NOTICE

AVIS

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

If pages are missing, contact the university which granted the degree.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

THE MOLECULAR CHARACTERIZATION OF THE ESSENTIAL GENE *LET-653* IN *CAENORHABDITIS ELEGANS*

by

Steven John Mathias Jones

B.Sc., University of Bristol, 1990

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the Department

of

Biological Sciences

© Steven John Mathias Jones 1994

SIMON FRASER UNIVERSITY

February 1994

Canada

APPROVAL

Name: **STEVEN JOHN MATHIAS JONES**

Degree: **Master of Science**

Title of Thesis:

THE MOLECULAR CHARACTERISATION OF THE LET-653 GENE IN
CAENORHABDITIS ELEGANS.

Examining Committee:

Chair: Dr. L. Albright, Professor

Dr. D. L. Baillie, Professor, Senior Supervisor
Department of Biological Sciences, SFU

Dr. B. Honda, Associate Professor
Department of Biological Sciences, SFU

Dr. B. Brandhorst, Professor
Department of Biological Sciences, SFU

Dr.C. Boone, Assistant Professor,
Department of Biological Sciences, SFU
Public Examiner

Date Approved _Feb. 8, 1994_____

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

THE MOLECULAR CHARACTERIZATION OF THE

ESSENTIAL GENE LET-653 IN CAENORHABDITIS

ELEGANS

Author:

(signature)

STEVEN J.M. JONES

(name)

FEBRUARY 16 1994

(date)

ABSTRACT

A correlation of the genetic and physical maps in the *unc-22* region of chromosome IV of *Caenorhabditis elegans* has been done. This was accomplished by germline transformation using cloned DNA from the physical map. The results allowed the physical positioning of the genes *let-92*, *let-660* and *let-653*. These results have also provided evidence for a maternal contribution of the *let-92* gene product.

Germline transformation rescue was able to position the *let-653* gene to two overlapping plasmid subclones. The rescuing plasmid subclones were used to isolate a cDNA subclone. Sequence data generated from both cDNA and genomic DNA subclones showed that the *let-653* gene encodes a mucin-like protein. Studies of the *let-653* phenotype revealed that lethal arrest is concurrent with the appearance of a vacuole anterior to the lower pharyngeal bulb. The position of the vacuole is consistent with a dysfunction of the secretory/excretory apparatus.

This thesis provides evidence that the Let-653 mucin-like protein is essential within *Caenorhabditis elegans* and the absence of functional Let-653 protein effects the function of the excretory/secretory apparatus.

"They are in you and in me; they created us,

body & mind; and their preservation is the

ultimate rationale for our existence."

    - Richard Dawkins (The Selfish Gene).

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES cont'd

INTRODUCTION

The information required for the development of a complex eukaryote is contained within its genome. This information is stored as the sequence of four nucleotide bases. Genome analysis attempts to elucidate the nature, organization and function of this information.

A molecular approach to genome analysis ultimately involves the determination of the nucleotide sequence of the genome. The sequence data can then be used to predict various genomic elements. Such genomic elements would include exons, introns, promoters, enhancers, structural RNA genes and repeated sequences. Several genome sequencing projects are attempting to achieve this, most notably that of *Caenorhabditis elegans* (Sulston *et al.* 1992) and yeast (Oliver *et al.* 1992; Ouellette *et al.* 1993). However analysis of a genome based purely on sequence data suffers two major drawbacks. Firstly, such sequencing projects are unable to correlate a biological function with the sequence they produce. They are merely able to infer the function based on similarity to genes where the biological function is clearly understood. Secondly, predictions for genomic elements are based upon our current knowledge. Thus sequencing projects will be unable to determine novel genomic elements or those elements which deviate from the current consensus or norm.

