elegans are being analysed and a similar comparision of their physical and genetic maps should be feasible.

In C. elegans, combined molecular and genetic approaches are being utilized for the study of genome organization and gene regulation. Available for these are restriction fragment length differences (RFLD's), chromosomal rearrangements (e.g. deficiencies) and transposon tagged genes. The analyses are being carried out in several laboratories. Some elegant examples of such analyses are associated with the following genes: ama-1(IV) (Rogalski et al., 1988; Riddle, personal communication); MSPs (Ward et al., 1988); dpy-5(I) (Howell et al., 1987; Babity, personal communication); lin-12(III)

(Greenwald et al., 1987); actin (Files et al., 1983); and unc-15(I) (Rose and Baillie, et al., 1980). The subject of this thesis is another small region, the unc-22(IV) region.

The region around the twitcher gene unc-22(IV) has been under extensive genetic analysis (Clark, Rogalski, Donati and Baillie, 1988; Rogalski and Baillie, 1985). The unc-22 gene which is central to this region codes for a component of nematode muscle structure (Moerman et al., 1988; Moerman et al., 1986; Rogalski et al., 1982; Waterston et al., 1980; Moerman and Baillie, 1979; Waterston et al., 1977). Mutations in this gene cause a characteristic twitching along the body wall musculature. Substantial evidence suggests that this twitching is a result of a disorder in the muscle cell itself rather than due to an abnormality in the nervous system (Lewis et al., 1980; Moerman, 1980). This region that is defined by the deficiency sDf2, is approximately 1.6 map units in length and extends in both directions from the unc-22 gene. The region around the unc-22(IV) gene has been approximately 65% saturated for mutations in essential genes (D.V. Clark, personal communication). So far, approximately 40 essential genes have been identified and a Poisson analysis of the data indicates that this small region may contain at least 55-65 essential genes. Furthermore, most of these genes lie in the approximately 1.0 map unit interval to the left of unc-22 and are more densely organized here than to the right. Their density decreases greatly at a greater recombination distance from unc-22. The mutations around unc-22 were mapped by the use of several deficiencies (Clark et al., 1988; see Figure 1).

Figure 1:

A detailed genetic map of the *unc-22* region on the linkage group *IV* (Courtesy of Denise Clark; derived from Clark et al., 1988). This region is distributed into zones by means of several mentioned deficiencies.

unc-43 (s1165, s1185) mec-3 let-61 let-69 let-72 let-63 let-96 let-307 let-308 let-312 lin-3 let-71 let-64 let-74 (s678, s753) let-59 let-70 let-73 let-98 let-100 let-311 let-65 dpy-20 let-60 let-92 let-56 unc-22 let-52 let-93 let-66 let-67 let-309 let-68 let-99 let-97 unc-31

sDf2

sDf7 sDf21

sDf8 sDf22

sDf9 (+sDf65)

sDf10

nDf27

mDf7

sDf60

sDf61

sDf62

sDf63

sDf64

sDf67

Zone:	1A	1B	1C	1D	2	3	4	5	6	7	8	9	10
	7	1	4	7	1	1	1	1	1	1	2	2	2

0.5 map units

Mutations in essential genes closely linked to unc-22 show an interesting feature, the blocking in early larval stages of development (Clark et al., 1988; Rogalski et al., 1985; Rogalski et al., 1982). Analyses of different alleles of the same gene appear to show that some alleles are hypomorphs in comparison to others that are apparently null alleles (Rogalski et al., 1985). A molecular analysis of the expression of these genes is crucial in order to establish the correlation with their mutant phenotypes.

In order to clone DNA from this region Baillie et al. (1985) used a method which involved the construction of a special genetic strain that contained the region around unc-22 from a Bergerac (BO) strain of *C. elegans* while the rest of the genome was from the Bristol (N2) strain of *C. elegans*. Since the Bergerac strain is rich in Tc1 (transposable element) copy number (approximately 300) while Bristol strain has only a few Tc1 copies (approximately 30), the majority of Tc1 bearing genomic fragments in the constructed strain were derived from the unc-22 region. This method has facilitated the isolation of several small fragments of DNA (flanking the Tc1 sites) in the region around the unc-22 gene, between unc-43 on the left and unc-31 on the right. The fragments provided probes for RFLDs between Bergerac and Bristol wild type strains. In order to accurately position the isolated fragments with respect to the unc-22 gene, recombinants from BO/N2 heterozygotes were obtained in the intervals unc-22 to unc-31 and unc-43 to unc-22. In this manner six cloned probes for N2-BO RFLDs

were obtained. The polymorphic sites (sP3 to sP8) were mapped with respect to unc-22 with the use of Southern blots of DNA from the recombinants (Baillie et al., 1985).

In order to construct a physical map of the genome of C. elegans, Coulson et al. (1986) have constructed cosmid banks of the C. elegans genome and have developed a fingerprinting method for matching clones to one another so that overlaps can be conveniently recognized. Each set of overlapping cosmids (contig) provides a large contiguous length of DNA in a particular region of the C. elegans genome. Cosmids covering several hundred kilobases of DNA around the unc-22 gene have been isolated. The genetic map position was determined using the probes isolated by Baillie et al. (1985). The unc-22 gene, which was cloned by Tc1 tagging, has been used to isolate cosmids adjacent to unc-22 (Moerman et al., 1986). For my analysis, two sets of cosmid contigs were kindly provided by Coulson et al. Described in the first section of this thesis is a strategy for identifying their coding sequences.

In order to establish correspondence between genetically and molecularly identified genes, the identification and positioning of transcribed regions within the cloned DNA is necessary. For large regions this task can be a very difficult one. The screening of individual fragments for their ability to hybridize to RNA (Northern) blots is not sensitive enough to detect low abundance mRNAs. The screening of cDNA libraries is both time consuming and limited by the representation of any particular cDNA in the library.

The central objective of this thesis is to locate, as far as possible, the genes in the dpy-20 - unc-22 region of the C. elegans genome. In order to achieve this goal, I have utilized a method of interspecies hybridization. Previous preliminary work by T. Snutch (1984) has shown that in the HSP70 (heat shock 70 kdal) gene region of C. elegans and Caenorhabditis briggsae, sequence homologies were limited to coding regions and regions adjacent to them. The noncoding regions did not demonstrate any hybridization signals with C. briggsae DNA, even at lower stringency. Therefore, it is reasonable to assume that the conserved regions between C. elegans and C. briggsae are likely to include gene sequences that are necessary for nematode growth, development and reproduction, whereas diverged regions are mainly noncoding. Using this strategy I have identified nine coding elements in the dpy-20 - unc-22 cosmid contigs.

The second objective of the thesis was to characterize the identified coding regions by first isolating them and then by examining their expression patterns during the various development stages on Northern blots. This was done in order to examine if there was any correlation between the times of expression of these genes and the times of developmental blockage by lethal mutations in genetically identified genes of the same region. These results are described in the second section of the thesis.

The third section of this thesis involved sequencing of two cDNA clones to infer the amino acid sequences that would

constitute the primary structure of the encoded protein. The C. elegans genomic sequences with homology to these cDNAs were determined to further confirm the isolation of correspondingly transcribed cDNA clones.

Furthermore, in order to initiate a subsequent analysis to test the prediction of gene order maintenance between C. elegans and C. briggsae, corresponding C. briggsae genomic clones were isolated. The possibility of a functional significance for certain genes to be adjacent in the genome of organisms is a question of particular interest. The Ddc locus described earlier suggests the validity of the above speculation (Wright, 1986). Similarily the Bithorax Complex loci have been shown to regulate development by cis-acting transcription units (Bender et al., 1983). If these observations are consistent for most developmentally regulated genes, it is reasonable to expect that such functional organization in the genome will be maintained by selection in evolution. Organization of the nested genes in the Gart locus has been shown to be evolutionarily stable between D. melanogaster and D. pseudoobscura which are thought to have shared a common ancestor about 45 million years ago (Henikoff and Eghtedarzadeh, 1987).

A C. briggsae region, with homology to one of the cDNA clones, was sequenced in order to determine the extent of sequence divergence between the two species. Earlier Emmons et al. (1979), by hybridizations of the randomly cloned DNA fragments from C. elegans genome to C. briggsae genomic DNA,

had roughly estimated the sequence divergence between these two species to be at least 20%. Based on this figure the corresponding divergence time of these two species from a common ancestor was estimated to be in tens of million years. In many organisms evidence for the accumulation of base substitution at steady rates was derived from the comparisons of proteins and nucleic acids from species with abundant and well studied fossil records. These complementary lines of evidence have been used to form a molecular clock for estimating the biological diversity on a temporal framework (Wilson et al., 1987). However fossil records are difficult to obtain in the case of C. elegans and other soft bodied invertebrates. In order to estimate the rates of molecular evolution without ample fossil records the method relies upon the comparison of sequence divergence in a gene or sets of genes from two species. The extent of divergence between one gene from C. elegans and C. briggsae could, for example, be compared to estimate the divergence between the two species. Furthermore this comparison can suggest a possible functional constraint by selection on certain regions of the encoded gene product. In Drosophila, similar studies have been recently reported in Adh and Gart loci of D. melanogaster and D. pseudoobscura (Schaeffer et al., 1987; Henikoff and Eghtedarzadeh, 1987).

The differences in DNA sequences between a pair of homologous genes from C. elegans and C. briggsae can be divided into those that alter amino acids (replacement sites) or those

that do not alter amino acids (synonymous, introns and flanking sites). A comparison of rates of replacement changes with silent site changes often distinguishes between natural selection and random genetic drift. By such a comparison between one of the genes in *C. elegans* and *C. briggsae*, an estimate on the sequence divergence and evolutionary divergence time is also provided.

B. Materials and Methods

Strains and Culture Methods:

The *Caenorhabditis elegans* strain N2 variety Bristol was from S. Brenner, Laboratory of Molecular Biology, Hills Road, Cambridge. They were maintained on nematode growth media (NGM) streaked with *E. coli* strain OP50 as described by Brenner (1974). Large numbers of worms for biochemical purposes were grown on NGM plates supplemented with four times the amount of bactopeptone and cholesterol used in NGM (Rose et al., 1982). Strains other than N2 were either constructed at SFU or obtained from the *Caenorhabditis* Genetics Stock Centre at the University of Missouri. The mutation, *let- (m435)* on the translocation, *nT1(IV; V)* and the deficiency *mDf7(IV)* were kindly provided by T. Rogalski (Columbia, Missouri).

Construction of Strains with Deficiencies Balanced over nT1 lethal:

In order to obtain genomic DNAs containing stable deficiencies it is essential that the deficiency chromosomes be maintained in balanced heterozygotes. The balancer used was the translocation *nT1(IV; V)* (Ferguson and Horvitz, 1985). In order to have a population of worms in which the number of chromosomes with and without the deficiency was equal, the mutation *let- (m435)* on *nT1(IV; V)* was used. The strains with the genotypes *Dfx; +/let- (m435) nT1 (IV; V)* were kindly

constructed by D.V. Clark.

Preparation of Genomic DNA:

Genomic DNAs of wild type *C. elegans* and *C. briggsae* for the use in blot hybridizations were prepared as described by Rose et al. (1982). Genomic DNAs of deficiency strains were isolated by a miniprep formamide procedure as decribed by Reymond (1987).

Restriction Enzyme Digestions:

Genomic and plasmid DNA samples were digested with a variety of restriction enzymes obtained from either Bethesda Research Laboratory or Pharmacia. Depending upon the enzyme used the digests were carried out in either high, medium or low salt buffers as described in Davis et al. (1980). Specifically, the genomic DNAs were digested in at least twice the recommended amount of enzymes and all digests were allowed to proceed for twice the recommended time. For digests requiring more than one enzyme, the first digests were done in low salt-requiring enzymes and subsequently salt concentrations were increased as required for other enzymes. All digests were stopped by heating the samples at 68°C for 10 minutes. Genomic DNAs were normally digested in large volumes, precipitated with ethanol and then redissolved in lower volumes for convenience in loading of the agarose gels.

Agarose Gel Electrophoresis:

All genomic DNA samples were completely digested with EcoRI restriction endonuclease and 4 ug of DNA was loaded onto 0.7% neutral agarose gels prepared in 1X TBE buffer [10X TBE is 89 mM Tris, 89 mM borate, 2.5 mM EDTA (pH 8.3)]. Samples other than genomic DNA (plasmids or cosmids) were electrophoresed on 0.5% to 1.2% agarose gels, depending upon the restriction fragments under study. The molten gel contained 1 ug/ml of ethidium bromide. Electrophoresis was performed in 1X TBE buffer overnight at 25 volts. Marker DNA was lambda DNA (strain cI857 Sam 7) digested with EcoRI/HindIII or BamHI. The separated DNA fragments were visualized and photographed with aid of a 300 nm wavelength transilluminator.

Transfer of DNA to Nitrocellulose:

Following electrophoresis, gel was soaked in four times the gel volume of 0.25 M HCl for 30 minutes. The gel was then rinsed with double distilled water (ddH_2O) and soaked in four times the gel volume of 0.5 M NaOH, 1.5 M NaCl for another 30 minutes. Finally, the gel was soaked in similar volume of 1 M ammonium acetate, 0.02 M NaOH for 60 minutes. DNA fragments were transferred from the gel to nitrocellulose by the bidirectional transfer method of Smith and Summers (1978).

Isolation of DNA Fragments From Agarose Gels:

The DNA fragments were separated on 0.7% Low Melting Point Agarose (BRL) gels and the desired fragments were cut out, the agarose melted at 65°C in approximately 5 vol. of 20 mM Tris, 1 mM EDTA, pH 8.0 for 10 minutes. Subsequently the DNA fragment was extracted with phenol following the method described in Maniatis et al. (1982).

Labelling of DNA Probes:

Either the gel purified fragments or subclones from cosmids were nick translated to a specific activity of approximately 2×10^7- 2×10^8 cpm/ug using ^{32}P-dCTP or ^{32}P-dATP as described by Rigby et al. (1977). The reaction was allowed to proceed for 2 to 4 hours at 12°C. Prior to hybridization the probes were digested with EcoRI in order to achieve the best hybridizations of insert DNAs. Labelling of DNA probes for genomic blot hybridizations to deficiency DNAs was done by the oligolabelling technique as described by Feinberg and Vogelstein (1983). Probes with a specific activity of 1×10^8- 1×10^9 cpm/ug were synthesized by this procedure. Each probe was millipore filtered with non-sterile filters and then denatured by immersion in a boiling water bath for ten minutes. When the probes were used immediately after boiling, cooling in an ice waterbath was found to be nonessential.

Hybridization of Probes to DNA Filters:

DNA filters were prehybridized in 5X SSPE, 0.3% SDS and 5X Denhardts (1X SSPE is 0.18 M NaCl, 10 mM sodium phosphate, 1 mM EDTA, pH 7.4; 1X Denhardts is 0.02% BSA, 0.02% Ficoll and 0.02% polyvinyl pyrrolidone) at the same temperatures as used for hybridization. Hybridization was carried out in a fresh aliquot of the above solution and included denatured ^{32}P-labelled DNA probes. The washing of filters was done in either 2X SSPE, 0.2% SDS at the same temperature as the hybridization or at higher stringency depending upon the filters.

Filters containing _C. briggsae_ genomic DNA were hybridized at 65°C and washed at 62°C in 1X SSPE, 0.2% SDS. The filters containing deficiency genomic DNAs were hybridized at 68°C and washed at 65°C in 1X SSPE, 0.2% SDS. Hybridization and wash condition of all other filters varied depending upon the required stringency as indicated in Results. The three different stringency conditions are given below:

a) _Low stringency_: Hybridization in 5X SSPE, 0.3% SDS and 5X Denhardts solution at 62°C and the filters were washed in 2X SSPE, 0.2% SDS at the same temperature.

b) _Moderate Stringency_: Hybridization was in the same buffer as above at 65°C and wash was done in the same wash buffer as above at 65°C.

c) _High Stringency_: Hybridization was in the same buffer at

$68^{o}C$ and wash was done in 0.2X SSPE at $68^{o}C$.

Dried filters were exposed to Kodak NS-2T film at $-80^{o}C$ for the required duration (1-3 days). The relative absorbance for the measurements of band intensities in the case of deficiency DNAs were obtained by scanning each band (experimental and control) on the same lane and a ratio of each deficiency band intensity was compared to that of bands resulting from the hybridization of a probe which is known to be present in two doses (i.e. situated somewhere else in the genome).

Subcloning of Cosmids DNA:

Cosmid DNAs were prepared by a rapid alkaline lysis method as described by Maniatis et al. (1982). One ug of each cosmid was completely digested with *Hind*III and *Pst*I separately. These digests were correspondingly ligated with 0.2 ug of pUC19 digested with similar restriction enzymes. Transformations were done into JM83 (ara, lac^{-}, pro, strA, thi, 080d lacZ MI5) cells. Positives were selected on ampicillin (Amp), 5-bromo-4-chloro-indolyl-B-D-galactosidase (X-gal) and isopropyl-thiogalactoside (IPTG) plates (Ruther, 1980). Mini plasmid DNA preparations were done as described by Maniatis et al. (1982) using the alkaline lysis method.

Isolation of cDNA Clones:

A library of *C. elegans* cDNA cloned in lambda *gt*10 (a gift

from B. Meyer) was screened with ^{32}P-labelled subclones of cosmids that showed homology to C. briggsae genomic DNA. Approximately 15,000 phage were screened three times as described by Benton and Davis (1977). In one case, a 24 hr stage specific cDNA library also constructed in lambda gt10 (kindly provided by J. Ahringer) was screened similarly using a 3.3 kb PstI fragment (designated "u") as a probe. Prehybridization and hybridization conditions were the same as described for Southern blots. Dried filters were exposed to Kodak NS-2T film at -80°C for the required duration (1-3 days). Phage were isolated on a 0.75 g/ml CsCl equilibrium gradient. Phage DNA was isolated according to the procedure described by Davis et al. (1980).

Isolation of Nematode RNA:

A population of worms at mixed stages of development was collected from plates with 0.05 M NaCl and pelleted lightly. The pellet was washed twice with 0.05 M NaCl to remove the bacterial suspension. Worms were resuspended into 6 M guanidium hydrochloride, 0.02 M sodium acetate and 0.1 M B-mercaptoethanol (pH 5.0) and passed twice through a French Press at 12,000 psi (pounds per square inch; 1 psi= 6.894757 kPa) on to ice. Worm carcasses were spun out and half a volume of 95% ethanol was added to the supernatant and RNA was precipitated overnight at -20°C. Subsequently, total nematode RNA was purified by the procedure as described by

Chirgwin et al. (1979). The final RNA pellet was washed twice with 95% ethanol, lightly dried under a stream of air and dissolved in sterile diethyl pyrocarbonate treated distilled water.

Total RNA of four developmental stages were isolated by T. Snutch by a similar method as above.

Northern Blot Hybridizations:

Total RNA of mixed stages or stage specific total RNA (10 ug) was denatured in 2.2 M formaldehyde, 50% formamide, 1 mM EDTA, 5 mM sodium acetate, 20 mM morpholinopropane sulphonic acid (pH 7.0) (MOPS) for 15 minutes at 55°C and electrophoresed through a 1.1% agarose gel containing 20 mM MOPS, 1 mM EDTA, 5 mM sodium acetate, 2.2 M formaldehyde (pH 7.0). Electrophoresis was at 80 volts for 4 to 4.5 hours on 20 cm gels (Maniatis et al., 1982) and 50 V for 3 hours on 10 cm minigels. The RNA samples for marker lanes contained 1 ug of ethidium bromide prior to loading on to gels. Thus, staining of lanes after electrophoresis was not necessary and ethidium bromide in the marker lanes did not interfere with subsequent transfer of RNA to membranes.

Electrophoresed RNAs were transferred to filters by two procedures. In the first procedure, the RNAs were transferred to nitrocellulose filters in 20X SSPE (1X SSPE is 0.18 M NaCl, 10 mM sodium phosphate, 1 mM EDTA, pH 7.4) (Thomas et al., 1979). RNAs were allowed to transfer overnight, the filters

were rinsed in 5X SSPE (RNA side up) for 5 minutes and the filters were baked at 80°C under vacuum for 2 hours.

The second procedure utilized the transfer of RNA to a nylon based membrane (Gene Screen, New England Nuclear). In this method, the gel was rinsed once in distilled water for 15 minutes and then rinsed twice (15 minutes each) in 0.025 M potassium phosphate buffer (PB). The nylon membrane was rinsed in distilled water and then soaked in 0.025 M PB before use for transfer. The transfer was carried out by a similar procedure as in the former procedure except this time it was done in 0.025 M PB instead of 20X SSPE. After transferring the RNA to nylon membranes, the membrane was briefly soaked in 0.025M PB and placed on a glass plate with RNA side up. This assembly was covered with Saran Wrap and exposed for 5 minutes to UV light (1200 uW/cm^2) by the use of 2 standard germicidal lamps at a distance of 15 cm. This treatment, which covalently crosslinks RNA to the membrane, increases the sensitivity of detection by at least ten fold and in addition permits the removal of bound probe and further hybridizations with other probes without significant loss of RNA (Khandjian, 1986).

After UV crosslinking of RNA, the nylon membranes were dried by baking under vacuum at 80°C. Prehybridizations and hybridization procedures were similar to Southern blot hybridizations as described for DNA. Since the oligolabelling procedure provided probes with specific activity as high as approximately 2 X 10^9 cpm/ug., this method was utilized in the synthesis of probes for Northern blot analysis.

Screening a Charon 4 C. briggsae Genomic library:

The library was constructed (by T. Snutch) by following the procedure as briefly described below. C. briggsae genomic DNA was partially cleaved with EcoRI restriction endonuclease and size selected for 15-23 kb. range of partial EcoRI fragments on low melting point agarose gels. Ligation of Charon 4 arms to C. briggsae genomic DNA was carried at the molar ratio of 0.85 ug of Charon 4 arms to 0.25 ug of insert DNA by the similar procedure as described for subcloning procedure earlier. Packaging of library was essentially done as described by Maniatis et al. (1982). The original titer of the library was approximately 2 X 10^6 pfu/ ml.

Approximately 25,000 plaques were kindly screened by B. Kuchinka with C2 and C5 cDNA inserts (see page 42 for the description on cDNAs) as probes together. Five positives were purified after three screens by the procedure of Benton and Davis (1977). Hybridization was carried out at moderate stringency.

The purified Charon phage clones were digested with EcoRI to completion, electrophoresed on a 0.7% agarose gel and transferred to nitrocellulose by the procedure of Southern (1977). The filters were baked at 80°C under vacuum to immobilize the transferred DNA, prehybridized and then hybridized with C5 and FH6 probes separately at moderate stringency.

DNA Sequence Analysis:

The plasmids containing fragments of inserts [C2, C5, CbC2 and CeC2] were isolated by the alkaline lysis minipreparation method as described in Maniatis et al. (1982). Approximately 5 ug of supercoiled plasmid DNAs were obtained by gel purification and dissolved in 16 ul of 1X TE buffer (1X TE is 10 mM Tris, 1 mM EDTA, pH 7.5). Each purified template was sequenced essentially as described by Hattori and Sakaki (1982). 16 ul (appoximately 5 ug) of template DNA was mixed with 4 ul of 1 M NaOH, 1 mM EDTA and denatured at room temperature for 5 minutes. 2 ul of filter sterilized 2 M ammonium acetate (pH 5.4) was added prior to precipitation of denatured templates with 150 ul of 95% ethanol. The pellets were rinsed with 300 ul of 70% ethanol and dried under vacuum for 10 minutes. The dried pellets were kept at -20°C and dissolved in 10 ul of ddH_2O just before sequencing. 9.5 ul of templates were mixed with 1 ul universal primer (reverse or forward) and incubated in 70 mM Tris (pH 7.5), 200 mM NaCl, 70 mM $MgCl_2$ and 1 mM EDTA for 15 minutes at 60°C. The reaction mix (12 ul) was kept at room temperature for another 15 minutes and then 20 uCi (2 ul) of alpha ^{32}P-dATP and 2 units of Klenow polymerase were mixed. 3.5 ul of this mixture was quickly aliquoted and mixed in four tubes containing 2 ul of ddGTP, ddATP, ddTTP and ddCTP mixes respectively (Pharmacia) and incubated at 45°C for 15 minutes. 1 ul of 0.5 mM dNTP was added and incubated at 45°C for another 15 minutes.

6 ul of formamide dye (95% formamide, 0.1% bromophenol blue, 0.1% xylenecyanol) was added and kept on ice. The samples were heated at 90°C for 3 minutes and immediately placed on ice before loading on to 6% polyacrylamide/urea gels.

All plasmids containing EcoRI fragments of interest were used to construct deletion derivatives for DNA sequencing. The procedure described by Henikoff et al. (1987b) was employed. A 2.1kb EcoRI/ClaI fragment of CeC2 was subcloned into the Bluescript (M13+) vector to construct further deletions to sequence the 3' end of the gene.

Comparative Sequence Analysis:

Sequence data was entered into a database and analysed using Delaney Sequence Program. Sequence alignments were done using the ESEE program (E. Cabot, personal communication). For comparision of C2 homologous genomic sequences from *C. elegans* and *C. briggsae*, the method as described by Schaeffer and Aquadro (1987) was used. The effective numbers of silent sites and replacement sites were tabulated for the coding regions of CeC2 and CbC2. The distribution of variable sites within the different exons was examined.

C. RESULTS:

Section I

Mapping of sP4 To Genetic Map and Positioning of Cosmids in the unc-22 region:

At the start of my study two separate sets of overlapping cosmids (contigs), known to be recombinationally near unc-22(IV), were obtained from A. Coulson and J. Sulston (MRC, Cambridge, England). One contig consisted of cosmids C11F2, C18D3, and C13G4 (with C18D3 being between the other two). D.M. Moerman (U.B.C., Vancouver, Canada) has established that this contig includes the coding region of the unc-22 gene itself (personal communication). He has shown that the 5′ end of unc-22 is on C13G4, the 3′ end is at one end of C18D3 and that the contig extends to the left of unc-22 with respect to the genetic map. The other contig, which includes the cosmids C02C4 and C08D7, had not been positioned as precisely. The cosmid C02C4 was known to include an N2/BO polymorphic fragment, sP4, that had been identified and positioned very close to unc-22 by Baillie et al. (1985). However, since sP4 had not been recombinationally separated from the unc-22 gene nor had its contig, as yet, been linked to the unc-22 contig, it was not known whether sP4 was to the left or right of unc-22. Therefore in this study, sP4 was mapped relative to various defined deficiencies in the unc-22 region, using strains with the genotypes Dfx; +/let- (m435) nT1(IV;V) (constructed by D.V. Clark; Dfx denotes deficiency sDf2, sDf7,

sDf8, sDf9, sDf10, or mDf7). These nematode strains were grown and genomic DNAs were prepared. Each DNA preparation was individually digested to completion with EcoRI, Southern blotted to nitrocellulose filters and hybridized to a 1.0 kb EcoRI fragment from a plasmid, pCes233 containing DNA which detects the sP4 RFLD on genomic Southerns. The same filter was also hybridized with a 5.8 kb EcoRI fragment from ps18 as an internal control, since ps18 was previously shown to be linked to linkage group I (Rose et al., 1983). Thus ps18 would not be deleted by any of the deficiencies on linkage group IV and therefore its band intensity would represent two copies in contrast to the pCes233 which, if uncovered by a deficiency, would demonstrate only one copy. Fragments from ps18 and pCes233 were chosen as probes because these fragments represent single copy fragments on the genomic blots and thus can be conveniently measured for their intensities. The band intensities resulting from each probe was scanned and their ratios were compared to the ratio in wild type DNA. Figure 2 and Table 1 show the results of these experiments and it is evident from the ratios that only sDf9 demonstrates the wild type pattern of hybridization while sP4 is deleted by all other deficiencies (sDf2, sDf7, sDf8, sDf10, and mDf7). This indicates that the sP4 cosmid cluster is situated to the left of the unc-22 gene on the gentic map (Figure 3).

Figure 2:

Hybridization of ^{32}P-labelled inserts from pCes233 (sP4) and ps18 clones to EcoRI digested genomic DNAs of strains containing various deficiencies (as indicated). The filter was washed at high stringency. The filter was exposed to Kodak Blue Brand film with an intensifying screen at -80°C for 3 days. Corresponding bands for all lanes (as indicated) are shown on the left side of the autoradiogram.

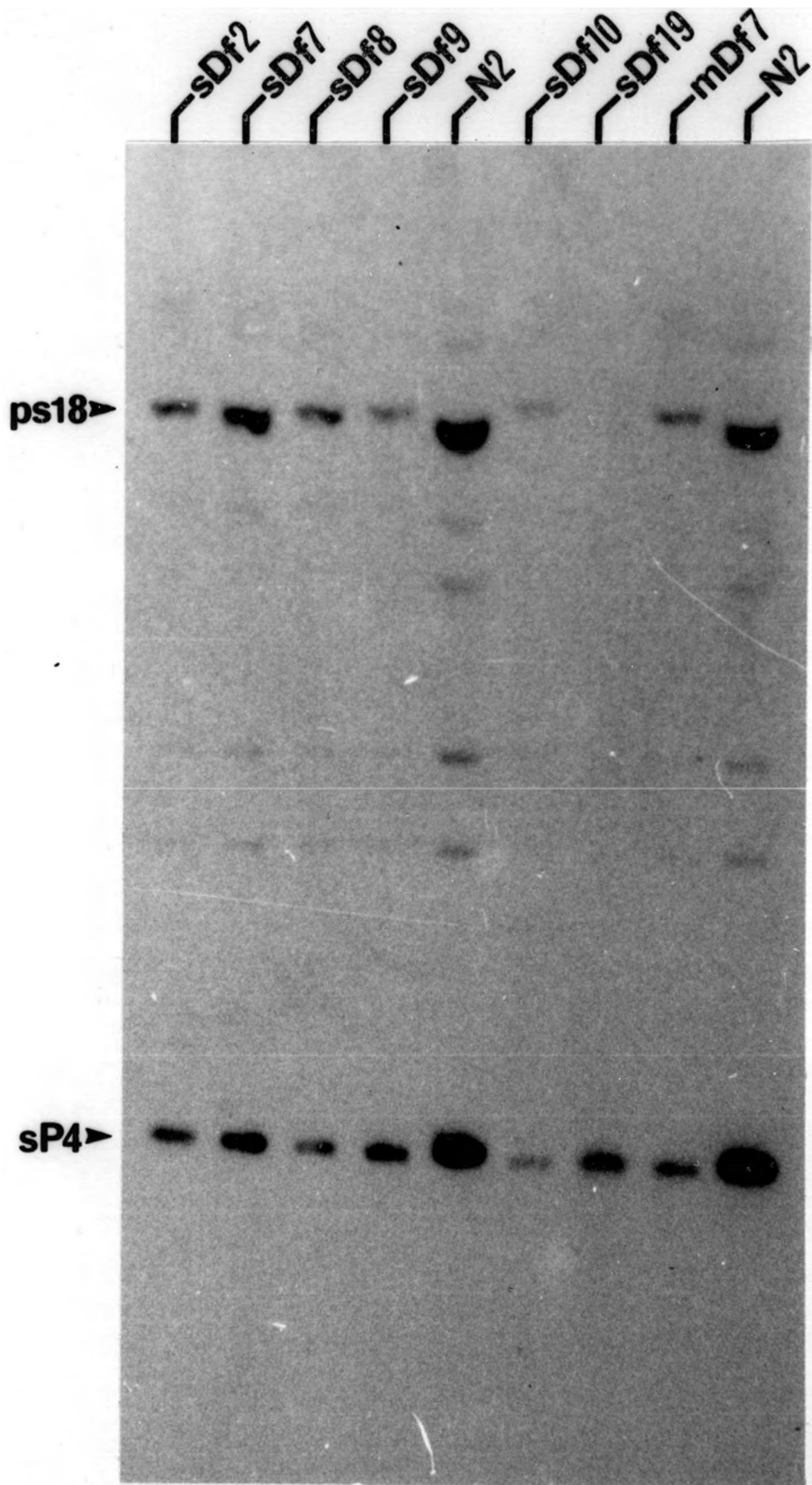

sDf2
sDf7
sDf8
sDf9
N2
sDf10
sDf19
mDf7
N2
ps18
sP4

Table 1:

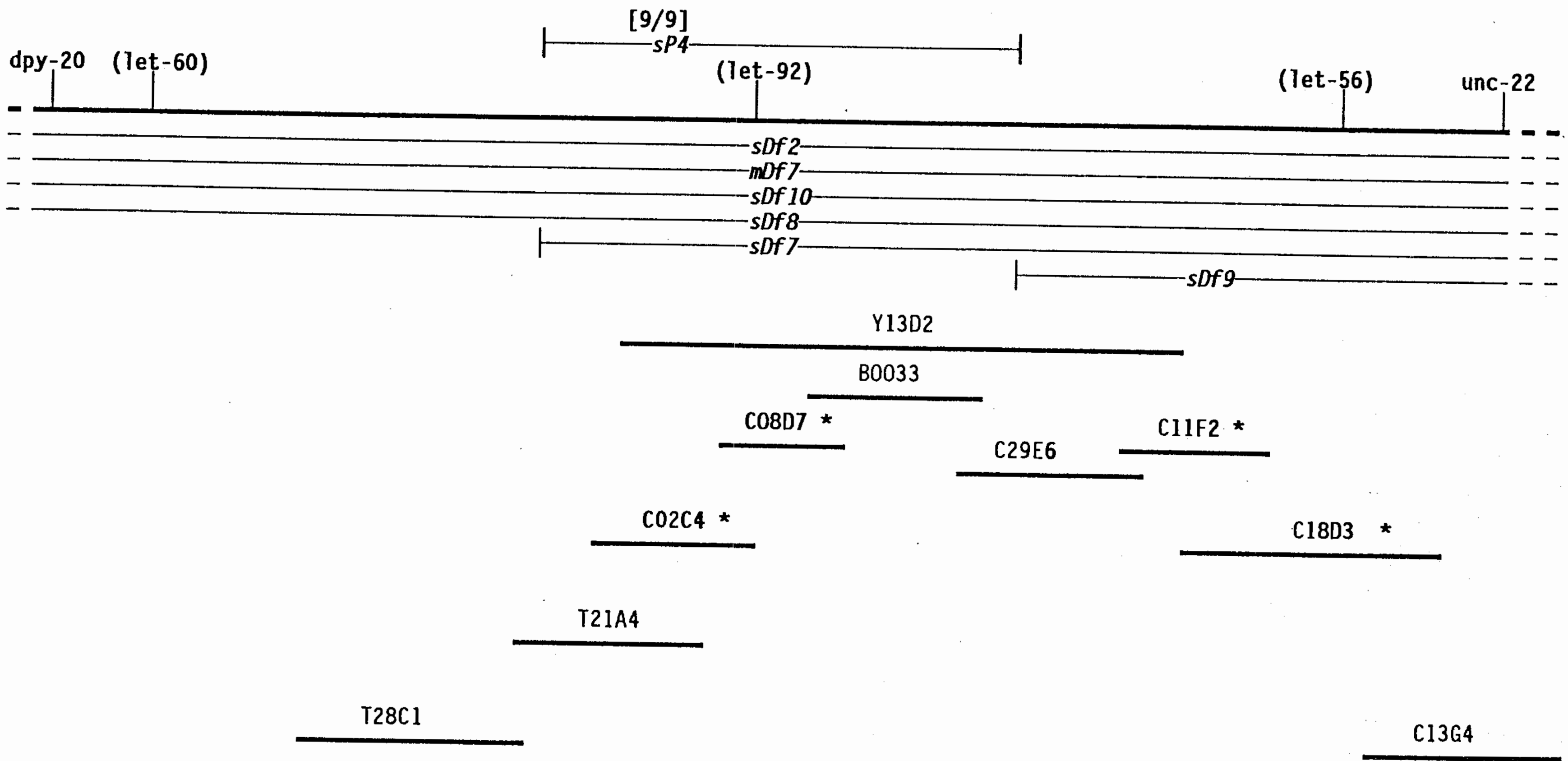
Southern blot analysis of deficiency DNAs with the sP4 fragment

	BAND INTENSITIES WITH PROBE:		
DEFICIENCY STRAIN DNA	pCes233(sP4)	ps18	RATIO OF sP4 vs ps18
sDf2	1.25	1.12	1.12 (1X)
sDf7	2.17	2.45	0.89 (1X)
sDf8	0.80	1.0	0.80 (1X)
sDf9	1.85	0.65	2.85 (2X)
sDf10	0.40	0.30	1.33 (1X)
mDf7	1.27	0.90	1.41 (1X)
WILD TYPE(N2)	9.50	3.65	2.60 (2X)

Band intensities were scanned for each band in the same lane at the same gain control. ps18 was used as control for the amount of DNA in the lanes.

Figure 3:

The alignment of the cosmid contigs (Sulston and Coulson, personal communication) to the genetic map (Clark, Rogalski, Donati and Baillie, 1988) in the dpy-20 - unc-22 interval on linkage group IV. The genetic map is above the physical map. Since the physical map is derived from cosmids (DNA content) while the genetic map is derived from recombination frequencies (map units) there is no direct correlation between the two maps. Cosmids analysed here are shown by asterisks. The position of sP4 with limits as determined by deficiency tests is shown above the genetic map. In the square bracket [] is the fraction of Southern blotted DNAs showing the BO pattern when probed with the sP4 polymorphic fragment (Baillie et al., 1985). The DNAs were prepared from unc-43 (N2);+ (BO) recombinant strains derived from unc-43(N2)unc-22(N2)/+(BO)+(BO) hermaphrodites. The sP4 site is within the C02C4 cosmid.



[9/9]
sP4
dpy-20
(let-60)
(let-92)
(let-56)
unc-22
sDf2
mDf7
sDf10
sDf8
sDf7
sDf9
Y13D2
B0033
C08D7 *
C11F2 *
C29E6
C02C4 *
C18D3 *
T21A4
T28C1
C13G4
C35H3
D1035

Restriction Mapping of Cosmids:

Each cosmid contains approximately 35 kilobases of C. elegans DNA. In order to obtain manageable sized fragments for restriction mapping and to provide a convenient source of DNA for the preparation of labelled probes, PstI and HindIII fragments of these cosmids were subcloned into the appropriate restriction sites of the pUC19 vector. In total approximately 150 kilobases of DNA, representing about 0.3% of C. elegans single copy DNA, were subcloned and restriction mapped with two hexanucleotide recognition endonucleases, namely KpnI and PstI (Figure 4). The fragments derived from the PstI digests were assigned letters "a" to "z" as shown in the Figure 4.

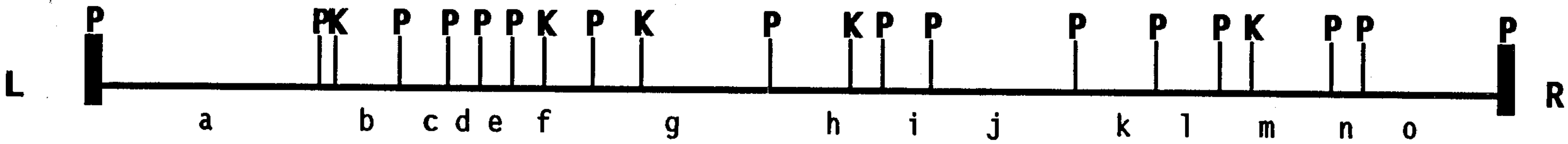
Overlapping cosmids were identified by digesting each cosmid with EcoRI, immobilizing the DNA fragments to nitrocellulose filters and hybridizing individually with nick translated cosmid DNAs. Restriction digestion with EcoRI releases a 5.4kb vector DNA from the rest of the insert DNA. On the basis of their homologies five cosmids have been verified (data not shown) to represent two separate contigs to the left of unc-22 (i.e., two cosmids in the sP4 contig and three cosmids in the unc-22 contig). No overlap was observed between the two sets of overlapping cosmids on hybridizations which confirms the description of cosmids on Figure 3.

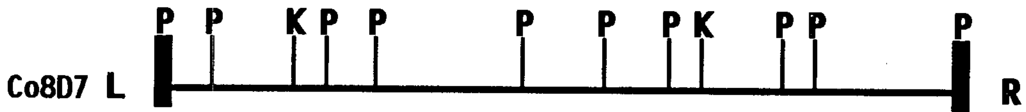
Subsequent to these analyses, two additional cosmids namely, C29E6 and BOO33, have been identified by Coulson et al. to extend to the right of the sP4 contig.

Figure 4:

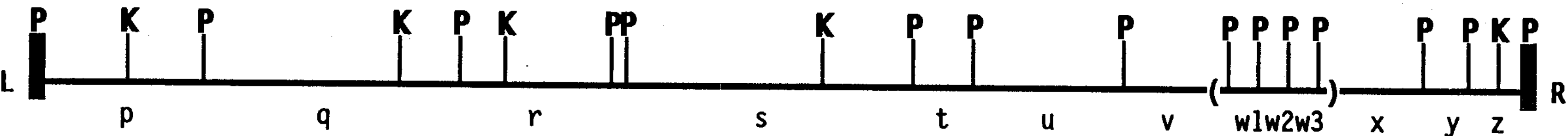
Restriction maps of cosmids. Above each set of overlapping cosmids is a merged restriction map of the contiguous region. The restriction map of cosmids from the unc-22 contig is shown below the restriction map of sP4 cosmids. R (right) and L (left) indicate the orientation of the cosmid with respect to the genetic map. P, PstI; K, KpnI. The PstI sites on each end of the individual as well as merged cosmids is derived from the PstI sites in the vector and is shown by the heavy vertical bars. For convenience in subsequent analyses each PstI fragment has been designated letters "a" to "z".

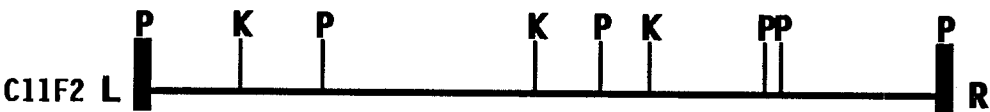
RESTRICTION MAP OF <u>sP4</u> COSMIDS

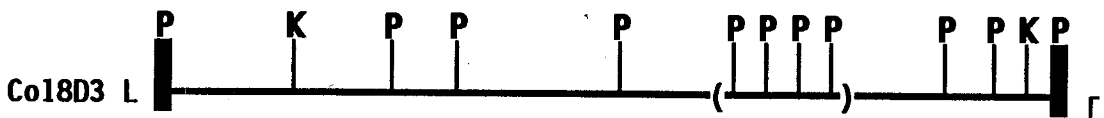




RESTRICTION MAP OF <u>unc22</u> COSMIDS







10 Kilobases

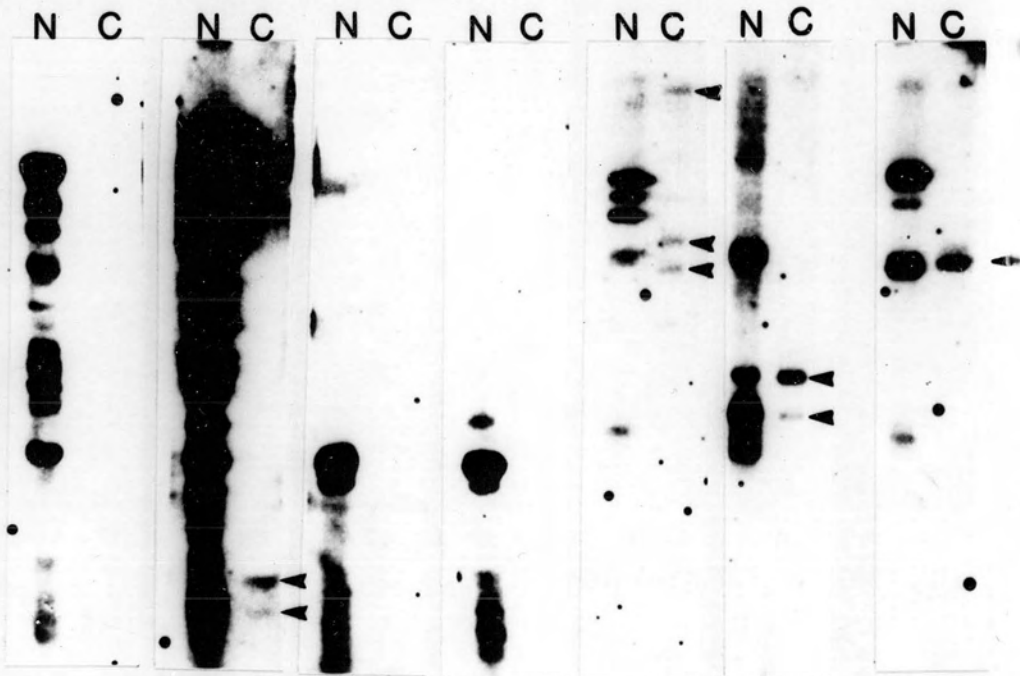
These cosmids have been confirmed to overlap the unc-22 contig by the isolation of a yeast artificial chromosome, Y13D2 (R. Waterston, personal communication).