The genetic approach to genome analysis involves determination of the function of nucleotide sequence through the disruption of the genome by mutation. Two distinct approaches to mutation exist. A targeted approach to mutagenesis involves the disruption of nucleotide sequence at specific sites within the genome. In *C. elegans* targeted gene disruption can be carried out by the insertion, detection and excision of the transposon Tc1 (Plasterk *et al.* 1992). Gene disruption in yeast can be carried out by homologous recombination between cloned DNA fragments and genomic sequence

1

(Rothstein 1983). The second approach to mutagenesis is non-targeted mutagenesis. Non-targeted mutagenesis employs mutagens such as ethyl methanosulfonate (EMS) (Coulondre and Miller 1977), UV light (Doudney 1976; Stewart *et al.* 1991), gamma radiation (Rosenbluth *et al.* 1985) and formaldehyde (O'Donnel *et al.* 1977) to mutate the genome at almost random positions. Targeted mutational analyses are biased toward our current understanding of the genome since such studies concentrate upon nucleotide sequence for which a function has already been predicted (Li *et al.* 1993; Charron *et al.* 1992). Due to this fact it is more likely that genetic studies which utilize the non-biased approach of non-targeted mutagenesis will discover novel genomic constituents which are currently outside our realm of knowledge.

The correlation of the genetic and physical maps enables the assignment of a biological function to a nucleotide sequence. The correlation of genetic and physical maps is difficult due to the fundamental differences in the basis of each map. Genetic map distances are based upon the recombinational frequencies between loci. Physical maps are measured in nucleotides and may be represented by contiguous overlapping clones (Coulson *et al.* 1986; Hartl *et al.* 1992; Riles *et al.* 1993), restriction endonuclease sites (Kohara *et al.* 1987) or sequence tag sites (STSs) (Olson *et al.* 1989; Weber and May 1989).

In this thesis I describe the correlation of the genetic and physical maps within the *unc-22* region of *Caenorhabditis elegans*. The approach used for this study was the creation of transgenic nematode strains containing cosmids and plasmids derived from the physical map. This approach also led to the cloning and sequencing of the essential gene, *let-653*, which encodes a mucin-like protein and also the elucidation of putative transcriptional control elements for *let-653*.

The free living soil nematode *Caenorhabditis elegans* is particularly amenable to both genetic and nucleotide sequence analysis. It has a short generation time

2

(3.5 days at 20°C) and small size which facilitates genetic analysis (Brenner 1974). The genetic map contains more than 1000 identified genetic loci most of which have been identified mutationally (J. Hodgkin, R. Durbin and M. O'Callahan, personal communication). Genetic study is aided by the fact that *C. elegans* strains may be stored in liquid nitrogen for an indefinite period of time and because *C. elegans* can exist as either a self-fertilizing hermaphrodite or as a male. The complete cell lineage of *C. elegans* has been determined and found to be invariant, with adult individuals possessing 959 somatic cells (Sulston and Horvitz 1977; Kimble and Hirsh 1979; Sulston *et al.* 1983) and the complete connectivety of the 302 neurons has also been determined (White *et al.* 1986). The haploid *C. elegans* genome has been estimated to contain 100 megabases of DNA, which is the smallest known genome of any metazoan (Coulson *et al.* 1991) and has been found to possess a gene density higher than mammalian genomes and also possesses smaller genes with fewer and smaller introns (Blumenthal and Thomas 1988). A physical map of the genome exists (Coulson *et al.* 1988; Coulson *et al.* 1991) in the form of a series of contiguous arrays of overlapping cosmid and yeast artificial chromosomal clones (YAC) (Burke *et al.* 1987). The *C. elegans* genome is currently being sequenced (Sulston *et al.* 1992) and over 2 megabases of sequence has currently been generated (M. Berks, personal communication). The protein coding elements of the *C. elegans* genome are also being elucidated through the generation of expressed sequence tags (ESTs) (Waterston *et al.* 1992). Information about both the genetic and physical maps is readily available through ACEDB, an integrated relational *C. elegans* database (R. Durbin and J. Thierry-Mieg, unpublished results).

Essential genes can be defined as those required for growth, development and reproduction. Such genes have been studied in many different organisms such as mice (Juriloff *et al.* 1992), *Drosophila* (Russell *et al.* 1992; Eberl *et al.* 1992; Hilliker *et al.* 1980) and yeast (Diehl and Pringle 1991). Studies of essential genes have been carried

out extensively in *C. elegans* (Rose and Baillie 1980; Howell *et al.* 1987; Howell and Rose 1990; Sigurdson *et al.* 1984; Rosenbluth *et al.* 1988; Meneely and Herman 1979, 1981; Rogalski and Riddle 1988; Johnsen and Baillie 1991). Essential genes in *C. elegans* have proven to be by far the richest source of genetic loci (Starr *et al.* 1989; Charest *et al.* 1990; Johnsen and Baillie 1991) and have been indispensable in the construction of a high resolution genetic map for *C. elegans*.