Detection of transcribed regions:

Having aligned the sP4 contig with the genetic map, and constructed a restriction map of it, I wanted to identify the positions of the transcribed regions within the DNA. Since Northern blot analyses do not allow for the detection of poorly expressed mRNAs, cloned regions were tested for any homology to genomic DNA from the related species, *Caenorhabditis briggsae*. According to a technique suggested by T. Snutch (1984), it was suspected that the highly conserved regions between *C. elegans* and *C. briggsae* DNA code for the proteins necessary for nematode growth, development and reproduction. Therefore, to search for potential coding regions, the labelled subcloned fragments of each cosmid were used as probes and hybridized to nitrocellulose filters with *Eco*RI restricted *C. briggsae* genomic DNA. By this approach nine regions of homology to *C. briggsae* DNA were detected among the four cosmids that were analysed (Figure 5; Table 2).

Figure 5:

Hybridization of ^{32}P-labelled subcloned PstI or HindIII fragments to EcoRI digested C. elegans (N) and C. briggsae (C) genomic DNA. Wash was at moderate stringency. The filters were exposed to Kodak blue brand film at -80°C for 3 days with an intensifying screen. Markers are EcoRI/HindIII digested lambda DNA. The homologies to the corresponding fragments in the C. briggsae genome are denoted by arrowheads.


N C
N C
N C
N C
N C
N C
N C

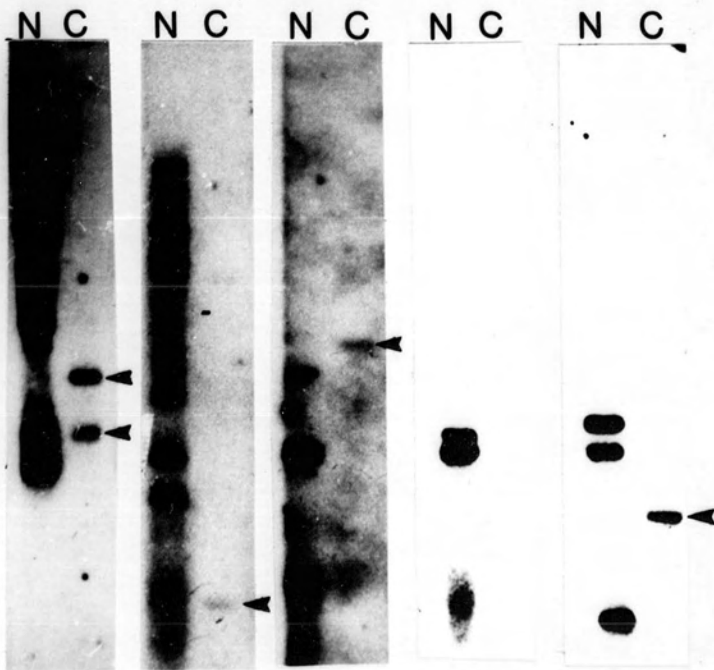

N C
N C
N C
N C
N C

Table 2:
Hybridization test of fragments derived from the restriction map in Figure 4

Fragment designation	*C. briggsae* Homology	Fragment designation	*C. briggsae* Homology
a	ND	p	Y
b	N	q	Y
c	N	r	Y
d	N	s	ND
e	N	t	Y
f	ND	u	N
g	Y	v	Y
h	N	w1-w3	ND
i	N	x	ND
j	ND	y	Y
k	Y	z	Y*
l	N		
m	Y		
n	N		
o	ND		

Y denotes the *PstI* fragment showing homology to *C. briggsae* genomic DNA. Y* is the homology due to the 3′ coding sequences of the *unc-22* gene. The fragments denoted N did not hybridize to *C. briggsae* genomic DNA. The fragment denoted ND were not used in these experiments either due to failure in subcloning or their sizes.

Section II

Northern Blot Analysis:

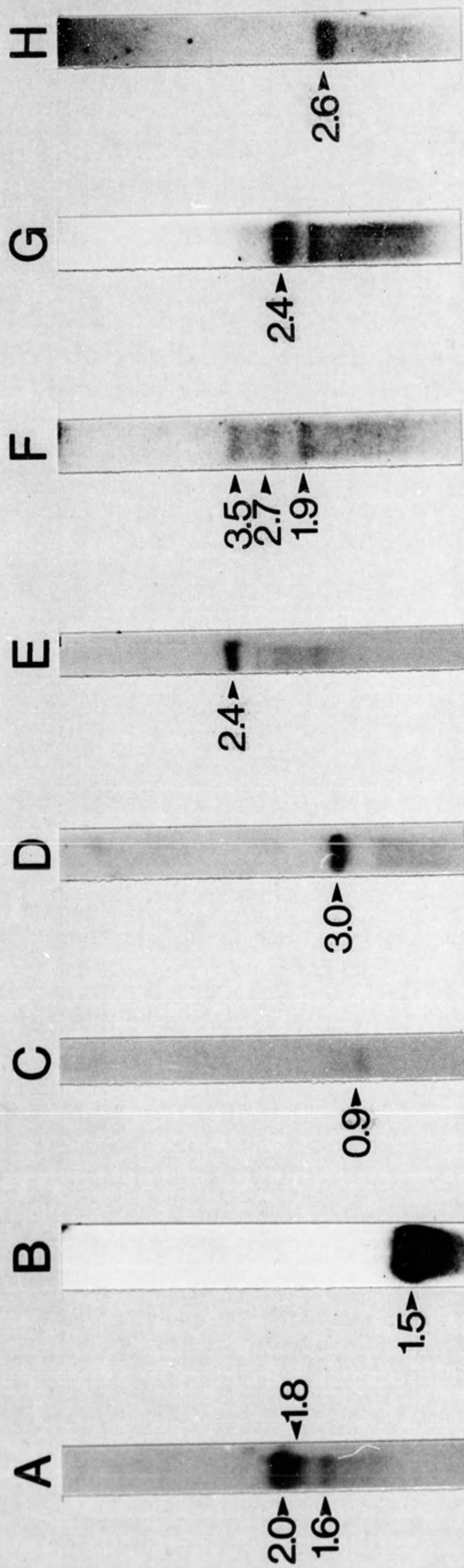
In order to determine if the *C. briggsae* homologous regions of DNA were transcribed, the subcloned fragments were used to hybridize to total RNA from a *C. elegans* population of mixed developmental stages. Out of nine *C. briggsae* homologous fragments, seven fragments showed homologies to the corresponding transcripts, when total RNA was crosslinked to nylon membrane to improve the retention of blotted RNA. g1 and g2 probes were derived from the same g fragments and were identified to contain two separate sources of transcripts. Probes for these experiments were labelled using oligonucleotide primed synthesis because nick translated probes provided inadequate hybridization signals and required much longer exposure time. With the oligonucleotide primed synthesis shorter exposure times of 3 to 24 hours were possible in comparision to 7-day exposures with nick translated probes. The oligonucleotide primed probes had a specific activity of approximately 2 X 10^9/ug. These results are shown in Figure 6.

Screening of a C. elegans Mixed Stage Lambda gt10 cDNA Library:

Having established that seven of the nine *C. briggsae*

Figure 6:

Northern blot hybridization. 10 ug of total RNA from a *C. elegans* populations of mixed developmental stages were blotted to nylon membrane and UV-crosslinked. Hybridization and wash was at high stringency. The filters were exposed to Kodak XAR-5 and Kodak blue brand X-ray films for 3 to 24 hours at -80°C with an intensifying screen. Migrations of the ribosomal RNA bands were used to estimate the sizes of the hybridizing bands with each probe and these sizes are given on the left side of the autoradiograph. Sizes are given in kilobases and the probes derived from the cosmids are A)"g1", B)"g2", C)"k", D)"p", E)"q", F)"r", G)"t", and H)"v".



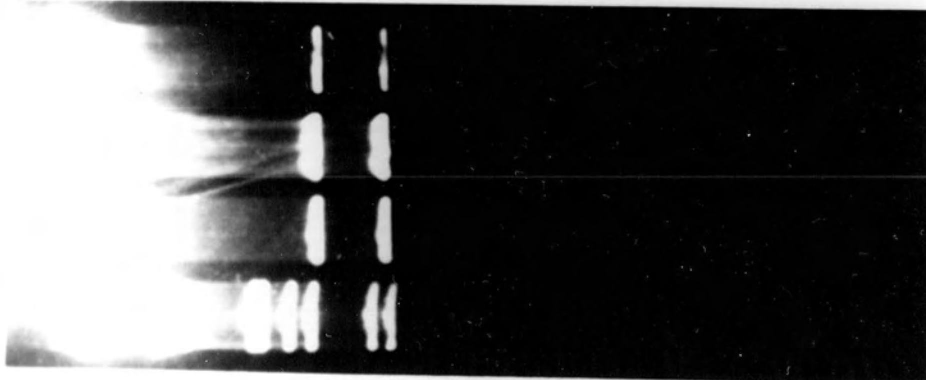
A
2.0
1.6
1.8
B
1.5
C
0.9
D
3.0
E
2.4
F
3.5
2.7
1.9
G
2.4
H
2.6

homologous fragments of cosmids contained the transcribed elements, I was interested in isolating corresponding cDNA clones. A lambda gt10 cDNA library containing reverse transcribed poly(A) RNA of mixed stages of C. elegans (a gift from B. Meyer) was used. Nine of the restriction fragments with homology to the C. briggsae genome (shown in Figure 4) were used as the probes to screen approximately 15,000 phage. Two of the nine C. briggsae homologous fragments hybridized to cDNAs. A 17 kb PstI fragment, designated fragment "q" (see Figure 4 and Table 2 for details) from cosmid C11F2 resulted in the purification of three phage, each of which contained two EcoRI fragments of sizes 0.9 kb and 1.4 kb (coding regions abbreviated C1 and C2 respectively) (Figure 7A). Similarily a 9.2 kb PstI fragment, designated fragment "g", from the cosmid C02C4 resulted in the isolation of three phage, one of which contained a 1.2 kb cDNA insert (abbreviated C4) whereas the other two each contained a 0.5 kb insert (abbreviated C5) (Figure 7B).

Each cDNA insert was subcloned into the EcoRI site of pUC19 for subsequent analyses. The C1 and C2 cDNAs were hybridized to filters containing C11F2 DNA, digested with several restriction endonucleases. Only C2 cDNA showed hybridization to the corresponding fragment(s) of the cosmid C11F2. The 6.2 kb HindIII fragment (FH6), demonstrating homology to C2 cDNA, was subcloned into pUC19 for subsequent analyses. C1 cDNA behaved anomalously in that it hybridized only to a PstI fragment of C11F2 (Figure 8). Therefore C1 was

Figure 7:

Ethidium bromide stained pattern of the cDNA clones isolated from a gt10 cDNA library using fragments A) "q" and B) "g", digested with EcoRI. Markers are EcoRI/HindIII digested lambda DNA.


λ 1 2 3
C2
C1
A

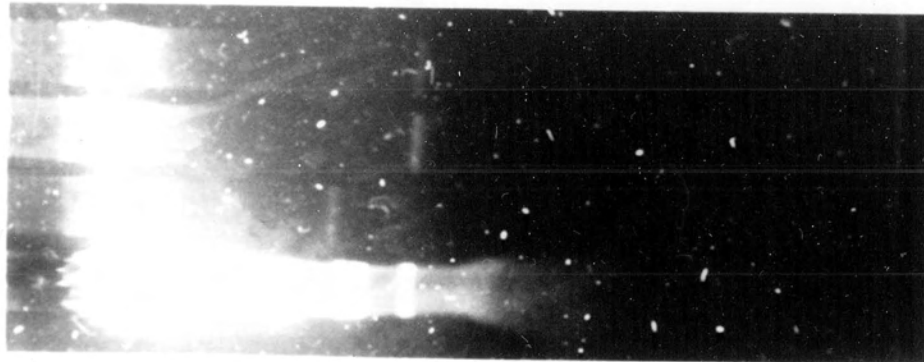

λ 1 2 3
C4
C5
B

Figure 8:

Hybridization of the subcloned *Eco*RI cDNA inserts C1 and C2 to the corresponding cosmid C11F2 DNA. C11F2 DNA was digested with *Pst*I (P), *Bam*HI (B), *Hind*III(H), and *Eco*RI (E). Wash was at high stringency. Filters were exposed to Kodak blue brand film at -80°C for 12 hours with intensifying screen. Sizes of the hybridizing fragments for the C2 cDNA are shown on the left of the autoradiograph.

C1
E H B P
C2
P B H E
17
7.0
6.2
3.2
17

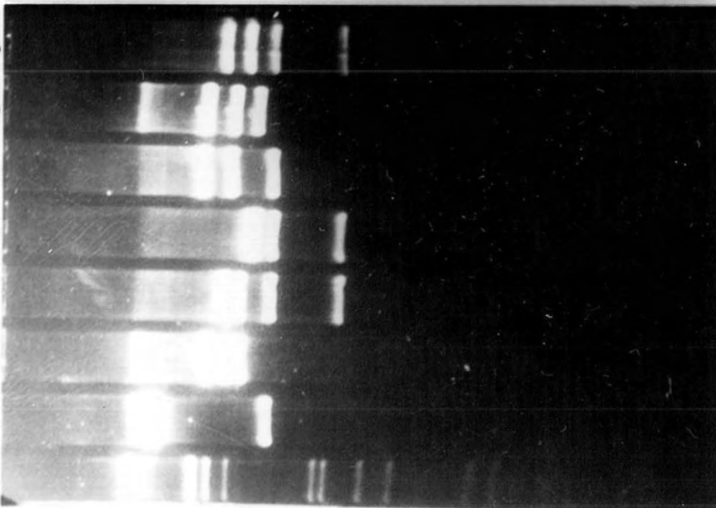
eliminated from further characterization. Similar experiments with C4 and C5 cDNAs showed that both cDNAs correspond to the 9.2 kb PstI fragment ("g") from cosmid C02C4. Further analysis positioned both regions of homology (i.e., with C4 and C5) in a 3.2 kb EcoRI/HindIII fragment of "g". However the two cDNAs hybridized to different subfragments. The C4 cDNA hybridized to a 1.9 kb ClaI fragment while C5 cDNA hybridized to a 3.3 kb ClaI/PstI fragment, thus indicating that C4 and C5 corresponded to two different yet very close regions of the C. elegans genome (Figure 9).

Since in the detection of coding regions, the cosmid probes were derived from either PstI or HindIII fragments of cosmids while the C. elegans and C. briggsae DNAs were digested with EcoRI, the copy number for each region of cosmids in the C. elegans genome could not be established from these experiments. Therefore, cDNA inserts were used as probes and hybridized back to C. elegans and C. briggsae genomic DNA at moderate and high stringencies. The results of these experiments are shown in Figure 10. From these experiments it was shown that C2 is a single copy gene in the C. elegans genome. C4 is encoded by a different gene than C5 but (as shown above) is closely linked to the C5 gene. The fact that C4 hybridizes to different EcoRI fragments on the C. briggsae genome than C5, supports the above notion. At high stringency, C4 and C5 both hybridized strongly to a 3.5kb EcoRI fragment on the Southern blots of C. elegans genomic DNA. However, at moderate stringency C5 hybridized strongly to at least four

Figure 9:

Hybridization of the gel isolated EcoRI cDNA inserts C4 and C5 to the 9.2kb PstI subclone "g". The subclone "g" DNA was digested with PstI (P), EcoRI (E), PstI/EcoRI (P/E), PstI/EcoRI/HindIII (P/E/H), PstI/HindIII (P/H), HindIII (H) and PstI/ClaI, and separated on a 0.9% agarose gel. The electrophoresed fragments were immobilized to nitrocellulose filters and hybridized at high stringency. Filters were exposed to Kodak Blue Brand film at -80°C for 3 hours with an intensifying screen. The photograph of the gel is shown in panel A while the autoradiograph is shown in panel B. Sizes of the hybridizing fragments on PstI/ClaI digestions are shown on the right while the other hybridizing fragments are shown on the left of the autoradiograph.

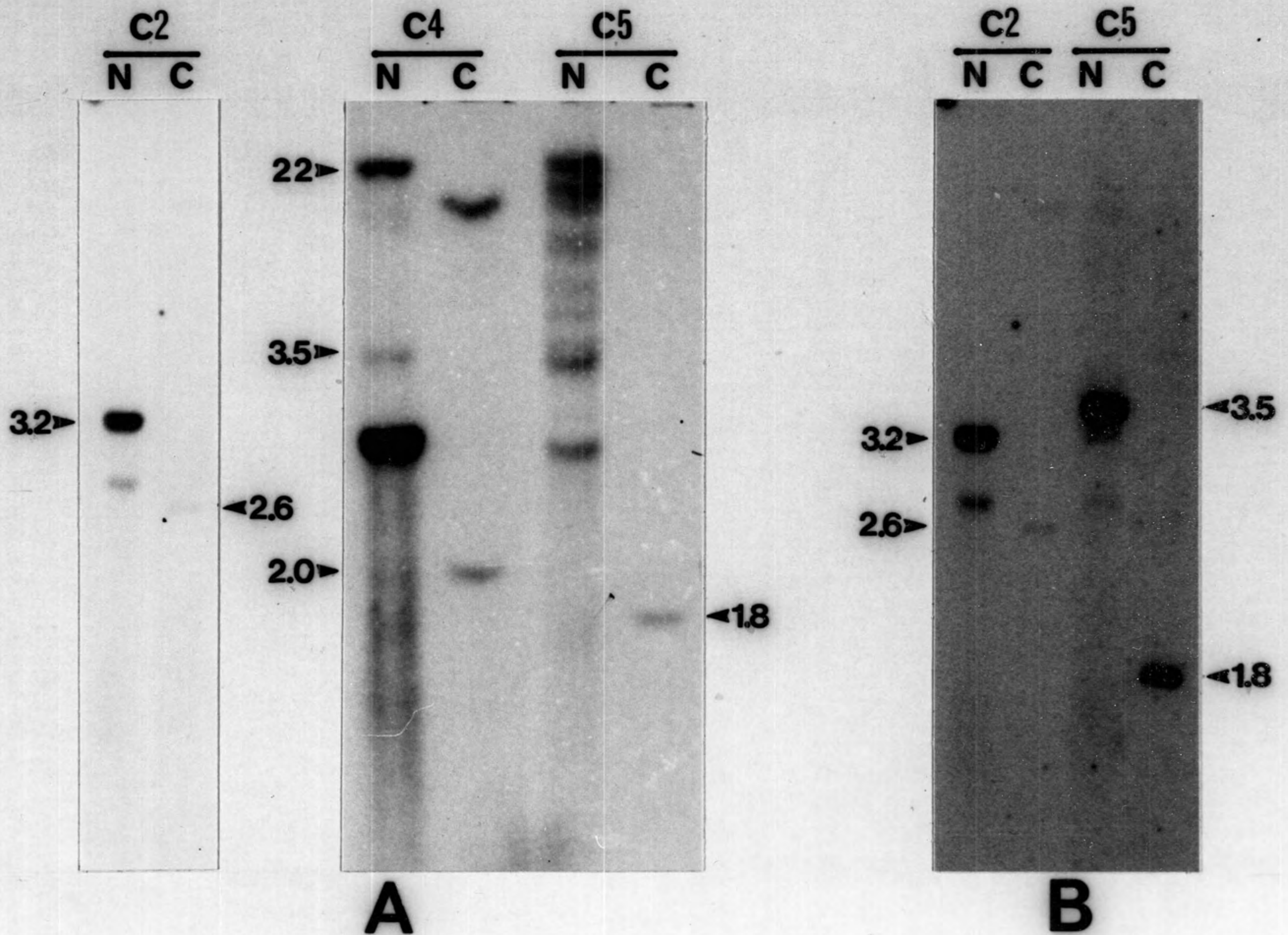
P
E
P/E
P/E/H
P/H
H
P/C
9.2
5.3
3.5
3.2
3.3
1.9
C4
C5
B



λE/H
P
E
P/E
P/E/H
P/H
H
P/C
A

Figure 10:

Hybridization of C2 (1.4 kb), C4 (1.2 kb) and C5 (0.54kb) cDNA inserts to EcoRI digested C. elegans (N) and C. briggsae (C) genomic DNA. Hybridizations and wash were under A) moderate stringency and B) high stringency conditions. The filters were exposed to Kodak Blue Brand film at -80°C for 48 hours with an intensifying screen. Sizes of the hybridizing fragments are shown on the both sides of the lanes. The 2.7 kb hybridizing fragment (unlabelled in this figure) is due to pUC19 DNA contamination in the genomic DNA and therefore must be ignored.


C2
N C
3.2
2.6
C4
N C
C5
N C
22
3.5
2.0
1.8
A
C2
N C
C5
N C
3.2
2.6
3.5
1.8
B

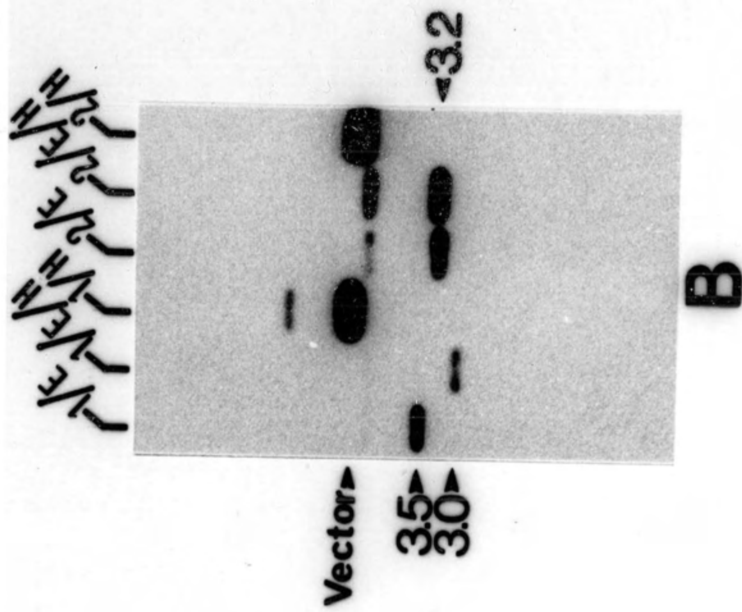
fragments suggesting the presence of a closely related gene family. Since C5 cDNA showed homology to four different EcoRI fragments on the genomic digest of C. elegans, the possibility of its being clustered in the genome of C. elegans was tested by hybridization of C5 to the pooled EcoRI digested cosmids C11F2, C02C4, C08D7, C18D3 and C13G4 at moderate and high stringencies. As expected at high stringency only a corresponding 3.5kb EcoRI fragment on the cosmid (C02C4) hybridized. Interestingly, at moderate stringency more than one fragment showed homology to C5. In order to determine the region on other cosmid(s) with homology to C5, EcoRI and EcoRI/HindIII digested individual cosmids were blotted and hybridized with C5 at moderate and high stringencies. The results of this experiment are shown in Figure 11a and Figure 11b at two different stringencies. Since at high stringency only the corresponding "g" fragment of C02C4 showed hybridization, it may be concluded that C5 is encoded by a gene on the fragment used to isolate the C5 cDNA. At moderate stringency the homology is clearly demonstrated on another cosmid, namely C11F2, suggesting the presence of a closely linked related member of C5 gene family.

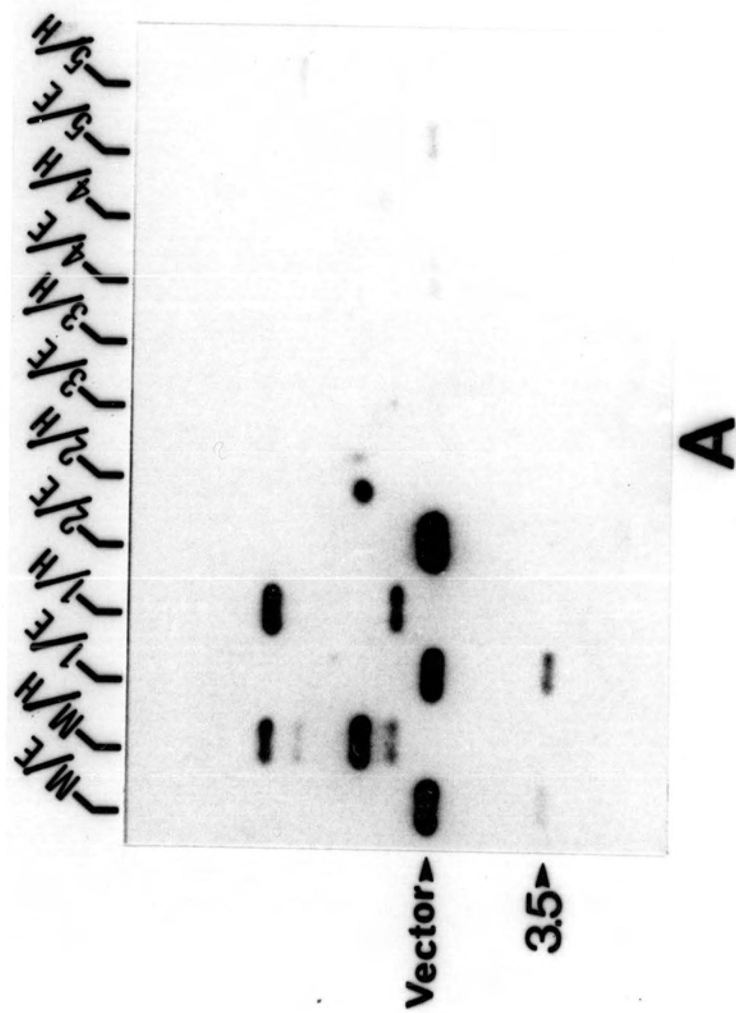
Northern Blot Analysis For The Developmental Stage Specific Expression:

Seven of the C. briggsae homologous fragments hybridized to total RNA from mixed developmental stages. In order to

Figure 11:

Hybridization of C5 cDNA insert to digested cosmids DNA. a) Mixed Cosmids (M), C02C4 (1), C08D7 (2), C11F2 (3), C18D3 (4) and C13G4 (5) were digested individually with *Eco*RI and *Hind*III and hybridized at high stringency. b) Cosmids C02C4 (1) and C11F2 (2) were digested with *Eco*RI (E), *Hind*III (H), and *Eco*RI/*Hind*III (E/H) and hybridized at moderate stringency. Filters were exposed to Kodak Blue Brand film at -80°C for 1 day with an intensifying screen. The homology to vector DNA is indicated on the left of the autoradiogram. Sizes resulting from the hybridization of the C5 cDNA are shown on the both sides of the autoradiogram.


1/E
1/H
1/E
1/H
2/E
2/H
2/E
2/H
Vector
3.5
3.0
3.2
B


M/E
M/H
1/E
1/H
2/E
2/H
3/E
3/H
4/E
4/H
5/E
5/H
Vector
3.5
A

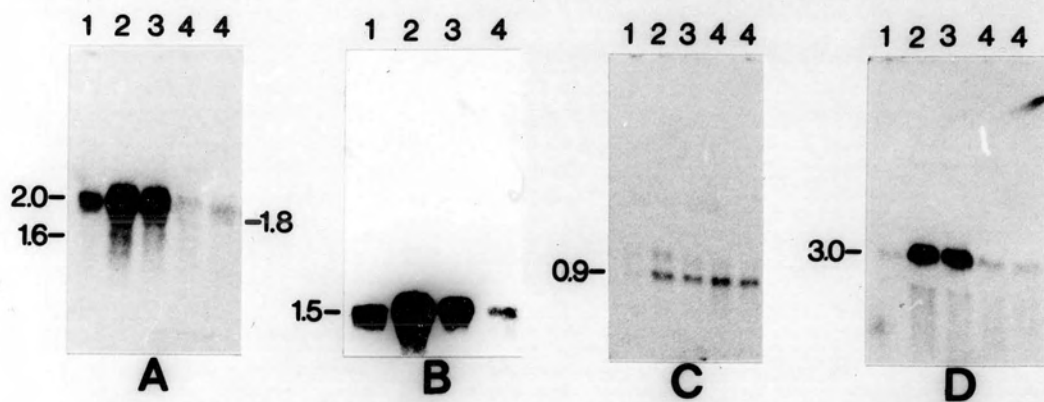
characterize the time of maximum expression for each of these coding regions, each probe was hybridized to RNA from three larval stages and the adult stage of C. elegans. The relative abundance of each transcript in each developmental stage was compared to actin (Krause et al., 1983), hsp-1(A) (Snutch et al., 1988) and P8 (rat Na^+ channel, kindly provided by T. Snutch, CALTECH) genes as shown in Figure 12. The relative level of expression for each gene was determined by scanning the bands from autoradiograms using a gel scanner. The absorbance value (Table 3) in each stage was plotted for expression profile analyses (Figure 13).

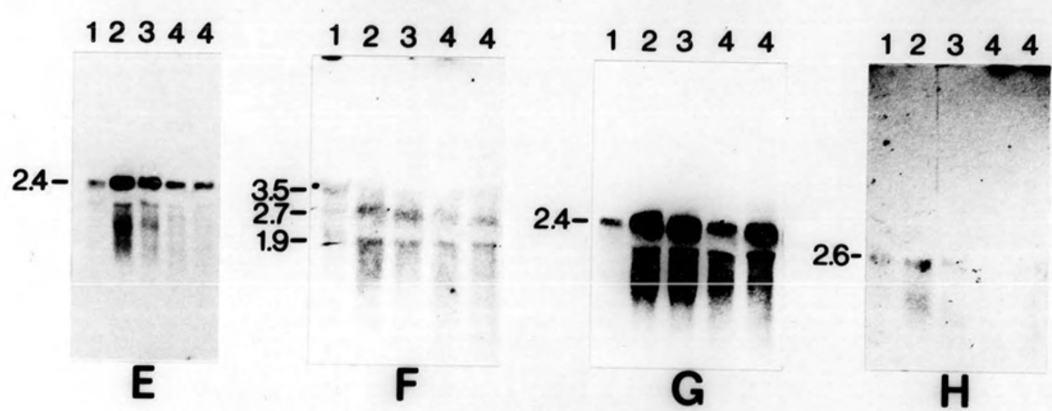
When the total RNA was electrophoresed for a longer duration than as indicated in the Northern blot procedure, C5 cDNA appeared to hybridize to three different lengths of transcripts (2.0 kb, 1.8 kb, and 1.6 kb). The 2.0 kb signal is the most prominent hybridizing band. It seems that these are different transcripts with sequence homology to C5, each of which shows a different pattern of expression during C. elegans development. C4 cDNA which corresponds to the same 3.2 kb EcoRI/HindIII genomic fragment, showed homology to a different transcript of size 1.5 kb with a distinctly different expression profile in comparision to the C5 transcripts (2.0 kb, 1.8 kb, and 1.6 kb). These observations support the previous conclusion that C4 and C5 are derived from two different genes which are closely linked and that C5 is a multi-copy. Six of the eight probes derived from the immediate left of unc-22 showed a peak of RNA expression at the L2 stage

Figure 12:

Developmental expression of identified coding regions on cosmids:

A) C5, B) C4, C)"k", D) "p", E) C2, F) "r", G)"t", H)"v" Controls are as indicated on the autoradiogram [Actin, hsp-1, and Rat sodium channel (P8)].
10 ug of total RNA from the larval stage L1 (1), L2 (2), L3 and L4 (3), Adult (4) and Mixed stages (M) were blotted to nylon membrane and UV-crosslinked. Hybridization and wash was at high stringency. The filters were exposed to Kodak XAR-5 and Kodak blue brand for 3 to 24 hours at -80°C with an intensifying screen. Sizes are given in kilobases.


1 2 3 4 4
2.0-
1.6-
-1.8
A
1 2 3 4
1.5-
B
1 2 3 4 4
0.9-
C
1 2 3 4 4
3.0-
D


1 2 3 4 4
2.4-
E
1 2 3 4 4
3.5-
2.7-
1.9-
F
1 2 3 4 4
2.4-
G
1 2 3 4 4
2.6-
H

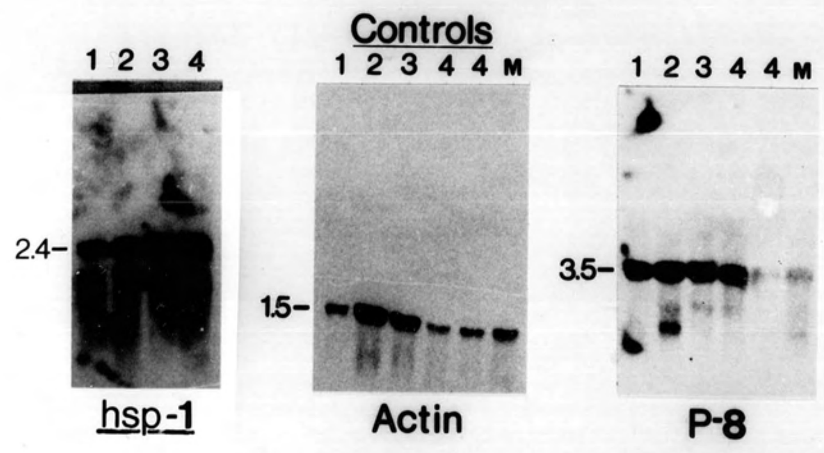

Controls
1 2 3 4
2.4-
hsp-1
1 2 3 4 4 M
1.5-
Actin
1 2 3 4 4 M
3.5-
P-8

Table 3

Relative Expression of C. elegans Probes In Total RNAs Of Different Stages In Development:

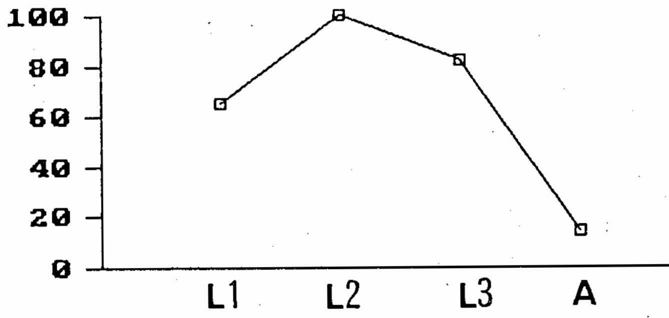
PROBE USED	RELATIVE BAND INTENSITITY AS A PERCENTAGE OF MAXIMUM				
	L1	L2	L3	ADULT	MIXED
g1 (C5)	65	100	82	14	69
g2 (C4)	27	100	69	9	87
k	8	76	61	100	40
p	81	100	94	81	87
q (C2)	13	100	79	18	67
r	20	100	59	40	47
t (C6)	76	91	100	55	76
v	57	100	57	25	100
hsp-1	ND	62	90	100	100
ACTIN	28	100	74	20	41
P8 (Rat)	100	77	74	74	36

Band intensities were scanned and the maximum absorbance was scored 100%. The relative values for other stages were calculated from the individual absorbances from each stage. Since the specific activities and duration of exposures were variable for different probes and filters, the level of expression between different transcribed regions cannot be compared in this experiment.

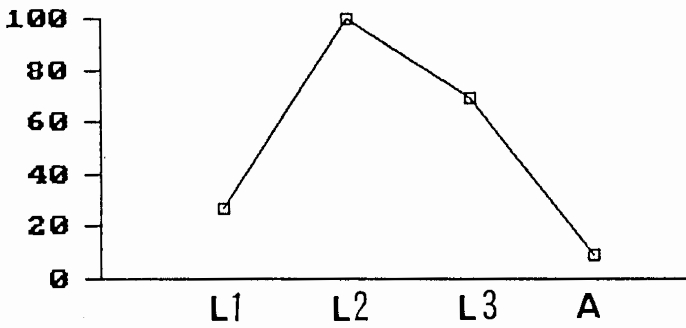
Figure 13:

Relative developmental stage specific expression of each *C. briggsae* homologous region. Hybridization signal with each probe is plotted relative to its maximum mRNA concentration. The level of expression between the different transcribed regions can not be compared in this experiment because different probes of variable specific activity were used and filters were washed and exposed for varying durations.

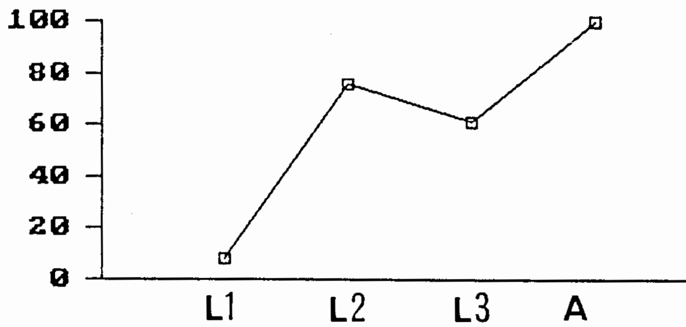
g1



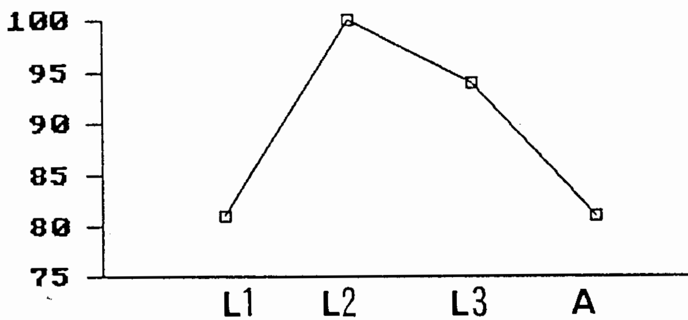
g2

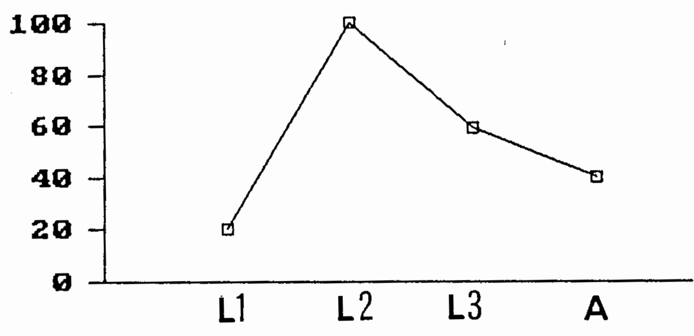


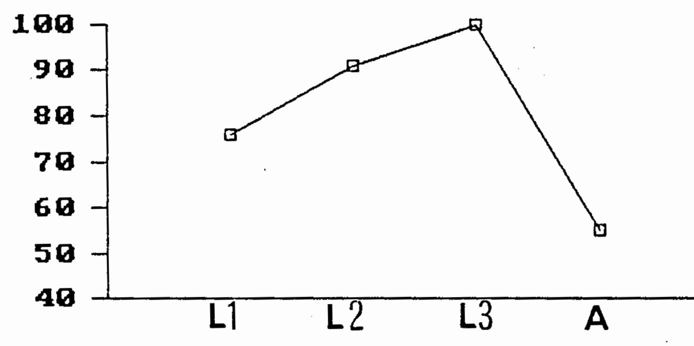
k

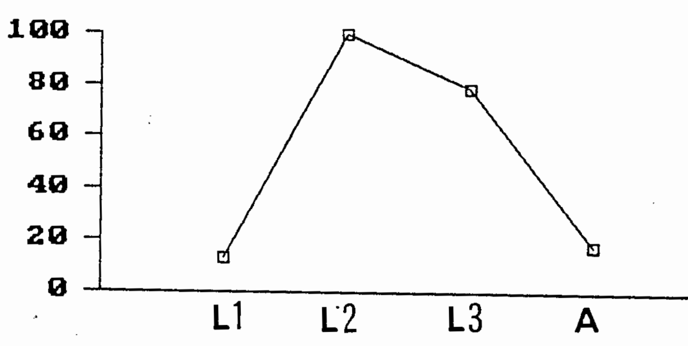


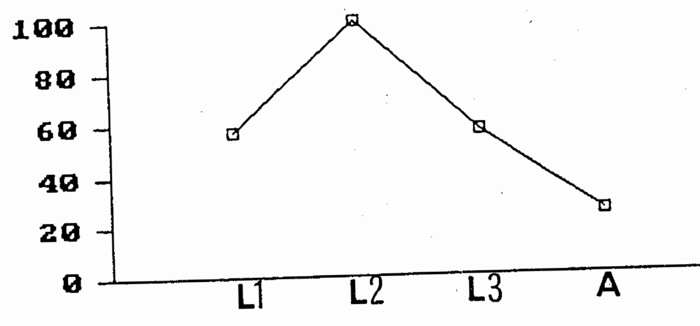
p

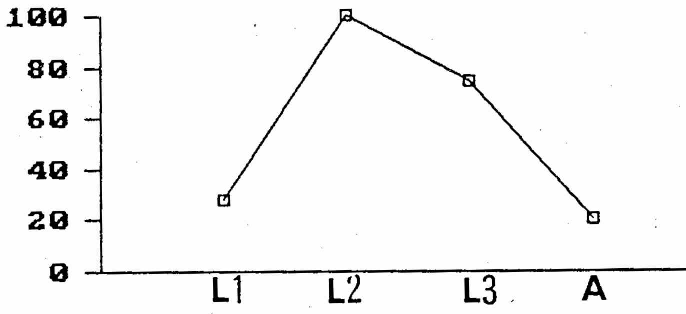



r
100
80
60
40
20
0
L1
L2
L3
A

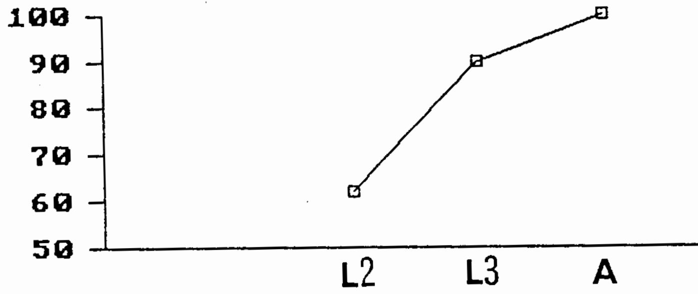

t
100
90
80
70
60
50
40
L1
L2
L3
A


q
100
80
60
40
20
0
L1
L2
L3
A

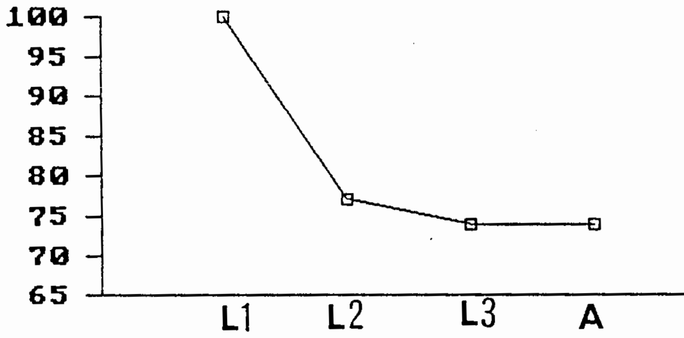

v
100
80
60
40
20
0
L1
L2
L3
A



Actin

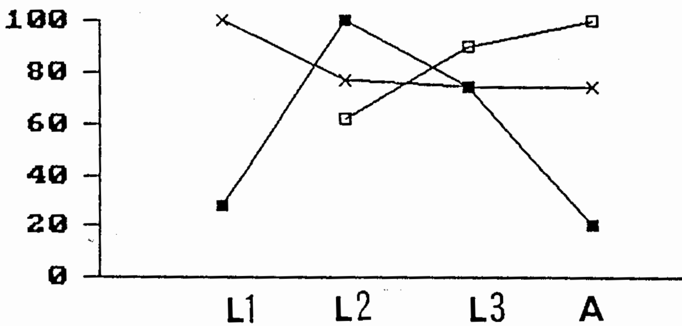


hsp-1



P8

control



hsp-1
Actin
P8 (Na^+)

of C. elegans which declined in later development. This pattern is similar to that of actin, but unlike that of heat shock or Na channel probes. Two others, "t" and "k" peak later at the L3 and adult stages respectively.

Screening of A 24 Hour Stage Specific cDNA Library:

Since most of the C. elegans fragments in this study, that showed homologies to stage specific RNA transcripts, appeared to be expressed in abundance at the mid larval (L2 - L3) stages (refer to Table 3), it was reasonable to expect that the representation of their corresponding cDNA clones in a 24 hour stage cDNA library would be abundant. In order to test this hypothesis, a lambda gt10 cDNA library constructed (by J. Ahringer, University of Wisconsin, Madison) from poly(A) RNA, isolated from the 24 hour stage of C. elegans, was screened with one of the fragments for which no cDNA had been obtained from the mixed stage cDNA library. Approximately 10,000 cDNA phage were screened with a 3.3 kb PstI fragment, designated "t". Three phage, each containing the same three EcoRI fragments, presumably due to the ligation of different cDNAs, were purified. In order to determine the cDNA insert(s) corresponding to the region around unc-22, the cDNA clones were blotted to nitrocellulose filters and hybridized with the "t" probe. Only one band 2.5 kb in size (designated C6) showed homology to the original fragment on the cosmid, hence the other EcoRI fragments are the result of the ligation of

differnt cDNA clones. For subsequent analysis, the 2.5 kb insert was subcloned into the EcoRI site of pUC19 vector. The 2.5 kb cDNA insert was used as a probe to hybridize to stage specific total RNA. As expected, it showed the same pattern of expression as the fragment "t".