The *unc-22* region on chromosome IV in *C. elegans* represents a 2 map unit interval defined by the extent of the deficiency *sDf2* (Rogalski *et al.* 1982). Genetic studies in this region have utilized the reciprocal translocation *nT1(IV;V)* as a recombinational suppressor (Ferguson and Horvitz 1985). This region has been studied extensively for the presence of essential genes (Rogalski *et al.* 1982; Rogalski and Baillie 1985; Clark *et al.* 1988, Clark and Baillie 1992). In the interval encompassed by *let-60* and *unc-22* (figure 1) all the genes except *unc-22* have been found to be essential. These essential genes are *let-60*, *let-56* (Moerman and Baillie 1981), *let-92* (Roglaski and Baillie 1985), *let-653* (Clark and Baillie 1988), *let-660* (M. Marra, unpublished results), *dpy-20* (S. Brenner and R. Edgar, unpublished results) which affects body morphology and where a single temperature sensitive lethal allele has been isolated (D. Suleman, personal communication) and the maternal effect lethal gene *par-5* (K. Kempheus, personal communication) which is required for fertility (figure 1). Mutants in the *unc-22* gene are viable and possess a characteristic uncoordinated phenotype with "twitching" of the body wall musculature (Brenner 1974). Prior to this study all of the genes within the *unc-22* region with the exception of *let-92* and *let-660* had been correlated to the physical map (figure 2).

In conjunction with the genetic analysis in this interval extensive molecular analysis has also been carried out. The *unc-22* gene has been sequenced and found to encode a muscle component possessing a myosin light chain kinase domain

4

Figure 1

The genetic map of the *let-60-unc-22* interval with flanking markers *let-65* (Rogalski *et al.* 1982) and *let-52* (Moerman and Baillie 1980).

let-65 let-60 dpy-20 par-5 let-92 let-653 let-56 unc-22 let-52

let-660

sDf19

sDf9,65

sDf61

sDf7

sDf8

0.04 m.u.