Positioning Of The C2 Gene on The Genetic Map:

In order to align the physical map of the unc-22 region more precisely with the genetic map, it is essential that the molecularly identified coding elements be positioned with respect to the zones defined by deficiency breakpoints on the genetic map. Therefore, genomic DNA from strains containing deficiencies around unc-22 (constructed as described earlier) were digested with EcoRI, Southern blotted to nitrocellulose and hybridized with the C2 cDNA insert. The same filter was also hybridized with a 5.8 kb EcoRI fragment from ps18 as an internal control. The ps18 corresponds to a site sP1, previously shown to be linked to the cluster of linkage group I (Rose et al., 1982). Its band intensities represent two copies (or the diploid complement) in contrast to C2 which when uncovered by deficiency(ies) would be present in one copy (the haploid complement). The band intensities resulting from hybridizations with each probe (C2 and ps18) were scanned and their ratios were compared to the ratio observed in wild type (N2) DNA. Table 4 shows the result of such an experiment which shows that C2 gene is uncovered by all the deficiencies (sDf2,

Table 4

Southern Blot Analysis Of Deficiency DNAs
With The C2 cDNA:

BAND INTENSITITIES WITH PROBE:

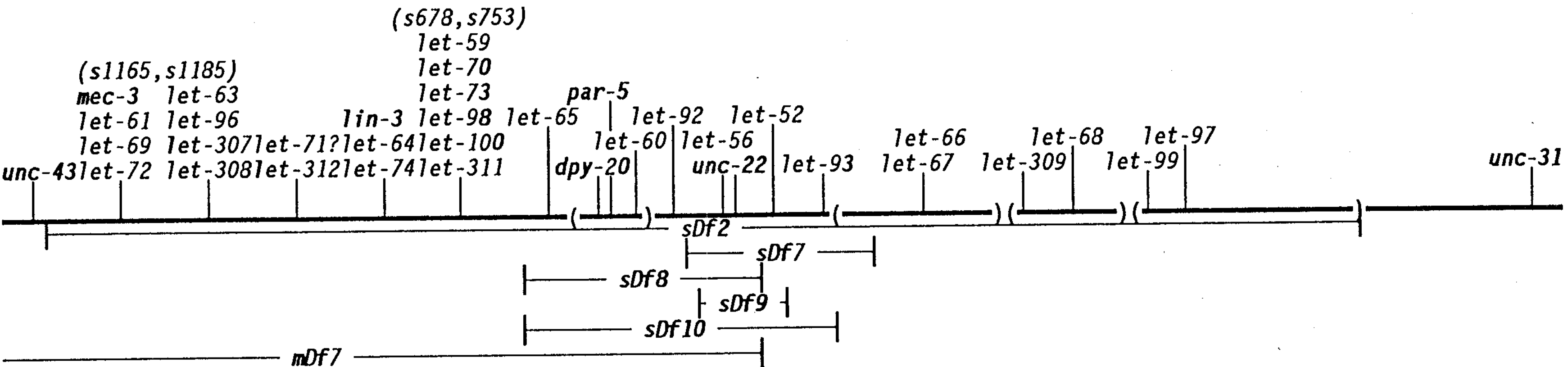
DEFICIENCY STRAIN DNA	C2	ps18	RATIO OF C2 vs pS18
sDf2	1.60	4.70	0.34 (1X)
sDf7	4.80	8.50	0.56 (1X)
sDf8	1.50	2.85	0.50 (1X)
sDf9	1.30	2.75	0.47 (1X)
sDf10	0.15	0.30	0.50 (1X)
mDf7	1.30	3.20	0.40 (1X)
WILD TYPE(N2)	7.60	7.40	1.02 (2X)

Band intensities were scanned for each band in the same lane at the same gain control. ps18 was used as control for the amount of DNA in the lanes.

sDf7, sDf8, sDf9, sDf10, mDf7). Therefore the C2 gene is situated very close to the unc-22 gene within zone 5 of the genetic map (Figure 14).

Figure 14:

The genetic map of the region around unc-22(IV) in *C. elegans* (from Clark et al., 1988). The essential genes are positioned in zones by the use of several mentioned deficiencies. At the bottom is the position of the C2 gene in zone 5.


(s1165,s1185)
mec-3
let-61
let-69
unc-43
let-72
let-63
let-96
let-307
let-308
let-71?
let-312
lin-3
let-64
let-74
(s678,s753)
let-59
let-70
let-73
let-98
let-100
let-311
let-65
par-5
dpy-20
let-60
let-92
let-56
unc-22
let-52
let-93
let-66
let-67
let-68
let-309
let-97
let-99
unc-31
sDf2
sDf7
sDf8
sDf9
sDf10
mDf7
Zone:
1A
1B
1C
1D
2
3
4
5
6
7
8
9
10
7
2
3
7
1
1
1
1
1
1
2
2
2
C2
0.5 map units

Section III:

Sequence Analysis of cDNA clones:

In an attempt to infer the primary amino acid sequence of the isolated cDNA clones, C2 (1.2 kb) and C5 (0.54 kb), both cDNAs were sequenced using the dideoxy method (Hattori and Sakaki, 1986). The entire sequence of C2 and C5 cDNAs is shown in Figure 15 and Figure 16 respectively. Both cDNAs contained an open reading frame and a poly (A) addition site. Translation of C2 cDNA terminates at a UGA (amber) codon which is followed by an AAUAAA sequence 70 nucleotides further on. Similar analysis of C5 showed that its translation terminates at a UAA (ochre) codon which is followed by an AAUAAA sequence six nucleotides downstream. The 3′ ends of both C2 and C5 cDNAs terminate 12 nucleotides downstream of the poly (A) addition sites where a poly A tail is observed. Neither of the isolated cDNA clones contained full length cDNA inserts as would be expected from mRNA transcript sizes based on Northern blots (C2=3.0 kb; C5=2.0 kb).

In order to confirm that C2 cDNA was derived from the corresponding genomic region of the cosmid, a 3.2 kb EcoRI fragment from FH6 (FH6E) earlier found to hybridize to C2 was subcloned into the EcoRI site of the pUC19 vector. The sequence of 2.0 kb from FH6E matched exactly with the sequence of C2 cDNA except where four introns were identified. The 5′ EcoRI site was located only 13 nucleotides upstream of the

Figure 15:

Sequence of C2 cDNA from *C. elegans*. The positions of introns as derived from the corresponding genomic fragment is indicated where present. The poly (A) addition site is underlined. The inferred amino acids are presented below each codon.

```
cDNAC2  ATTCTTGCTCCGGTCTTTATTTTTGTGTTACCTTACATGGCCTACCTTACAGCTGAAATGGTTTCACTTTCCTCCATCATAGCAATCGCAATTTGTGGAATGCTC 105
Protein IleLeuAlaProValPheIlePheValLeuProTyrMETAlaTyrLeuThrAlaGluMETValSerLeuSerSerIleIleAlaIleAlaIleCysGlyMETLeu   35

cDNAC2  ATGAAACAATACATTAAAGGAAATGTTACTCAAGCAGCTGCCAATTCTGTAAAATATTTCACAAAAATGCTTGCTCAATCTTCTGAAACTGTCATCTTCATGTTC 210
Protein METLysGlnTyrIleLysGlyAsnValThrGlnAlaAlaAlaAsnSerValLysTyrPheThrLysMETLeuAlaGlnSerSerGluThrValIlePheMETPhe   70

cDNAC2  CTCGGGCTCTCCACAATTTCTTCTCAACATCACTTTGACCTCTATTTCATTTGCGCAACACTATTCTTCTGTCTCATTTATAGAGCCATTGGAATTGTTGTTCAA 315
Protein LeuGlyLeuSerThrIleSerSerGlnHisHisPheAspLeuTyrPheIleCysAlaThrLeuPhePheCysLeuIleTyrArgAlaIleGlyIleValValGln 105

cDNAC2  TGTTATATTTTGAACCGATTCCGTGCCAAAAAGTTCGAAATGGTTGATCAATTCATCATGTCATATGGAGGTCTTCGTGGAGCCATTGCCTATGGTCTTGTCGTC 420
Protein CysTyrIleLeuAsnArgPheArgAlaLysLysPheGluMETValAspGlnPheIleMETSerTyrGlyGlyLeuArgGlyAlaIleAlaTyrGlyLeuValVal 140

cDNAC2  TCAATTCCAGCTTCAATTACTCGTAAACCAATGTTTATCACTGCGACAATTGCATGGATCTACTTCACTGTATTCCTTCAAGGAATCACAATTCGACCATTGGTC 525
Protein SerIleProAlaSerIleThrArgLysProMETPheIleThrAlaThrIleAlaTrpIleTyrPheThrValPheLeuGlnGlyIleThrIleArgProLeuVal 175

cDNAC2  AACTTTTTGAAGATTAAGAAGAAGGAGGAAAGAGATCCAACGATGGTTGAAAGTGTTTACAATAAATATTTGGATTACATGATGTCTGGAGTGGAAGATATTGCT 630
Protein AsnPheLeuLysIleLysLysLysGluGluArgAspProThrMETValGluSerValTyrAsnLysTyrLeuAspTyrMETMETSerGlyValGluAspIleAla 210

cDNAC2  GGACAGAAGGGGCATTACACTTTCATTGAGAATTTCGAGAGATTCAATGCAAAAGTAATAAAACCAGTATTGATGAGACACCAGAAAAGAGAAAGTTTCGATGCT 735
Protein GlyGlnLysGlyHisTyrThrPheIleGluAsnPheGluArgPheAsnAlaLysValIleLysProValLeuMETArgHisGlnLysArgGluSerPheAspAla 245

cDNAC2  TCATCGATTGTTCGTGCTTACGAGAAAATCACATTGGAAGATGCCATCAAACTTGCCAAAGTCAAGAATAATATTCAAAATAAGCGTCTCGAACGAATTAAGAGC 840
Protein SerSerIleValArgAlaTyrGluLysIleThrLeuGluAspAlaIleLysLeuAlaLysValLysAsnAsnIleGlnAsnLysArgLeuGluArgIleLysSer 280

cDNAC2  AAAGGTAGAGTTGCACCAATTCTTCCCGATAAAATATCCAATCAAAAGACGATGACACCGAAGGATCTTCAATTGAAGAGGTTTATGGAATCTGGTGAAAACATT 945
Protein LysGlyArgValAlaProIleLeuProAspLysIleSerAsnGlnLysThrMETThrProLysAspLeuGlnLeuLysArgPheMETGluSerGlyGluAsnIle 315

cDNAC2  GATTCTCTGTACACGCTCTTCAGTGATCTGCTTGATAGAAAGTTACACGAAATGAATAGACCATCAGTTCAAATTACGGACGTTGATGGACAGGATGATATTCAA 1050
Protein AspSerLeuTyrThrLeuPheSerAspLeuLeuAspArgLysLeuHisGluMETAsnArgProSerValGlnIleThrAspValAspGlyGlnAspAspIleGln  350

cDNAC2  GACGATTACATGGCTGAAGTGTCACGATCGAACCTCTCAGCAATGTTCCGAAGTACGGAACAACTGCCATCAGAAACGCCATTTCATAGTGGTAGAAGACAATCG 1155
Protein AspAspTyrMETAlaGluValSerArgSerAsnLeuSerAlaMETPheArgSerThrGluGlnLeuProSerGluThrProPheHisSerGlyArgArgGlnSer  385

cDNAC2  ACAGGAGATTTAAATGCAACACGAAGAGCTGATTTCAATGTTTGACATTCAATTGAAACCTCTTCGATATGTTTTTGTAGTAATTCTCGCCATAACGAAAAAAAA 1260
Protein ThrGlyAspLeuAsnAlaThrArgArgAlaAspPheAsnVal***---------------------------------------------------------------  399

cDNAC2  CTGTAATAATAATAAATTTATCGCTTTTAAAAAAAA .                                                                   1296
Protein ---Poly A addition site----Poly A tail                                                                  1223
```

Figure 16:

Sequence of C5 cDNA from *C. elegans*. The poly (A) addition site is underlined. The inferred amino acids are presented below each codon.

```
cDNAC5  ATCTATTCATCGGTGTTCGAGTATTCGCCGGTATCGCGTCCGATCGTATTCGTACACTGCCGGAAAAGTCAAAGTTGCGGTTATTCAACACGATCTCATTGCAAG 105
Protein    LeuPheIleGlyValArgValPheAlaGlyIleAlaSerAspArgIleArgThrLeuProGluLysSerLysLeuArgLeuPheAsnThrIleSerLeuGlnV  34

cDNAC5  TACCCGCGTTTTTTCTCATGCTCGTAGTCTTGCTGCCACGCGAGCATCCATATCTACACGTCATCTGCATCACTTTCTATCAGGCTTCATTCGGGTTCAATTGTG 210
Protein alProAlaPhePheLeuMETLeuValValLeuLeuProArgGluHisProTyrLeuHisValIleCysIleThrPheTyrGlnAlaSerPheGlyPheAsnCysG  68

cDNAC5  GAGGATTCTATAAGGGCGCCGCTTTGATATCAAGACAATTCTCTCACTTTGTCATCGGGTATATTCAATTGTTCAAATCGTCAGCAACACTGCTCGAACCTGTTC 315
Protein lyGlyPheTyrLysGlyAlaAlaLeuIleSerArgGlnPheSerHisPheValIleGlyTyrIleGlnLeuPheLysSerSerAlaThrLeuLeuGluProValL 102

cDNAC5  TTTTTTCTCTTCTTGTGCTTCCGGGAAGTGAAGATTCTGAACTCTCATGGACCAGCTACTTCCTGATCCATGCTTTAACACTTACAGTAGCCAACACAGTCTATG 420
Protein euPheSerLeuLeuValLeuProGlySerGluAspSerGluLeuSerTrpThrSerTyrPheLeuIleHisAlaLeuThrLeuThrValAlaAsnThrValTyrV 136

cDNAC5  TTCTCCTAGCTCGTTCAGAACCAGCCGACTTTGTGTTGAACGCTGAATTAGAGCTCAGCAAGCCTCAGCTTCCCTGCGAAACTTCAATCAATTAATTATTTAATA 525
Protein alLeuLeuAlaArgSerGluProAlaAspPheValLeuAsnAlaGluLeuGluLeuSerLysProGlnLeuProCysGluThrSerIleAsn***-----Poly 166

cDNAC5  AATTATTCATATCAG                                                                                          540
        A addition site
```

available length of cDNA insert. The position of introns are indicated in Figure 15.

Since C4 and C5 had earlier been shown to be very closely linked on a 3.2 kb EcoRI/HindIII fragment, this fragment was subcloned and sequenced from both ends to determine the direction of transcription for both genes. This sequencing, in addition, confirmed that the C5 cDNA, a member of a closely related gene family, is expressed from this region of the cosmid. Furthermore it showed that the 3′ end of the gene extends from the EcoRI site towards the HindIII site of the 3.2 kb EcoRI/HindIII fragment. An open reading frame in the opposite direction was identified on sequencing from the HindIII site of the same fragment, suggesting that the C4 gene may be transcribed from the opposite strand and that its 3′ end is close to the 3′ end of the C5 gene. These results are summarised with a detailed restriction map of the 9.2 kb PstI fragment ("g") in Figure 17.

Isolation and Characterization of the C2 and C5 Related Sequence From C. briggsae:

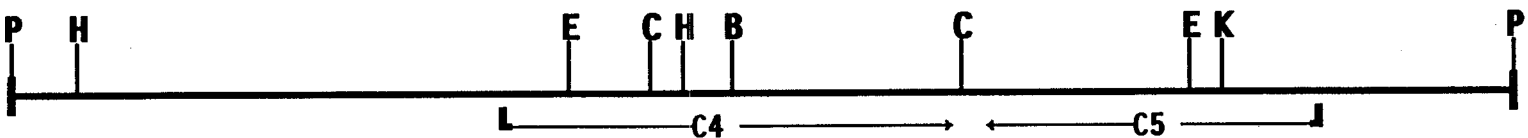
In order to study organization and sequence divergence in C2 and C5 related genes of *C. briggsae*, a Charon 4 *C. briggsae* genomic library was screened with a mixed C2 and C5 probe. Screening of 25,000 plaques resulted in the final purification of five different recombinant phage. Figure 18a shows the ethidium bromide stained pattern of all five phage after

Figure 17:

A detailed restriction map of the fragment "g" and fragment "q". The direction of transcription for each gene (as denoted) is shown by an arrowhead. PstI (P), BamHI (B), ClaI (C), HindIII (H), and KpnI(K).

RESTRICTION MAP OF FH6 ("q") SHOWING C2 HOMOLOGY

RESTRICTION MAP OF CP8 ("g") FRAGMENT SHOWING CODING REGIONS



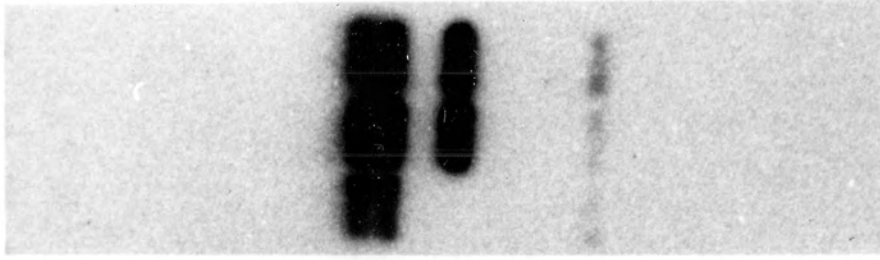
restriction digest with EcoRI and electrophoresis on a 0.7% agarose gel. Hybridization of blots of the gels with the individual C2 and C5 cDNA probes demonstrated that three phage had C2 homologous regions while one phage contained the C5 homologous region. The C2 cDNA hybridized to only a 2.6 kb EcoRI fragment of all three phage while C5 hybridized to a single 1.8kb EcoRI fragment on another phage (Figure 18b). A 6.2 kb HindIII (FH6) fragment, obtained from the original PstI ("q") subclone, was used as a probe to determine the extent of homology towards the 5′ end of the C2 gene. Three EcoRI fragments (2.8 kb, 2.6 kb and 1.3 kb) showed homologies to the FH6 fragment on all three phage, and two phage showed homology to another EcoRI fragment (1.9 kb) (Figure 19b). A restriction map showing the position of C2 and FH6 homologies in the C. briggsae genomic region derived from three overlapping phage is shown in Figure 19.

Sequence Comparision of C2 Homologous Regions Between C. elegans and C. briggsae:

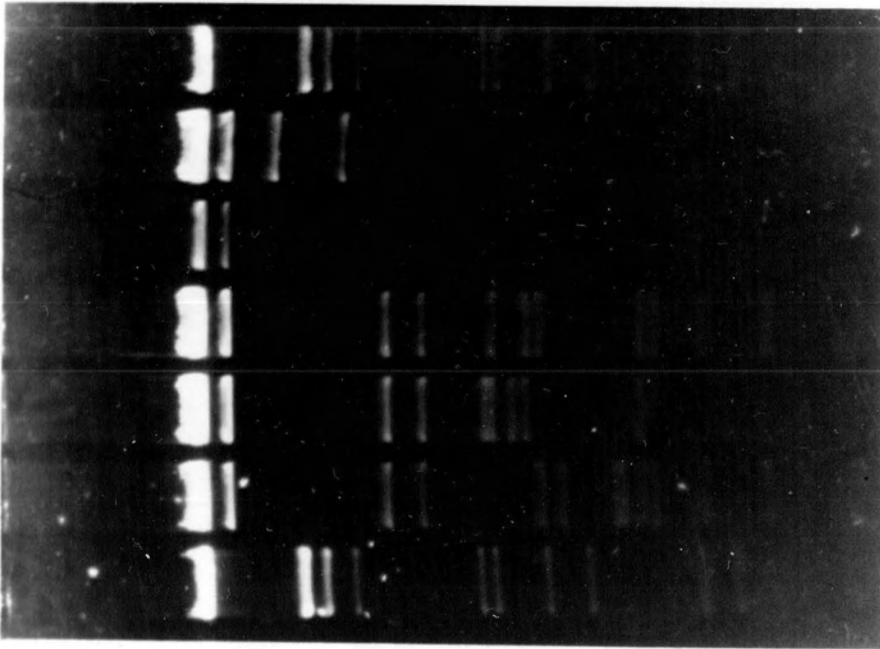
The 2.6 kb EcoRI fragment from the C. briggsae Cb8 phage was subcloned into a pVZ1 plasmid for further analysis. The entire fragment was sequenced and aligned with the sequence of C. elegans fragment FH6E. Nucleotide 1 corresponds to the first base of the C2 homologous region in the FH6E fragment (Figure 20). The 5′ EcoRI sequence and the intron positions of the gene are conserved among these two species. The coding

Figure 18:

Isolation of C2 and C5 related clones from a C. briggsae genomic library. A) Ethidium bromide stained pattern of the five isolated phage digested with EcoRI and electrophoresed on a 0.7% agarose gel. B) Hybridization of ^{32}P-labelled FH6 and C5 sequences to the blots of the gel shown in A). Filters were washed at moderate stringency and exposed to Kodak Blue Brand film at -80°C for 12 hours with an intensifying screen. Sizes of the hybridizing fragments are shown on the left for FH6 (8, 12, 15) and on the right for C5 (3) on the autoradiograph.



Cb
8
Cb
12
Cb
15
Cb
3
2.8
2.6
1.9
1.2
1.8
B



λ
Cb
8
Cb
12
Cb
15
Cb
3
Cb
9
λ
A

Figure 19:

An *EcoRI* restriction map of three overlapping phage (Cb8, Cb12, and Cb15) isolated using C2 as a probe. The regions with homology to C2 cDNA on all phage are shown by striped lines. The region with homologies to FH6 has been indicated by heavy lines. At the bottom is a merged restriction map of the three phage.

RESTRICTION MAP OF C. brigssae GENOMIC REGION WITH C2 & FH6 HOMOLOGY
Cb12
Cb15
Cb8
CB LEFT
RIGHT
L
R
E
KEY:
C2 & FH6
FH6
1 Kb.

Figure 20:

Sequence alignment of C2 homologous regions from *C. elegans* and *C. briggsae*. The amino acids are presented in the coding regions above each sequence. Replacement changes have been underlined in both sequences. The introns are shown where the amino acids are absent. The matches in coding sequences are shown by vertical bars.

```
            IleLeuAlaProValPheIlePheValLeuProTyrMETAlaTyrLeuThrAlaGluMETValSerLeuSerSerIleIleAl------------------- 83
C. elegans  GAATTCTTGCTCCGGTCTTTATTTTTGTGTTACCTTACATGGCCTACCTTACAGCTGAAATGGTTTCACTTTCCTCCATCATAGCgtgagtcttatattgataga 105
C. briggsae GAATTCTTGCACCTGTGTTCATCTTTGTGCTGCCATACATGGCATATCTAACTGCTGAGATGGTTTCACTTTCTTCGATCATCGCgtaagtgttacactggtttc 105
            IleLeuAlaProValPheIlePheValLeuProTyrMETAlaTyrLeuThrAlaGluMETValSerLeuSerSerIleIleAl------------------- 34

            -----INTRON I---------------------------------------------------------------------------------      90
C. elegans  acaatctcgacaactgctgaagttggttgagagttgctacgttagttgacttaaacagatgggatacagaagttttataaaaagccagaattaa        200
C. briggsae aactagaacgaaaaaaatttcctcaatgttttgttttcggtttagaaatgctttcgcatatttgtgaaccagtccctggtagccctcctcgagccctgtcgtag 210
            ------------------------------------------------------------------------------------------------------ 69

                                                                                                                   90
C. elegans              smaller intron                                                                             213
                              |
C. briggsae tttcttttgggatcagtattttggactaacttaatttcatatcagaactcaaatgatatattttgtagtactcgtcgagttctatcatatttagtttagctgg 315
            ------------------------------------------------------------------------------------------------------ 104

                                                                                  aIleAlaIleCysGlyMETLeuMETLysGlnTyrIle 127
C. elegans                           END OF INTRON I--------------------tttcagAATCGCAATTTGTGGAATGCTCATGAAACAATACATT 268
C. briggsae tc---------------------------tttgtattcatcatcattacaaacttttacaatttcagAATCGCAATATGCGGAATGCTAATGAAACAATATATC 392
            -- END OF INTRON I------------------------------------------------aIleAlaIleCysGlyMETLeuMETLysGlnTyrIle 138

            LysGlyAsnValThrGlnAlaAlaAlaAsnSerValLysTyrPheThrLysMETLeuAlaGlnSerSerGluThrValIlePheMETPheLeuGlyLeuSerThr 232
C. elegans  AAAGGAAATGTTACTCAAGCAGCTGCCAATTCTGTAAAATATTTCACAAAAATGCTTGCTCAATCTTCTGAAACTGTCATCTTCATGTTCCTCGGGCTCTCCACA 373
C. briggsae AAAGGAAACATTACCCAAGCAGCTGCAAATTCTGTGAAATACTTTACCAAAATGCTTGCTCAATGCTCCGAAACTGTAATTTTCATGTTCCTTGGTCTATCCACA 497
            LysGlyAsnIleThrGlnAlaAlaAlaAsnSerValLysTyrPheThrLysMETLeuAlaGlnCysSerGluThrValIlePheMETPheLeuGlyLeuSerThr 173

            IleSerSerGlnHisHisPheAspLeuTyrPheIleCysAlaThrLeuPhePheCysLeuIleTyrArgAlaIleGlyIleValValGlnCysTyrIleLeuAsn 337
C. elegans  ATTTCTTCTCAACATCACTTTGACCTCTATTTCATTTGCGCAACACTATTCTTCTGTCTCATTTATAGAGCCATTGGAATTGTTGTTCAATGTTATATTTTGAAC 478
C. briggsae ATTTCATCTCAACATCATTTCGATCTATGGTTCATCTGTGCCACTCTTTTCTTCTGTTTGATCTACCGTGCCATCGGAATCGTTGTTCAATGCTACATTTTGAAT 602
            IleSerSerGlnHisHisPheAspLeuTrpPheIleCysAlaThrLeuPhePheCysLeuIleTyrArgAlaIleGlyIleValValGlnCysTyrIleLeuAsn 208
```

```
            ArgPheArgAlaLysLysPheGluMETValAspGlnPheIleMETSerTyrGlyGlyLeuArgGlyAlaIleAlaTyrGlyLeuValValSerIleProAlaSer 442
C. elegans  CGATTCCGTGCCAAAAAGTTCGAAATGGTTGATCAATTCATCATGTCATATGGAGGTCTTCGTGGAGCCATTGCCTATGGTCTTGTCGTCTCAATTCCAGCTTCA 583

C. briggsae CGCTTCCGTGCCAAAAAATTCGAAGTTGTTGATCAATTCATTATGTCTTATGGTGGCCTTCGTGGAGCCATTGCATATGGTCTTGTGGTTTCCATTCCCGCATCA 707
            ArgPheArgAlaLysLysPheGluValValAspGlnPheIleMETSerTyrGlyGlyLeuArgGlyAlaIleAlaTyrGlyLeuValValSerIleProAlaSer 243

            IleThrArgLysProMETPheIleThrAlaThrIleAlaTrpIleTyrPheThrValPheLeuGlnGlyIleThrIleArgProLeuValAsnPheLeuLysIle 547
C. elegans  ATTACTCGTAAACCAATGTTTATCACTGCGACAATTGCATGGATCTACTTCACTGTATTCCTTCAAGGAATCACAATTCGACCATTGGTCAACTTTTTGAAGATT 688

C. briggsae ATAACTGCTAAACCAATGTTCATCACAGCCACAATTGCAGTCATCTACTTTACAGTATTCCTTCAAGGAATCACTATTAAACCATTAGTAAACTTGCTGAAAGTG 812
            IleThrAlaLysProMETPheIleThrAlaThrIleAlaValIleTyrPheThrValPheLeuGlnGlyIleThrIleLysProLeuValAsnLeuLeuLysVal 278

            LysLysLysGluGluArgAspProThrMETValGluSerValTyrAsnLysTyrLeuAspTyrMETMETSerGlyValGluAspIleAlaGlyGlnLysGlyHis 652
C. elegans  AAGAAGAAGGAGGAAAGAGATCCAACGATGGTTGAAAGTGTTTACAATAAATATTTGGATTACATGATGTCTGGAGTGGAAGATATTGCTGGACAGAAGGGGCAT 793

C. briggsae AAGAGAAAGGAGGAGAGAGATCCAACAATGGTTGAAAGTGTTTACAACAAGTACTTGGACTATATGATGTCAGGTGTTGAAGATATTGCCGGTCAGAAAGGACAT 917
            LysArgLysGluGluArgAspProThrMETValGluSerValTyrAsnLysTyrLeuAspTyrMETMETSerGlyValGluAspIleAlaGlyGlnLysGlyHis 313

            TyrThrPheIleGluLy------------------INTRON II----------------------------------------------------------- 677
C. elegans  TACACTTTCATTGAGAAgtaagttctttttttgaagtaggtggat----gcttatgtaggcataaaatactctcgcctacctacacctgcattgtgcctatacgc 894

C. briggsae TATACTTTTATTGAAAAgtaagttcaactgaaaaatatcttttgttt  SMALLER INTRON                                          977
            TyrThrPheIleGluLy------------------INTRON II----------------------------------------------------------- 348

            ------------------------------------------------------------------------------------------------------- 677
C. elegans  aggcctacagcaaaagtgcacttgactactttattcctttcacagttctaggtgtcatacctataattttctgtttatacatttttctgtaaatgcctaaacaa 999

C. briggsae                                                                                                         977
            ------------------------------------------------------------------------------------------------------- 383

            ------------------------------------------END OF INTRON II--sPheGluArgPheAsnAlaLysValIleLysProValLeuME 732
C. elegans  taactgtatttcggagtatacggtagtctattattatcttttcaataactattataatttcagTTTCGAGAGATTCAATGCAAAAGTAATAAAACCAGTATTGAT 1104

C. briggsae                                           acatcttattatttctagTTTTGAAAGATTCAATGCAAAAGTTATCAAACCTGTTTTGAT 1037
            ---------------------------------------------------------sPheGluArgPheAsnAlaLysValIleLysProValLeuME 418
```

```
            TArgHisGlnLysArgGluSerPheAspAlaSerSerIleValArgAlaTyrGluLysIleThrLeuGluAspAlaIleLysLeuAlaLysValLysAsnAsnIl 837
C. elegans  GAGACACCAGAAAAGAGAAAGTTTCGATGCTTCATCGATTGTTCGTGCTTACGAGAAAATCACATTGGAAGATGCCATCAAACTTGCCAAAGTCAAGAATAATAT 1209
            |||||| || ||||||||||| || |||||||| || ||||||||||||||||||||||||| ||| | |||||||| |||||    | || ||||| ||||||| |||
C. briggsae GAGACATCAAAAAAGAGAAAGCTTTGATGCTTCTTCAATTGTTCGTGCTTACGAGAAAATTACACTTGAAGATGCGATCAAGTTGGCAAAAGTTAAGAATACTAT 1142
            TArgHisGlnLysArgGluSerPheAspAlaSerSerIleValArgAlaTyrGluLysIleThrLeuGluAspAlaIleLysLeuAlaLysValLysAsnThrIl 453

            eGlnAsnLysArgLeuGluArgIleLysSerLysGlyArgValAlaProIleLeuProAspLysIleSerAsnGlnLysThrMETThrProLysAspLeuGlnLe 942
C. elegans  TCAAAATAAGCGTCTCGAACGAATTAAGAGCAAAGGTAGAGTTGCACCAATTCTTCCCGATAAAATATCCAATCAAAAGACGATGACACCGAAGGATCTTCAATT 1314
              || |||||| | || ||| |||| |||||   |      |  || ||||||||||| || |||||||| || |||||||  |||||||||| || |||||||||||
C. briggsae CCAGAATAAGAGACTAGAAAGAATAAAGAGTGAGAACAAGGTAGCACCAATTCTCCCGGATAAAATTTCGAATCAAACAACGATGACACCTAAAGATCTTCAATT 1247
            eGlnAsnLysArgLeuGluArgIleLysSerGluAsnLysValAlaProIleLeuProAspLysIleSerAsnGlnThrThrMETThrProLysAspLeuGlnLe 488

            uLysArg--------INTRON III------------END OF INTRON III----PheMETGluSerGlyGluAsnIleAspSerLeuTyrThrLeuPheS 1018
C. elegans  GAAGAGGtttgtagatggaaatagaatatggactttaatttcaaaaag----tttcagTTTATGGAATCTGGTGAAAACATTGATTCTCTGTACACGCTCTTCA 1415
            ||| | |||    |   | | | |   |   |  |  |    ||  | ||       |||||||| ||||||||||| || ||||| |||||  ||||||| | ||||
C. briggsae GAAAAAGgtaaatcaaagaatttcataattatataaaaaaatgaatgtatgaatttcagTTCATGGAATCTGGGGAGAACATCGATTCCTTGTACACTTTGTTCA 1352
            uLysLys----------------------------END OF INTRON III----PheMETGluSerGlyGluAsnIleAspSerLeuTyrThrLeuPheS 522

            erAspLeuLeuAspArgLysLeuHisGluMETAsnArgProSerValGlnIleThrAspValAspGlyGlnAspAspIleGlnAspAspTyrMETAlaGluVal- 1122
C. elegans  GTGATCTGCTTGATAGAAAGTTACACGAAATGAATAGACCATCAGTTCAAATTACGGACGTTGATGGACAGGATGATATTCAAGACGATTACATGGCTGAAGTGg 1520
            ||||| || | ||||| ||| |||| |||||||| | | |  ||| |||||  | ||| ||||| ||   | || |||||||| || || ||||||||||||||||
C. briggsae GTGATTTGTTGGATAGGAAGCTACATGAAATGAACACAACGCCAGCTCAAAAGAAGGAAGTTGAGGGTGAAGACGATATTCAGGATGACTACATGGCTGAAGTGg 1457
            erAspLeuLeuAspArgLysLeuHisGluMETAsnThrThrProAlaGlnLysLysGluValGluGlyGluAspAspIleGlnAspAspTyrMETAlaGluVal- 557

            -------------------------INTRON IV-------------------------------------------------------------------- 1130
C. elegans  ttagttttattttagtaccaagatttcgaaaaactttttaaaaatattgactttctgaaaacatcttctgaactttctaaaaatctagatatatgatctcgatc 1625
            | |||||| ||| |            |    ||| ||||    ||    |               || |          |   || |    ||       ||   ||
C. briggsae taagttttgtttaatatttggcccataaggaaaagtttcctaaatctcgaaaaaacgaatccaaaaaatagtattatcgtaaaacctgtaatagaagagtcct 1562
            -------------------------INTRON IV-------------------------------------------------------------------- 591

            ----------------------------------------------------------------------------                          1130
C. elegans  aagttcaatgtttatcaattttctttatcgaattttagaa-----elegansgap1---------------------------------------------- 1677
                         | | | |   | |          ||||   |
C. briggsae ctaactttgaactttgagtcaatttaccgtatttttctaatagaaccccacctctatttgaagcccacttctaatagaagcccaccctggctgtgcatttttct 1667
            ------------------------------------------------------------------------------------------------------ 626
```

```
                                                                                                                  1130
 C. elegans  ---------------------------------------------------------------------------------------------------- 1677

C. briggsae  ttttgtttagtgcatacatatttttttgtttattttctggtaattcccgttaattcttgtcagttcagttaaaaaccgaaaccctctaatttgagagtttc 1772
             ----------------------------------------------------------------------------------------------------  661

                                                                                                                  1130
 C. elegans  ---------------------------------------------------------------------------------------------------- 1677

C. briggsae  tctaaaatttgttcatccccagaaggaagttaacaaatagaagcccatgggcctctaatagaggaaataaggtatctcaatacctgtcaatgatataataaagct 1877
             ----------------------------------------------------------------------------------------------------  696

                                          END OF INTRON IV                   SerArgSerAsnLeuSerAlaMETPheArgSerT 1177
 C. elegans  ---------------------------------------------------------------nnnnngaTCACGATCGAACCTCTCAGCAATGTTCCGAAGTA 1718
                                                                                                          | ||||
C. briggsae  gtcatctgataagttgttcatcataacatcagaagttttgtcagttatcacttttggtatcttca.................................CCAAGTG 1951
             ------------------------------END OF INTRON IV----------------                                ProSerA  720

             hrGluGlnLeuProSerGluThrProPheHisSerGlyArgArgGlnSerThrGlyAspLeuAsnAlaThrArgArgAlaAspPheAsnVal***---------- 1269
 C. elegans  CGGAACAACTGCCATCAGAAACGCCATTTCATAGTGGTAGAAGACAATCGACAGGAGATTTAAATGCAACACGAAGAGCTGATTTCAATGTTTGAcattcaattg 1823
             | |||||||||||||||| |  |||||||| | || ||||||||||||| |||||||| ||||||||   |||||| |||||| |||||||| ||    ||
C. briggsae  CTGAACAACTGCCATCAGAAGCAGCATTTCATCGAGGAAGAAGACAATCGACCGGAGATTTGAATGCAACGAAAAGAGCCGATTTCGATGTTTGAaatgttttta 2056
             laGluGlnLeuProSerGluAlaAlaPheHisArgGlyArgArgGlnSerThrGlyAspLeuAsnAlaThrLysArgAlaAspPheAspVal***----------  755

             ----------------------------------------------------------Poly A addition----                         1282
 C. elegans  aaacctcttcgatatgtttttgtagtaattctcgccataacgaaaaaaaactgtaataataataaatttatcgcttta                        1902
             ||  |       | |||  | ||| | ||   | ||  |      |   | || | | |  |||
C. briggsae  aactttaaattcgaaatttaatttgttaaactttgaacacagatttcaaaaacgactaataaattatgtatatttgcgaacacaacttctaattgaatttcta 2161
             ------------------------------------------------------- Poly A addition ------------------------------  789

 C. elegans

C. briggsae  attagactgtttttgttttgaaagaattc                                                                              2191
             --------------------------                                                                                 798
```

sequences showed the highest degree of identity (80% at the nucleotide level and 91% at the amino acid level) between the two species. The introns in *C. elegans* gene range from 48 to 253 nucleotides while in *C.briggsae* these range from 48 to >450 nucleotides.

Only a small proportion of nucleotide changes (synonymous or silent substitutions) in coding sequences can change without altering the primary structure of the gene product. This proportion was calculated for the five such amino acid codons which only allow a change in the third position without replacing the amino acids. 26.6% of all codons (106 of 399) in the available C2 homologous sequences from *C. elegans* and *C. briggsae* are synonymous codons. Thirty-nine of such codons have accumulated nucleotide differences between the C2 genes of these two species. Therefore, 45.3% of silent sites (48 of 106) have observed changes between the species being examined. The sequence divergence, using the formula (Jukes and Cantor, 1969) which corrects for multiple hits between two species being compared, is 70%. With the assumption that both species have evolved at a constant rate the sequence divergence between *C. elegans* and *C. briggsae* since the latest common ancestor is about 35% (Nei, 1987). Therefore with the assumption of divergence in time scale to be about 1%/MYr. (Ochman and Wilson, 1988), the divergence time since the latest common ancestor for *C. elegans* and *C. briggsae* would be in the order of 35 MYr. Silent substitutions were distributed evenly across the four exons except the first exon had more changes (60%).

These results are summarized in Table 5.

Thirty-two amino acid replacements out of 399 amino acids have accumulated between the two available C2 sequences. The effective number of replacement sites are derived by subtracting the number of effectively silent sites (106) from total number of coding sites (399). Thus, an average of 10.9% replacement sites are observed between C. elegans and C. briggsae. Replacement sites between these two species were unevenly distributed among the four exons; the first exon had no amino acid replacement while the fourth and fifth exons had a high number of amino acid replacement (Table 6). Most amino acid replacements were of a conservative type i.e., the amino acids of similar charge were substituted.

The positions of introns in the C2 genes are identical in C. elegans and C. briggsae although sizes of three introns are larger in the C. briggsae gene than in the C. elegans gene. The presence of each intron was directly inferred from the sequences of cDNA clones which overlap these regions. Each intron contains consensus splice sites at their 5′ and 3′ boundaries in both species (Table 7); each intron contains stop codons in all three reading frames and are at least 70% A+T. Comparision of introns, allowing for insertions and deletions, showed homologies between 60-67%. In addition to the expected conserved regions at the ends of the introns, the sequence AATTTCA was found to be conserved in three of the four introns examined. Similar alignments of 3′ flanking sequences showed a homology of 68%.

Table 5:

Percent Divergence Between C. elegans and C. briggsae:

Region	Length In bp.	No. of Synonymous Codons	No. of Substitu-itions	% Divergence Uncorrected	Corrected[1]
Exon I	84	10	6	60.0	120.7
Exon II	579	64	26	40.6	58.5
Exon III	258	18	9	50.0	82.4
Exon IV	150	6	3	50.0	82.4
Exon V	126	8	4	50.0	82.4
Total	1197	106	48	45.3	69.6

Only five synonymous codons namely, threonine, valine, proline, glycine and alanine were counted. Number of third position changes in these amino acids were used to determine the divergence between the two sequences shown in Figure 21.

[1] Values are corrected for multiple hits by the formula, $d = - 3/4 \ln (1 - 4/3f)$, where f is the uncorrected fraction of divergence.

Table 6:

Percent Divergence In the Replacement Codons:

Region	Length In bp.	Effective No. of Replacement Codons[1]	No. of Altered Replacement Codons	% Divergence Uncorrected	Corrected[2]
Exon I	84	18	0	0	0
Exon II	579	129	10	7.8	8.2
Exon III	258	68	6	8.8	9.4
Exon IV	150	44	9	20.5	23.9
Exon V	126	34	7	20.6	24.9
Total	1197	293	32	10.9	11.8

[1] Effective number of replacement codons was determined by subtracting the synonymous codons from the total number of amino acid codons.

[2] Values are corrected for multiple hits by the formula, $d = -3/4 \ln(1 - 4/3f)$, where f is the uncorrected fraction of divergence.

Table 7:

Intron Boundaries In C. elegans and C. briggsae:

Region	Species	5'	3'
Intron I:	*C. elegans*	AGC/gtgagt------------------	atttcag/AAT
	C. briggsae	CGC/gtaagt------------------	atttcag/AAT
Intron II:	*C. elegans*	GAA/gtaagt------------------	atttcag/TTT
	C. briggsae	AAA/gtaagt-----------------	atttctag/TTT
Intron III:	*C. elegans*	AGG/gtttgt-----------------	atttctag/TTT
	C. briggsae	AAG/gttagt------------------	atttcag/TTC
Intron IV:	*C. elegans*	GTG/gttagt------------------	N/D
	C. briggsae	GTG/gtaagt------------------	N/D

The codon usage for *lin-12* (Greenwald, *et al.*, 1986), *vit-5* (Spieth, *et al.*, 1985), CeC2 and CbC2 genes were observed to be variable for the different genes of same species. There is no considerable variation in codon usage for C2 genes of *C. elegans* and *C. briggsae*. These results are tabulated in Table 8.