5b
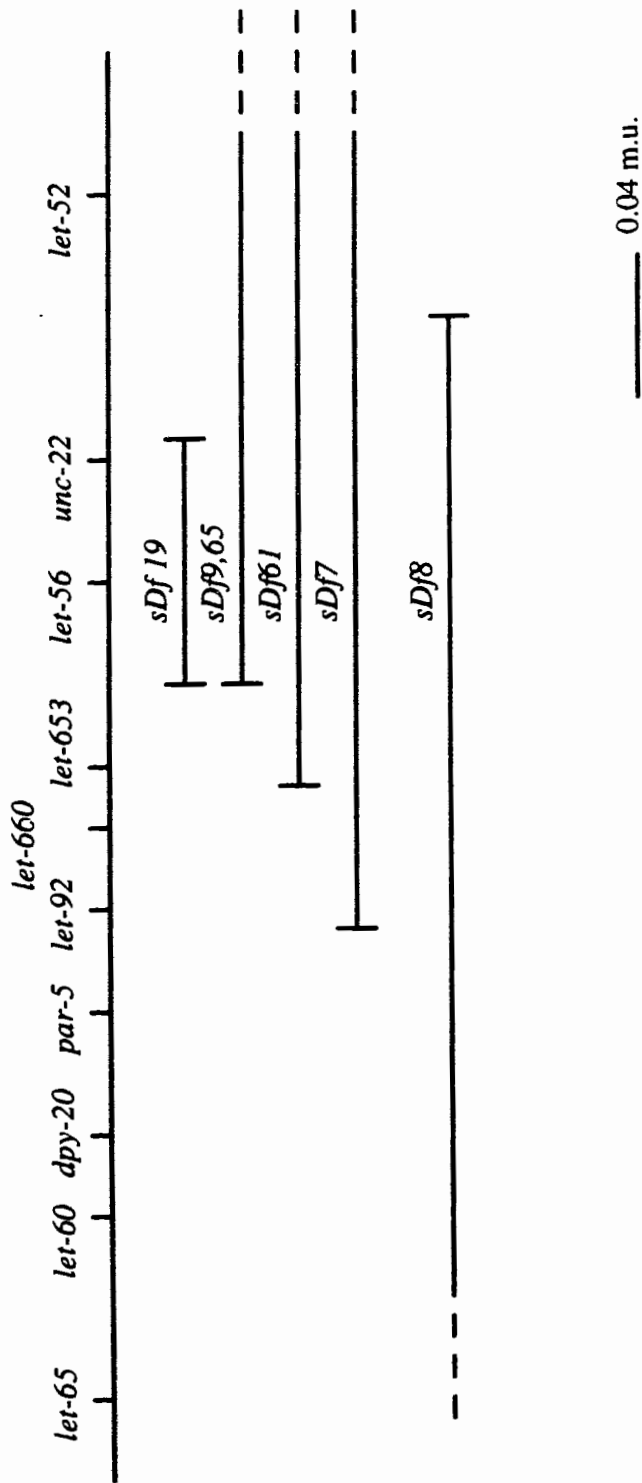
Figure 2

The genetic and physical correlation of genes within the *let-60-unc-22* interval prior to this study. The clones represented by bold lines are those which have been correlated to the genetic map. C33B8 contains *let-60* (Han and Sternberg 1990), C35H3 contains *dpy-20* (Clark 1990), C38H7 tentatively contains *par-5* (D. Shakes, personal communication), C29E6 contains *let-653*, C11F2 contains *let-56* (Clark and Baillie 1992) and LDM22 contains *unc-22* (Benian *et al.* 1989).

6b
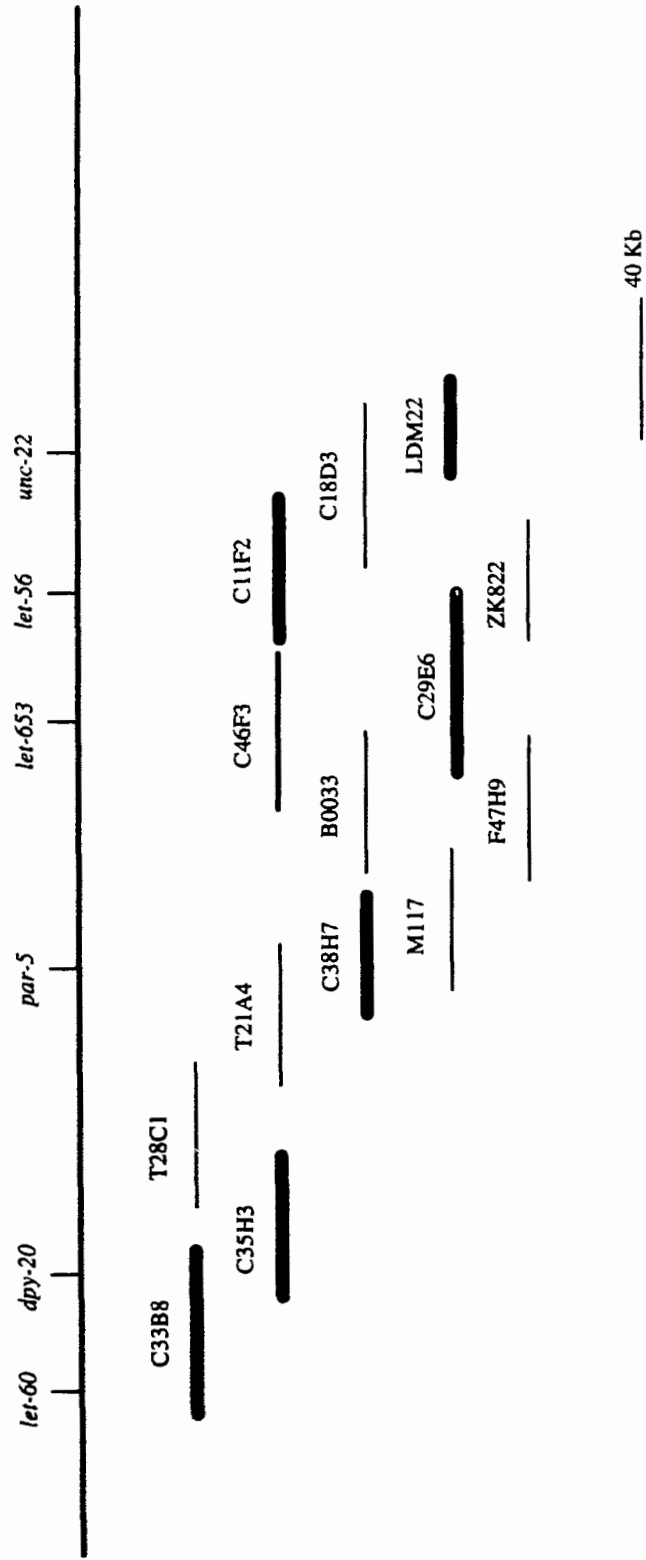
(Benian *et al.* 1989). *let-60* has been found to encode the *C. elegans* homologue of the

human N-*ras* protein (Han and Sternberg 1990). The *dpy-20* gene has been sequenced (D.