Table 8:

Codon	amino acid	lin-12	CeC2	CbC2	vit-5
UUU	Phe	8	7	5	11
UUC	Phe	9	16	15	59
UUA	Leu	1	3	1	1
UUG	Leu	4	7	10	18
CUU	Leu	4	11	7	76
CUC	Leu	2	8	1	44
CUA	Leu	4	1	6	0
CUG	Leu	3	3	3	2
AUU	Ile	11	25	14	40
AUC	Ile	4	9	12	51
AUA	Ile	2	3	3	1
AUG	Met	6	18	15	35
GUU	Val	4	12	11	68
GUC	Val	3	6	1	48
GUA	Val	2	4	4	1
GUG	Val	3	3	5	2
UCU	Ser	6	8	5	51
UCC	Ser	0	4	4	38
UCA	Ser	10	9	5	13
UCG	Ser	6	3	4	3
AGU	Ser	4	5	4	4
AGC	Ser	2	1	0	9
CCC	Pro	2	1	1	0
CCA	Pro	13	9	8	61
CCG	Pro	1	2	1	2
ACU	Thr	4	6	7	47
ACC	Thr	3	0	3	49
ACA	Thr	3	10	9	5
ACG	Thr	1	6	3	1

Codon	amino acid	lin-12	CeC2	CbC2	vit-5
GCU	Ala	2	10	6	46
GCC	Ala	3	8	6	28
GCA	Ala	5	9	11	5
GCG	Ala	0	1	0	1
UAU	Tyr	6	7	5	8
UAC	Tyr	3	10	9	45
CAU	His	6	3	5	8
CAC	His	3	3	0	25
CAA	Gln	9	14	11	90
CAG	Gln	4	3	2	10
AAU	Asn	17	12	5	17
AAC	Asn	7	4	5	56
AAA	Lys	14	15	14	16
AAG	Lys	7	13	10	108
GAU	Asp	23	17	14	27
GAC	Asp	1	3	3	31
GAA	Glu	17	15	13	66
GAG	Glu	8	4	6	106
UGU	Cys	43	3	2	1
UGC	Cys	10	1	3	19
UGG	Trp	1	1	1	14
CGU	Arg	1	4	2	45
CGC	Arg	0	0	1	34
CGA	Arg	5	6	2	3
CGG	Arg	1	0	0	1
AGA	Arg	3	11	7	14
AGG	Arg	0	1	1	1
GGU	Gly	6	5	5	2
GGC	Gly	1	0	0	0
GGA	Gly	31	10	7	41
GGG	Gly	4	2	1	1

D. Discussion:

Correlation of the Physical and the Genetic Maps:

One prerequisite for studying genome organization is the alignment of the genetic and molecular maps. This requires placing cloned fragments of DNA within narrow genetically defined intervals. At the genetic level the unc-22(IV) region has been under extensive study for several years. It has been saturated to about 65% for mutations in essential genes and these genes have been genetically characterized (Clark et al., 1988; Rogalski and Baillie, 1985; and Rogalski et al., 1982). Several alleles have been obtained for many of the genes that are uncovered by sDf2 deficiency and a molecular study should ultimately permit the functional analyses of these alleles at the sequence level.

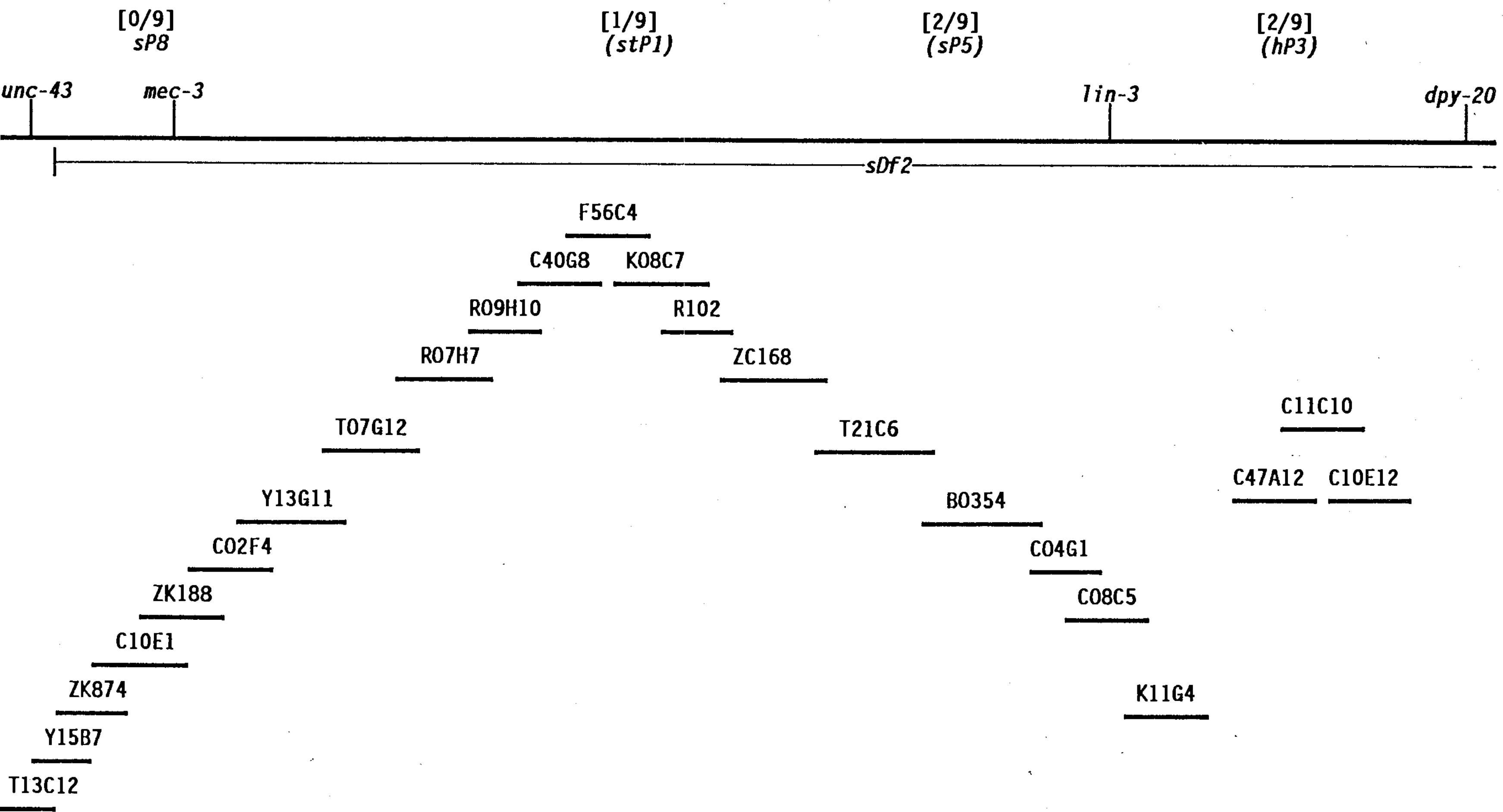
The subjects of this thesis were four cosmids in the unc-22 (IV) region: C11F2 and C18D3 in the unc-22 contig and C02C4 (carrying the sP4 polymorphism) and C08D7 constituting the sP4 contig. Prior to this study, the position of the sP4 contig relative to unc-22 was not known. Genetically the region around unc-22 is being divided into zones delineated by deficiency breakpoints (Clark et al., 1988; Rogalski and Baillie, 1985; and Rogalski et al., 1982). In the first section of this thesis I have shown that by mapping sP4 relative to these breakpoints I was able to position C02C4 between the left hand breakpoints of sDf7 and sDf9 (Figure 3, Section I). The position of the C02C4 was, therefore, established to be to the left of unc-22 and within the

recombinationally small (approximately 0.2 m.u.) interval between dpy-20 and unc-22. This result has recently been confirmed by the isolation of a yeast artificial chromosome (Y13D2) (R. Waterston, personal communications). The yeast chromosome has also established the orientation of the cosmids within the sP4 contig relative to the genetic map (i.e. left - right positions) and the contig has now been linked to the unc-22 contig by other overlapping cosmids as shown in Figure 3.

Based on data obtained from A. Coulson and J. Sulston, Figure 3 also shows the minimum number of cosmid lengths that span the sP4-unc-22 contigs. Assuming an average insert size of 35 kb for each cosmid, I estimate that the contig covers about 300 kb of the C. elegans genome. Similarily, Figure 21 shows the minimum number of cosmids within the unc-43 - sP4 interval. Here the communal genome mapping project has ordered and linked three other known polymorphic regions (i.e., sP8, stP1 and sP5) into a contig of approximately 450 kb that is anchored to the genetic map by mec-3. Recently, by means of another polymorphism, namely hP3, a contig of at least 70 kb has been identified between the mec-3 contig on the left and the sP4 - unc-22 contig on the right (Harris, 1988). Together, the three contigs between unc-43 and unc-22 span at least 820 kb. Based on the recombination map this interval represents approximately 1.0 m.u. (D.V. Clark, personal communication). Therefore at least 820 kb of C. elegans DNA are in this region, defined by 1.0 m.u. This amount of DNA already represents more

Figure 21:

The region between unc-43 and dpy20 genes on the linkage group IV. Top line shows the genetic map defined by deficiencies below it. In the square brackets [] are the fraction of Southern blotted DNAs showing the BO pattern when probed with the indicated polymorphic fragment (Baillie et al., 1985). The DNAs were prepared from unc-43;+ recombinants derived from unc-43(N2)unc-22(N2)/+(BO)+(BO) hermaphrodites.



[0/9]
sP8
[1/9]
(stP1)
[2/9]
(sP5)
[2/9]
(hP3)
unc-43
mec-3
lin-3
dpy-20
sDf2
F56C4
C40G8
K08C7
R09H10
R102
R07H7
ZC168
T07G12
T21C6
C11C10
C47A12
C10E12
Y13G11
B0354
C02F4
C04G1
ZK188
C08C5
C10E1
ZK874
K11G4
Y15B7
T13C12

than 1% of the entire C. elegans molecular map, and occupies only 0.3% (1/300) of the recombinational map.

As explained in the Introduction the calculated average of DNA/m.u. in the C. elegans genome is 267 kb/m.u. Therefore, a higher value, as in the region of unc-22 (>820 kb/m.u.), would support the proposal that the apparent clustering of genes in the C. elegans genetic map is, at least in part, due to decreased recombination frequencies in clusters instead of a non-random distribution of genes. In the regions on the autosomal arms and on the X- chromosome, which do not exhibit marked clustering of genes, DNA/m.u. will be expected to be closer to or less than the average value of 267 kb/m.u. These values should be taken into consideration when attempts are made to clone C. elegans genes by chromosomal "walking" strategies.

Identification Of Coding Regions In Cosmids:

The cosmids as described above provide an elegant basis for studying several genetically characterized genes molecularly. These correlations can be used to examine the complexity of gene structure and regulation in detail. The second part of section I of this thesis involved the search for potential coding regions in the four cosmids being studied.

Previously cloning of C. elegans genes has been carried out successfully either 1) by screening for specific genes using homologous probes from other organisms (e.g.,hsp-1 genes,

Snutch et al., 1988), 2) by transposon tagging genetically identified genes [examples are unc-22 (IV) (Moerman et al.,1986) and lin-12 (III) (Greenwald et al., 1986)], or 3) by identifying injected DNA clones that rescue mutant phenotypes (Fire, 1986). These methods require previous identification of a gene. Furthermore, not all genes may be amenable to the production of transposon tagged alleles. The methods were, therefore, not suitable for identifying all coding sequences in the four cosmids. Required was a method that detected coding sequences *per se*. Such a method could be Northern blot hybridizations and/or screening of cDNA libraries. In a large region such as that of unc-22 these methods are both time consuming and rely on the representation of all transcripts. Additionally, only a small fraction of the eucaryotic genome is transcribed. Therefore, a major purpose of this thesis was to develop an alternate efficient method to identify genes in the unc-22 (IV) region. The method described here relies on the divergence of genomes between *C. elegans* and *C. briggsae*. Based on the study with hsp-1 genes from these two species, (explained in the Introduction), it was believed that mainly coding regions would be conserved. Most of the non-coding DNA sequences in these two species should be diverged to a greater degree than coding sequences and therefore hybridization of fragments from these two genomes would allow a rapid means of detecting coding regions. With the consideration that other evolutionary conserved regions would also be detected, I have utilized this method in an attempt to identify coding regions

in the cosmids positioned to the left of unc-22 gene.

With the hybridization of fragments from the cosmids around the unc-22 gene to C. briggsae genomic DNA, nine regions of homology were detected (Table 1). Impressively, none of C. elegans fragments that hybridized to C. briggsae genome at moderately high stringency were repetitive elements. On the basis of these hybridizations it would appear that specific regions of genomes have been conserved in evolution and these regions have been selectively constrained for neutral changes due to their important functions. Furthermore, Heine et al. (1986) in a converse experiment showed that only coding regions of cloned genes hybridize to the labelled C. briggsae genome. Therefore the nine regions of the cosmids that hybridize to the C. briggsae genome are putative coding regions. Hence, C. briggsae homologies to C. elegans probes can be used to identify the potential gene sequences when a large region of the genome such as unc-22 is being analysed. In addition, this technique aids in determining whether linkage in a set of genes is conserved evolutionarily. Such conservation may be indicative of coregulated gene functions.

Isolation Of cDNA Clones Representing The unc-22 Region:

The essential genes around unc-22 IV have been studied quite extensively by genetic analysis (Clark et al., 1988; Rogalski and Baillie, 1985; Rogalski et al., 1982). Twenty three essential genes and at least three visible genes (dpy-20,

him-8, and mec-3) to the left of unc-22 have been identified (see Edgley and Riddle, 1987 for information on visible genes). In order to facilitate a molecular analysis in the region which is between dpy-20 and unc-22, I have isolated four cDNA clones with the C. briggsae homologous fragments of C. elegans cosmids in this region. On screening a cDNA library constructed with the transcripts from a population of mixed stages of development, only two out of nine homologous regions were found to be expressed or represented in this specific cDNA library. The examination of the other seven C. briggsae homologous fragments revealed that at least five other regions hybridize to transcript(s), when total RNA from various stages of C. elegans development were analysed. These observations suggest that the inability to isolate other cDNA clones from the above-mentioned cDNA library was due to the lack of representation of individual cDNA clones in the library instead of an absence of expression. Two other regions of C. briggsae homologies did not hybridize to any transcript at any stage of development. I believe these two regions are either evolutionarily conserved regulatory sequences or are expressed in too low an abundance during C.elegans development to detect with Northern blots.

The results presented here support the conclusion that most genomic sequences of C. elegans with homology to C. briggsae genome are actually coding sequences. Therefore, C. briggsae homologies to C. elegans probes can be initially used to identify potential gene sequences and later can be confirmed

by Northern blots and screening of cDNA libraries. I would like to emphasise that not all hybridization signals were clearly detected on RNA blots of mixed stages of development but on stage specific RNA blots these signals were distinctly detected. Thus, with merely a Northern blot of mixed population RNA, such signals would be missed. The possibility of isolating other cDNA clones from a cDNA library constructed from mRNA isolated from the L2 stage of development was tested with a 3.3 kb *Pst*I fragment ("t") from C18D3. As expected from the stage specific Northern blots, a cDNA clone (C6) for this specific gene was isolated from this library. Thus, I believe that the other *C. briggsae* homologous regions, for which no cDNA clone was isolated from the mixed stages cDNA library, may be used as probes with a stage specific cDNA library to isolate other cDNA clones.

Characterization Of The Identified Coding Regions:

The analyses in section II of this thesis have facilitated the determination of the relative expression of seven coding regions in various stages of development in *C. elegans*. Interestingly, six of the eight coding regions show an increase in the abundance of transcript at L2 (24 hrs.) stage and then either gradually or sharply decrease at the later stages of development. Similar to these observations is the fact that all mutations isolated so far in this region arrest development of *C. elegans* in early to mid-larval stages (Clark et al.,

1988; Rogalski and Baillie, 1985). Therefore, the products of the genes identified by these mutations are essential either at or before the mid larval stages and the observation of abundance of trascripts at those stages seems to correlate well with this. Furthermore the developmental profile of expression of six of the eight genes in this region is much more similar to that of the muscle actin than those of the hsp-1 or Na-channel genes. Since the unc-22 gene has been earlier shown to code for a component of muscle thick filament (Moerman et al., 1988), it is reasonable to suggest that the genes clustered in this region may have muscle related functions. A further analysis of the conserved regulatory sequences of these genes should address the question if the clustered genes are due to the co-ordinated gene expression of functionally similar and/or tissue specific genes.

The isolated C2 cDNA represents a single copy gene with at least one EcoRI site. The 3′ ends of both C4 and C5 cDNAs lie close to each other in a 3.2 kb EcoRI/HindIII fragment and these genes are transcribed in the opposite orientation. C5 cDNA is encoded by a member of a gene family, another member of which is closely linked in the genome near unc-22, represented by cosmid C11F2. The expression of the C4 gene is much lower at L1 and L3 stages of development than that of C5.

The C2 Gene Lies In Zone 5 Of The Genetic Map:

Since alleles genetically defining several essential genes

in this region have been isolated and characterized, it would be interesting to correlate those genes with the molecularly identified genes. The first problem in making these correlations is the fact that this region appears to be recombinationally suppressed and therefore the amount of DNA in a map unit is above average (at least 300 kb in 0.2 m.u. between dpy-20 and unc-22). The second problem is the possibility of mutationally silent genes that would not be identified through EMS mutagenesis because the function of gene product may not be essential for development or may be compensated for by another gene. For example, in the case of C5, it may be difficult to obtain a mutant phenotype if another member of its gene family is wild type and fully functional. Thus it is desirable to localize the molecularly identified genes to as small a genetically defined region as possible. Fortunately, the identified essential genes in the unc-22 region (left and right) have been distributed into ten zones on the basis of breakpoints of nine overlapping deficiencies. I have attempted to position C2 gene using the deficiency strain DNA. Through this approach, C2 was demonstrated to be deleted by deficiencies sDf2, sDf7, sDf8, sDf9, sDf10, and mDf7. Therefore, it is possible to conclude that this gene is present in zone 5 in the genetic map. So far, the only genetically identified gene other than unc-22 in this zone is let-56. Therefore let-56 is a candidate for the C2 gene. The other possibility that the C2 gene may encode for a non-essential gene product cannot be ruled out because the genetic analysis

for lethal mutations would not identify such genes. In the latter situation it is possible that other coding regions which may lie near the C2 gene in zone 5 may encode the gene product of let-56. Recently seven more small deficiencies in the unc-22 region have been obtained (D.V. Clark, personal communication). A similar construction of strains with their genotype + Dfx +/let- (m435) nT1(IV);+/nT1(V) may be used to isolate DNA and position molecularly identified genes in individual zones. The transformation system developed by Fire (1986) should be able to resolve the question of assigning a particular genetically identified gene to a molecularly identified gene. This technique would involve rescuing a mutant phenotype with the injection of wild type DNA from the corresponding region into oocytes.

Sequence Analysis:

In the third section of this thesis the DNA sequences of two cDNAs were obtained and the EMBL DNA data bank was searched for DNA identities. No detectable homology was found. The sequences of both cDNAs (C2 and C5) showed a remarkably high number of residues with hydrophobic amino acids suggesting the possibility of their association with the membrane spanning domains. On the basis of Northern blot hybridization, it had been evident that both cDNAs were of incomplete lengths. The sequences now showed that the cDNAs contained only the 3′ end of the transcript by the presence of a consensus poly (A)

addition site (AATAAA).

Since these cDNAs were isolated by hybridizations of the subcloned genomic fragments, a 3.2 kb EcoRI genomic fragment (derived from FH6) hybridizing to C2 was sequenced. The observation that the sequences of C2 matched perfectly to the coding sequences of FH6 derivative confirmed its location near the unc-22 gene. Furthermore, the sequencing of genomic fragment deduced the direction of transcription for the C2 gene with respect to that of unc-22. The C2 gene was found to be transcribed in an opposite orientation to that of the unc-22 gene. The unc-22 gene has earlier been shown to be transcribed from right to left on the genetic map (Moerman et al., 1986). In C2 four introns of sizes ranging from exceptionally small (48 nucleotides) to exceptionally large (253 nucleotides) with normal consensus splice junctions were identified.

Similarly the C5 identity was confirmed by sequencing from both ends of the 3.2 kb EcoRI/HindIII genomic fragment. In addition, the sequence showed that C5 was transcribed from left to right (with respect to the genetic map) and that C4 was transcribed in the opposite direction. Consequently the 3′ ends of the two genes are positioned very close to each other.

The *C. briggsae* regions with homology to the C2 and C5 cDNAs of *C. elegans* were isolated and characterized. Based on hybridization and sequencing data I showed that, for C2, homology between the two species extends from the 3′ end toward the 5′ end for about 8.3 kb (Figure 19).

The sequences of the C2 genes from *C. elegans* and *C.*

briggsae were compared to determine the divergence of sequences because similarily diverged genomes between other organisms may be used to detect the coding regions. It is clear that silent substitutions in exons are more frequent than ones resulting in amino acid replacements between C2 genes of C. elegans and C. briggsae. The corrected change for synonymous codons is 70% while for replaced amino acids it is 11.8%. Strong purifying selection seems to have constrained replacement substitutions from the time of the evolution of these two species. Some sections of the C2 protein appear to be more selectively constrained than others since amino acid replacements are distributed unevenly across the four exons. More changes are observed in the fourth and fifth exons than would be expected on the average whereas no changes are observed in the first exon suggesting that purifying selection is limiting replacement substitutions in the functionally important domains. Another possibility for these observations may be due to a smaller sample size in the first exon. The high level of sequence identity suggests that the C2 gene may have similar functions in the two species.

The estimation of the time of divergence between C. elegans and C. briggsae is difficult due to the lack of any fossil records. On the basis of their present day morphological differences and geographical distribution, it is believed that these two species were probably separated from a common ancestor and became reproductively isolated about 20-40 million years ago (G. Poinar, personal communication). Since

this estimate is only based on detectable morphological differences, it may under estimate the divergence time for these closely related species. Recently it has been shown that a silent substitution rate of 1%/Myr is consistent for bacterial protein coding genes and nuclear genes of mammals, invertebrates and flowering plants (Ochman and Wilson, 1987). Hence a rough estimate of the time of divergence since a last shared common ancestor can be obtained from the corrected nucleotide divergence. Based on corrected nucleotide divergence from the silent changes in the C2 genes for _C. elegans_ and _C. briggsae_ of 70%, their time of divergence since a common ancestor can be estimated to be at least 35 Myr. It must be noted that this is a minimum estimate, since any back substituitions in the synonymous codons can not be accounted for by this procedure. Furthermore, any functional constraint on a particular preferred structure of the C2 transcript would limit the amount of nucleotide divergence. However, any similarly diverged species should provide an analogous means for identifying coding sequences among other organisms when large regions, such as the _unc-22_ region, are being analysed. I believe hybridization of cloned DNA from _D. melanogaster_ to the genome of _D. pseudoobscura_ may be used to identify most coding sequences of these two species. This technique will be especially applicable in the large regions, where the genes are developmentally regulated because such genes are expressed in abundance only at specific times in development.

An ultimate goal of studying genome organization is to

determine whether there is any functional significance in the relative gene positions. Support for such significance could be in the form of detecting gene linkages that are conserved between _C_. _elegans_ and _C_. _briggsae_. Selective maintenance of such organization in the evolution of two diverged species must play an important role in their function. The work carried out in this thesis constitutes the first step towards assessing the degree of evolutionary linkage stability in the _unc-22_ region.

E. Summary and Conclusions

The objective of this thesis has been to establish the application of interspecies hybridization as a tool to identify various coding regions rapidly and efficiently. I undertook this task with the cosmids that were positioned to the left of unc-22 in *Caenorhabditis elegans* because this region has been intensively studied through genetic means.