Suleman, D.V. Clark and D.L. Baillie, manuscript in preparation). At least 9 coding

elements exist in this region (Prasad and Baillie 1989; Schein *et al.* in press). These

coding elements include sodium proton antiporter, pyridoxal cofactor dependent

decarboxylase, L-proline transporter and a glucose transporter like protein (Marra *et al.*

1993; Schein *et al.* in press).

The molecular analysis of the region has revealed few clues to any pattern

of genomic organization which may exist within this region. The coding elements studied

so far have shown little functional relatedness. Prasad and Baillie (1989) showed that of

seven coding elements detected through Northern blot analysis five showed maximal

hybridization to RNA from L2 stage animals. This indicates that this region may be to

some degree transcriptionally synchronized.

Clark and Baillie (1992) estimated that the *unc-22* region is at most 73%

saturated for mutations in essential genes. This suggests that many of these coding

elements identified do not possess an essential role in *C. elegans* development and have a

dispensable or redundant function. Indeed this prediction has been confirmed by Schein

et al. (1993) who showed by means of a homozygous deficiency strain that neither the

L-proline nor glucose transporter like proteins were essential for development and

reproduction. These results indicate that many of the coding elements in *C. elegans*

when mutated may not effect the viability of the organism under laboratory conditions.

Thus essential genes in *C. elegans* represent an important subset of genes whose products

are indispensable for the appropriate development of a fertilized egg to a fertile adult.

The work presented in this thesis was carried out in order to determine the

molecular nature of the *let-653* gene. The phenotype of let-653 homozygotes is

suggestive of a dysfunction of the secretory/excretory apparatus of *C. elegans*, therefore a

7

molecular dissection of this locus would provide insights into the function and mode of

action of the secretory/excretory apparatus.

MATERIALS AND METHODS

Nematode strains and culture conditions

C. elegans strains were maintained on petri plates containing nematode growth medium (NGM) streaked with *Escherichia coli* strain OP50 (Brenner 1974). The genetic nomenclature used follows that of Horvitz *et al.* (1979).

Strains used

The wild type strain N2 (var. Bristol) was obtained from the stock collection at the Medical Research Council, Cambridge, England. All the chromosome IV mutations except for *unc-31(e169)* (Ferguson and Horvitz 1985) were isolated at Simon Fraser University: *unc-22(s7)* (Moerman and Baillie 1979), *let-653(s1733)* (Clark and Baillie 1992), *let-92(s504)*, *let-92(s677)* (Rogalski and Baillie 1985), *let-660(s1996)* (Marco, Marra personal communication). *unc-22(s7)* and *unc-31(e169)* were used as markers. *unc-22(s7)* is a conditionally semi-dominant mutation whereby *unc-22(s7)/+* individuals twitch in 1% nicotine solution (Sigma). *unc-22(s7)* homozygotes are unconditional twitchers.

The lethal mutations studied were previously induced on chromosome IV possessing the markers *unc-22(s7)* or *unc-22(s7)* and *unc-31(e169)*. To prevent the recombination of the lethal alleles away from these markers the lethal bearing chromosomes were balanced over the reciprocal translocation *nT1(IV;V)* (Ferguson and Horvitz 1985).

Microscopy

Arrested larvae were photographed using Normarski optics (Sulston and Horvitz 1977). Worms were mounted on 5% agar pads containing 10µl S buffer (Brenner 1974). Photographs were taken using Kodak Technical Pan film at A.S.A. 32 on an Olympus AHBS3 microscope. To prevent unc-22(s7) homozygotes from twitching during photography 1µl of 10mM sodium azide solution was added to the mounted specimen (Sulston and Horvitz 1977).

Origin of cosmid clones

Cosmid clones used in this study were obtained from A.R. Coulson and J.E. Sulston (Coulson et al. 1986) at the MRC, England: they are C46F3, C11F2 and BOO33. These cosmids contain the vector pJB8 (Ish-Horowicz and Burke 1981) which contains an ampicillin resistance gene.