Two cosmids each from two sets of cosmid islands (four cosmids), namely unc-22 and sP4 were examined in this study. Earlier Baillie et al. (1985) had shown that the sP4 polymorphic fragment, which was later used to isolate several overlapping cosmids, mapped on the genetic map extremely close to the unc-22 gene. Since neither end of the sP4 cosmid contig showed any overlap with the unc-22 contig, its left-right position with respect to unc-22 was not apparent. In order to map the sP4 region more precisely, I have utilized deficiency strain DNA around unc-22 and have demonstrated that this contig is positioned immediately left of the unc-22 contig. Recently the orientation of this contig has been established by the isolation of yeast artificial chromosome (R. Waterston, personal communication) which also confirms its position as observed with the deficiency DNAs.

Through these analyses and the estimation from the number of existing overlapping cosmids (including sP8, stP1, sP5 and hP3), it is evident that the region between unc-43 and unc-22 contains greater than 820 kb of *C. elegans* DNA. This region has been defined by the 0.9 m.u. left half of sDf2 deficiency. Thus, it can be concluded that this region is apparently

recombinationally suppressed on the basis of expected DNA content (on average 267 kb/m.u) for the C. elegans genome.

The region between unc-22 and dpy-20 has been shown by genetic analysis to be gene dense and lethal mutations in this region block development of this organism at early to mid larval stages. Since this region is excessively large, the use of interspecies hybridization for identifying those regions on the cosmids that are coding is an ideal one. The results shown in this thesis indicate that the hybridization of C. elegans probes to C. briggsae genomic DNA is indeed a rapid and quite sensitive method for detecting coding elements. Through this technique, nine coding elements were detected and later characterized. It should be noted that not all of these homologous regions detected a transcript when total RNA from mixed stages of nematodes were screened. Since initial indication of these coding elements were derived by their homology to C. briggsae genome, four stages of nematode RNA were used for Northern blots. Seven of the nine regions with homology to C. briggsae genome showed homology to a transcript which is expressed in high abundance at L2 stage of C. elegans development. Furthermore, the expression profile of each region resembled the expression profile observed with muscle actin probe. These observations are particularly interesting because the closely linked unc-22 gene has been shown to code for a muscle component in C. elegans (Moerman, et al., 1988). Therefore, the speculation of the identified coding regions being expressed in muscle specific cells must be considered.

An initial test would be to sequence and compare the 5' regions of each gene in this region.

Two of the nine homologous regions identified and purified corresponding cDNA clones when a mixed population cDNA library was individually screened. I believe that this inability to obtain any cDNA clones for other seven regions of C. briggsae homology is due to the low representation of corresponding cDNAs in the mixed population cDNA library. I have tested this hypothesis by screening a 24 hour (L2) stage cDNA library with one such region. As expected a corresponding cDNA clone was isolated from this library, thus suggesting that the other six regions can be used similarly to isolate corresponding cDNA clones from a stage specific cDNA library where it is shown to be expressed in abundance.

Two isolated cDNAs were characterized in detail. The C2 gene is clearly present as a single copy in both C. elegans and C. briggsae genomes. The C5 gene appears to represent a member of closely related gene family; one of its member is closely linked to unc-22 in the C. elegans genome. Since C5 showed three different transcripts, two of which being in extremely low abundance, it would be interesting to determine whether the other clustered region corresponds to any of those transcripts.

Both of the isolated cDNAs were entirely sequenced to determine the amino acid sequence of protein encoded by them. No detectable homology to the existing DNA sequence data bank was found. The C2 gene was hybridized to the deficiency strain DNAs to position it to a single zone, so that this gene may be

correlated to any identified mutant phenotype. It was shown that the C2 gene lies in the zone 5 and is a likely candidate for let-56, since the only genetically identified essential gene in zone 5 is let-56. The possibility of the C2 gene corresponding to a nonessential gene has not been ruled out.

Each gene identified in this study showed a developmental expression profile very similar to the actin gene and very dissimilar to hsp-1 and Na^+ channel. This observation is particularly interesting in light of the possibility that the organization of these genes may reflect some functional similarity. This possibility can be directly tested by studying similar genomic regions from C. briggsae and determining whether a specific gene order is maintained in evolution. I have begun such analysis by first estimating the sequence divergence using C2 genes between C. elegans and C. briggsae. The sequence analysis showed that these two species have diverged from a last shared common ancestor at least 35 million years ago. The nucleotide sequences in the coding regions are 80% conserved while the introns are highly diverged. The amino acids are about 91% conserved and the replacement changes in amino acids are mainly of conservative nature.

F. <u>PROPOSALS FOR FURTHER RESEARCH</u>:

The experiments carried out during the course of this thesis provide the foundation for a detailed molecular study in the *unc-22(IV)* region of *C. elegans*. The following experiments are suggested as a means for furthering some of these analyses.

1) Isolate cDNA clones for the other five *C. briggsae* homologous regions from one of the stage specific cDNA libraries where the corresponding cDNA is expected to be in abundance on the basis of Northern blots.

2) Sequence the 5' non-coding regions of each of the identified genes to determine if there are any concensus promoter regions for their expression.

3) Screen the charon 4 *C. briggsae* genomic library using the entire cosmids as probes to determine whether any gene order is maintained between *C. elegans* and *C. briggsae*.

4) Sequence and compare 5' non-coding regions for each pair of homologous genes between *C. elegans* and *C. briggsae*.

5) With the use of *insitu* hybridization, determine if the identified genes in this region are expressed in a tissue specific manner.

6) Position each of the identified coding sequences in the genetically defined zones by the use of deficiency strain DNAs.

7) The genes molecularly positioned in zones can then be individually used to rescue the corresponding lethal phenotypes in the same zone by transformation.

8) Once the molecularly identified genes have been correlated to the lethal mutations, the analogous genes from <u>C</u>. <u>briggsae</u> genome can be used in transformation studies to address a variety of interesting questions.

9) Use <u>in</u> <u>vitro</u> site specific mutagenesis to study structure and function related questions for individual genes in the <u>unc-22</u> region.

G. REFEERENCES:

Baillie, D. L., K.A. Beckenbach, and A.M. Rose. 1985. Cloning within the unc-43 to unc-31 interval (linkage group IV) of the Caenorhabditis elegans genome using Tc-1 linkage selection. Can. Jnl. of Gen. and Cytol. 27(4):457-466.

Beermann, W. 1956. Nuclear differentiation and functional morphology of chromosomes. Cold Spring Harbor Symp. Quant. Biol. 21:217-232.

Bender, W., M. Akam, F. Karch, P.A. Beachy, M. Peifer, P. Spierer, E.B. Lewis and D.S. Hogness. 1983a. Molecular genetics of the bithorax complex in Drosophila melanogaster. Science 221:23-29.

Bender, W., P. Spierer and D.S. Hogness. 1983b. Chromosomal walking and jumping to isolate DNA from the Ace and Rosy loci and the bithorax complex in Drosophila melanogaster. J. Mol. Biol. 168:17-33.

Benton, W. D. and R.W. Davis. 1977. Screening gt10 recombinant clones by hybridization to single plaques in-situ. Science 196:180-182

Blattner, F.R., B.G. Williams, A.E. Blechl, K. Denniston-Thompson, L.A. Furlong, D.J. Grunwald, D.O. Kiefer, D.O. Moore, J.W. Schumm, E.L. Sheldon and O. Smithes. 1977. Charon phages: safer derivatives of bacteriophage lambda for DNA cloning. Science 196:161-169.

Blumenthal, T., M. Squire, S. Kirtland, J. Cane, M. Donegan, J. Spieth, and W. Sherrock. 1984. Cloning of a yolk protein gene family from Caenorhabditis elegans. J. Mol. Biol. 174:1-18.

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

Burke, D. J. and S. Ward. 1983. Identification of a large multigene family encoding the major sperm protein of Caenorhabditis elegans. J. Mol. Biol. 171:1-29.

Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochem. 18:5294-5299.

Clark, D.V., T.M. Rogalski, L.M. Donati, and D.L. Baillie. 1988. The unc-22(IV) region of Caenorhabditis elegans: Genetic analysis of lethal mutations. Genetics, In Press.

Clark, S.H., A.J. Hilliker, A. Chovnik. 1986. Genetic analysis of the right 3'end of the Rosy locus in Drosophila melanogaster. Genet. Res. Camb. 47:109-116.

Cote, S., A. Preiss, J. Haller, R. Schuh, A. Kienlin, E. Seifert and H. Jackle. 1987. The gooseberry-zipper region of Drosophila: five genes encode different spatially restricted transcripts in the embryo. EMBO Journal 6(9):2793-2801.

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

Crowley, T.E., M.W. Bond and E.M. Meyerowitz. 1983. The structural genes for three Drosophila glue proteins reside at a single polytene chromosome puff locus. Mol. Cell. Biol. 3:623-634.

Davis, R. W., D. Botstein, and J.R. Roth. 1980. A manual for genetic engineering: advanced bacterial genetics. Cold Spring Harbor Laboratory, New York.

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. USA. 76:1333-1337.

Feinberg, A. and B. Vogelstein. 1983. Atechnique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal.Biochem. 132:6-13.

Ferguson, E.L. and H.R. Horvitz. 1985. Identification and characterization of 22 genes that affect the vulval cell lineages of the nematode Caenorhabditis elegans. Genetics 110:17-72.

Files, J. G., S. Carr, and D. Hirsh. 1983. Actin gene family of Caenorhabditis elegans. J. Mol. Biol. 164:355-375

Fire, A. 1986. Integrative transfomation of Caenorhabditis elegans. EMBO Journal 5:2673-2680.

Garrells, J.I. 1979. Changes in protein synthesis during myogenesis in a clonal cell line. Dev. Biol. 73:134-152.

Greenwald, I., A. Coulson, J. Sulston and J. Priess. 1987. Correlation of the physical maps in the lin-12 region of Caenorhabditis elegans. Nucl. Acids Res. 15(5):2295-2307.

Greenwald, I. 1985. lin-12, a nematode homeotic gene, is homologous to a set of mammalian proteins that includes epidermal growth factor. Cell 43:583-590.

Gurdon, J.B. 1962. The developmental capacity of nuclei taken from intestine epithelial cells of feeding tadpoles. J. Embryol. Exp. Morphol. 10:622-640.

Harris, L.J. 1988. Transposable elements in Caenorhabditis elegans Var. Bristol and Caenorhabditis briggsae. Ph.D. Thesis, University of British Columbia, Vancouver, British Columbia, Canada.

Hattori, M. and Y. Sakaki. 1986. Dideoxy sequencing method using denatured plasmid templates. Anal. Biochem. 152:232-238.

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

Henikoff, S. and M.K. Eghtedarzadeh. 1987. Conserved arrangement of nested genes at the Drosophila Gart locus. Genetics 117:711-725.

Henikoff, S. 1987. Unidirectional digestion with Exonuclease III in DNA sequence analysis. Methods Enzymol. 155:156-165.

Hochman, B. 1978. In: Ashburner, M. and E. Novitzki (Eds.). The genetics and biology of Drosophila. vol. 1b, Academic Press, London.

Hough-Evans, B.R., M. Jacobs-Lorena, M.R. Cummings, R.J. Britten and E.H. Davidson. 1980. Complexity of RNA in eggs of Drosophila melanogaster and Musca domestica. Genetics 95:81-94.

Howell, A.M., S.G. Gilmour, R.A. Mancebo and A.M. Rose. 1987. Genetic analysis of a large autosomal region in Caenorhabditis elegans by the use of a free duplication. Genet. Res. Camb. 49:207-213.

Ish-Horowicz, D. and J.F. Burke. 1981. Rapid and efficient cosmid cloning. Nucl. Acids Res. 9:1323-1338

Izquierdo, M., and J.O.Bishop. 1979. An analysis of cytoplasmic RNA populations in Drosophila melanogaster, Oregon R. Biochem. Genet. 17:473-497.

Papajian, D.M., T.L. Schwarz, B.L. Tempel, Y.N. Jan and L.Y. Jan. 1987. Cloning of genomic and complementary DNA from Shaker, a putative potassium channel gene from Drosophila. Science 237:749-753.

Jukes, T.H. and C.R. Cantor. 1969. Evolution of protein molecules. In: Munro H.N. (ed) Mammalian protein metabolism, vol 3. Academic Press, New York.

Karlik, C.C, J.W. Mahaffey, M.D. Coutre, and E.A. Fyrberg. 1984. Organization of contractile protein genes within the 88F subdivision of the D. melanogaster third chromosome. Cell 37:469-481.

Khandjian, E.W. 1986. UV cosslinking of RNA to nylon membranes enhances Hybridization signals. Molec. Biol. Rep. 11:107-115.

Kramer, J.M., G.N. Cox, and D. Hirsh. 1982. Comparisions of the complete sequences of two collagen genes from Caenorhabditis elegans. Cell 30:599-606.

Laskey, R.A. and A.D. Mills. 1977. Enhanced autoradiographic detection of ^{32}P and ^{125}I using intensifying screens and hypersensitive film. FEBS. Lett. 82:314-316.

Lefevre, G. and W. Watkins. 1986. The question of the total gene number in Drosophila melanogaster. Genetics 113:869-895.

Lewis, E.B. 1978. A gene complex controlling segmentation in Drosophila. Nature 276:565-570.

Lewis, J.A., C.H. Wu, J.H. Levine, and H. Berg. 1980. Levamisole-resistant mutants of the nematode Caenorhabditis elegans appear to lack pharmacological acetylcholine receptors. Neuroscience 5:967-989.

Lifton, R.P., M.L. Goldberg, R.W. Karp, D.S. Hogness. 1977. The organization of the histone genes in Drosophila melanogaster: Functions and evolutionary implications. Cold Spring Harbor Symp. Quant. Biol. 42:1047-1051.

Lindsay, S. and A.P. Bird. 1987. Use of restriction enzymes to detect potential gene sequences in mammalian DNA. Nature 327:336-338.

Long, E.O., and I.B. Dawid. 1980. Repeated genes in eukaryotes. Annu. Rev. Biochem. 49:727-764.

MacLeod, A.R., R.H. Waterston, R.M. Fishpool, and S. Brenner. 1977. Identification of the structural gene for a myosin heavy chain in Caenorhabditis elegans. J. Mol. Biol. 114:133-140.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, New York.

Maxam, A.M. and W. Gilbert. 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA. 74:560-565.

Meyerowitz, E.M. and D.S. Hogness. 1982. Molecular organization of a Drosophila puff site that responds to ecdysone. Cell 28:165-176.

Miller, J.H. and W.S. Reznikoff (Eds.). 1978. The operon. Cold Spring Harbor Laboratory, New York.

Moerman, D.G., G.M. Benian, R.J. Barstead, L.A. Schriefer, L.A. and R.H. Waterston. 1988. Identification and intracellular localization of the unc-22 gene product of Caenorhabditis elegans. Gene Dev. 2:93-105.

Moerman, D.G., G.M. Benian, and R.H. Waterston. 1986. Molecular cloning of the unc-22 gene in Caenorhabditis elegans by Tc-1 transposon tagging. Proc. Natl. Acad. Sci. USA. 83:2579-2583.

Moerman, D.G., S. Plurad, R.H., Waterston, and D.L. Baillie. 1982. Mutations in the unc-54 myosin heavy chain gene of Caenorhabditis elegans that alter contractility but not muscle structure. Cell 29:773-781.

Moerman, D.G. 1980. A genetic analysis of the unc-22 region in Caenorhabditis elegans. Ph.D. Thesis, Simon Fraser University, Burnaby, B. C., Canada.

Moerman, D.G. and D. L. Baillie. 1979. Genetic organization in Caenorhabditis elegans: fine structure analysis of the unc-22 gene. Genetics 91:95-103.

Morgan, T.H. 1934. Embryology and Genetics. Greenwood Press, Westport, Conn.

Murray, N.E. and K. Murray. 1974. Manipulations of restriction targets in phage lambda to form receptor chromosomes for DNA fragments. Nature 251:476.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Ochman, H. and A.C. Wilson. 1987. Evolution in bacteria: Evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. 26:74-86.

Proudfoot, N.J., M.H.M. Shander, J.L. Manley, M.L. Gefter and T. Maniatis. 1980. Structure and in-vitro transcription of human globin genes. Science 209: 1329-1336.

Reymond, C. D. 1987. A rapid method for the preparation of multiple samples of eukaryotic DNA. Nucl. Acids Res. 15(19):8118.

Riddle, D.L. 1978. The genetics of development and behavior in Caenorhabditis elegans. J. Nematol. 10:1-16.

Rigby, P.W., M. Dieckmann, C. Rhodes, and P. Berg. 1977. Labelling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase. J. Mol. Biol. 113:237-251.

Rogalski, T.M. and D.L. Riddle. 1988. A Caenorhabditis elegans RNA polymerase II gene, ama-1 IV, and nearby essential genes. Genetics 118:61-74.

Rogalski, T.M. and D.L. Baillie. 1985. Genetic organization of the unc-22 IV gene and the adjacent region in Caenorhabditis elegans. Mol. Gen. Genet. 201:409-414.

Rogalski, T.M. 1983. A genetic characterization of the region around the unc-22 gene in Caenorhabditis elegans. Ph.D. Thesis. Simon Fraser University, Burnaby, B. C., Canada.

Rogalski, T.M., D.G. Moerman, and D.L. Baillie. 1982. Essential genes and deficiencies in the unc-22 IV region of Caenorhabditis elegans. Genetics 102:725-736.

Rose, A.M., D.L. Baillie, E.P.M. Candido, K.A. Beckenbach and D. Nelson. 1982. The linkage mapping of cloned restriction fragment length differences in Caenorhabditis elegans. Mol. Gen. Genet. 188:286-291.

Rose, A.M. and D.L. Baillie. 1980. Genetic organization of the region around unc-15 I, a gene affecting paramyosin in Caenorhabditis elegans. Genetics 92:409-418.

Rose, A.M. and D.L. Baillie. 1979. The effect of temperature and parental age on recombination and nondisjunction in Caenorhabditis elegans. Genetics 92:21-30.

Sanger, F., S. Nicklen and A.R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA. 74:5463-5468

Schaffer, S.W. and C.F. Aquadro. 1987. Nucleotide sequence of the Adh gene region of Drosophila pseudoobscura: Evolutionary change and evidence for an ancient gene duplication. Genetics 117:61-73.

Scott, M.P. 1987. Complex loci of Drosophila. Ann. Rev. Biochem. 56:195-227.

Scott, M.P. and P.H. O'Farrell. 1986. Spatial programming of gene expression in early Drosophila embryogenesis. Ann. Rev. Cell Biol. 2:49-80.

Smith, G.E. and M.O. Summers. 1980. The bidirectional transfer of DNA and RNA to nitrocellulose or DBM paper. Anal. Biochem. 109:123-129.

Snutch, T.P., M.F.P. Heschl and D.L. Baillie. 1988. The Caenorhabditis elegans hsp70 gene family: A molecular genetic characterization. Gene, In Press.

Snutch, T.P. 1984. A molecular and genetic analysis of the heat shock response of Caenorhabditis elegans. Ph.D. Thesis. Simon Fraser University, Burnaby, B. C., Canada.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517.

Spieth, J., K. Denison, E. Zucker and T. Blumenthal. 1985. The nucleotide sequence of a nematode vitellogenin gene. Nucl. Acids. Res. 13(19):7129-7138.

Spradling, A.C. and G.M. Rubin. 1982. Science. 218:341-347.

Spradling, A.C. and G.M. Rubin. 1981. Drosophila genome organization: conserved and dynamic aspects. Ann. Rev. Genet. 15:218-261.

Spradling, A.C. 1981. The organization and amplification of two chromosomal domains containing Drosophila chorion genes. Cell 27:193-201.

Spradling, A.C., M.E. Digan, A.P. Mahowald, M. Scott and E.A.Craig. 1980. Two clusters of genes for major chorion proteins of Drosophila melanogaster. Cell 19:905-914.

Sulston, J.E. and H.R. Horvitz. 1977. Post-embryonic cell lineages of the nematode Caenorhabditis elegans. Develop. Biol. 56:110-156.

Sulston, J.E. and S. Brenner. 1974. The DNA of Caenorhabditis elegans. Genetics 77:95-104.

Swanson, M.M., M.L. Edgley and D.L. Riddle. 1984. The nematode C. elegans. In Genetic Maps. Vol. 3 Edited by S.J. O'Brien. Cold Spring Harbor, New york.

Thomas, P.S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA. 77:5201-5205.

Ward, S., D.J. Burke, J.E. Sulston, A.R. Coulson, D.G. Albertson, D. Ammons, M. Klass and E. Hogan. 1988. Genomic organization of major sperm protein genes and pseudogenes in the nematode Caenorhabditis elegans. J. Mol. Biol. 199:1-13.

Waterston, R.H., J.N. Thompson and S. Brenner. 1980. Mutants with altered muscle structure in Caenorhabditis elegans Develop. Biol. 77:271-302.

Waterston, R.H., R.M. Fishpool, and S. Brenner. 1977. Mutants affecting paramyosin in Caenorhabditis elegans. J. Mol. Biol. 117:679-698.

Wilson, A.C., H. Ochman and E.M. Prager. 1987. Molecular time scale for evolution. TIG 3(9):241-247.

Woodruff, R.C. and M. Ashburner. 1979. The genetics of a small autosomal region of Drosophila melanogaster containing the structural gene for alcohol dehydrogenase. II. Lethal mutations in the region. Genetics 92:133-149.

Wright, T.R.F. 1986. The genetic and molecular organization of the dense cluster of functionally related, vital genes in the DOPA decarboxylase region of the Drosophila melanogaster genome. In: Structure and function of eukaryotic chromosomes, Vol. 14 Edited by W. Hennig. Springer-Verlag Press.

Wright, T.R.F., W. Beermann, J.L. Marsh, C.P. Bishop, R. Steward, B.C. Black, A.D. Tonsett and E.Y. Wright. 1981. The genetics of DOPA decarboxylase in Drosophila melanogaster. IV. The genetics and cytology of the 37B10-37D1 region. Chromosoma 83:45-58.

Young, R.A., O. Hagenbuchle and U. Schibler. 1981. A single mouse alpha-amylase gene specifies two different tissue-specific mRNAs. Cell 23(2):451-458.

Young, M.W. and B.H. Judd. 1978. Nonessential sequences, genes and the polytene chromosome bands of Drosophila melanogaster. Genetics 88:723-742.

Zhimulev, I.F., G.V. Pokholkova, A.V. Bgatov, V.F. Semeshin, and E.S. Belyaeva. 1981. Fine cytogenetic analysis of the band 10A1-2 and the adjoining regions in the Drosophila melanogaster chromosome. II. genetic analysis. Chromosoma 82:25-40.