Preparation of DNA for germline transformation

Cosmid DNA was prepared using the large scale alkaline lysis method as described by Sambrook et al. (1989). DNA was prepared from 500ml LB liquid cultures containing 30µg/ml ampicillin. These cultures were incubated overnight in a shaking incubator at 37°C and 300 r.p.m. The DNA was purified on CsCl density gradients (Sambrook et al. 1989). The gradients were spun for 24 hours at 60,000 r.p.m. using a Beckman L8-80 ultracentrifuge and Ti-70.1 rotor. Cosmid DNA bands were collected from centrifuge tubes and the ethidium bromide was removed by several extractions with water saturated butanol. CsCl was removed by precipating the DNA in 4 volumes of

water and 8 volumes of 95% ethanol. Precipitated DNA was redissolved in 500μl TE buffer.

Quantification of the purified DNA was performed by measuring the absorbance of the DNA solution at 260nm using an Ultrospec III UV/visible spectrophotometer (Pharmacia).

Plasmid DNA for germline transformation was prepared by alkaline lysis miniprep method outlined in Sambrook *et al.* (1989) although the phenol/chloroform purification step was omitted. No further purification step was found to be necessary apart from centrifugation of the DNA solution at 13,000 r.p.m. for 20 minutes in a benchtop centrifuge to pellet cellular debris. Quantification of the plasmid DNA was carried out by electrophoresing a *Eco*RI or *Pst*I restriction enzyme digested sample on an ethidium bromide stained 0.7% agarose gel and comparing the intensity of the sample bands to a known quantity of 1Kb ladder (Bethesda Research Laboratories).

Germline transformation techniques

The methodology of germline transformation adopted was that described by Mello *et al.* (1991). This involves a single injection of DNA solution into the large syncytium of one or both of the gonadal arms. The injection solution contained a total of approximately 100ng/μl of DNA in TE buffer, considered to be optimal for the formation of heritable extrachromosomal arrays (Mello *et al.* 1991). To detect the presence of extrachromosomal arrays a plasmid containing the dominant roller gene *rol-6(su1006)* (Kramer *et al.* 1990) was included in the injection solution. The *rol-6(su1006)* bearing plasmids used were pRF4 (Kramer *et al.* 1990) and pCes1943 (S. Jones and B. Barbazuk, unpublished results). The presence of *rol-6(su1006)* in extrachromosomal arrays acts as a dominant morphological marker as individuals which possess the extrachromosomal

11

array will display the Rol-6 phenotype. This allows easy identification of transformed individuals and maintenance of transformed strains. The relative DNA concentration in the cosmid injection solutions was usually 20:1 pRF4: experimental DNA. However, in instances where Rol-6 tranformants were not obtained, presumably because the experimental DNA was encoding a protein at poisonous levels, the concentration of experimental DNA was decreased and the pRF4 concentration increased accordingly to maintain the total DNA concentration of the injection solution at 100-150 ng/μl.

Restriction enzyme digestions

Restriction enzymes were obtained from either Bethesda Research Laboratories (BRL) or from Pharmacia. The buffer used was 1-Phor-all as recommended by Pharmacia. Restriction enzyme digestions were carried out for a minimum of one hour at the recommended temperature with at least a twofold excess of restriction enzyme.

Subcloning of cosmid and plasmid DNA

Subcloning was carried out by the shotgun method. The target DNA and vector DNA was digested with the appropriate restriction enzyme(s) and heated to 68°C for 20 minutes. Vector and target DNA were ligated together as outlined by Snutch (1984). The vector used was pBluescript I (SK⁻) (Stratagene). Completed ligations were transformed into *E. coli* strain DH5α (Hanahan 1983). Transformed cells containing vector and insert were identified as white colonies on agar plates containing ampicillin (100μg/ml), X-gal (5-bromo-4-chloro-indolyl-B-D-galactoside) and IPTG (isopropyl-thiogalactoside). Plasmid DNA was prepared by the alkaline lysis miniprep method as described previously.

Agarose gel electrophoresis

Gel electrophoresis was carried out using 0.7% w:v agarose gels in 0.5X TBE

buffer [10X TBE is 89mM Tris, 89mM Borate, 2.5mM EDTA (pH8.3)] with 0.2ug/ml

ethidium bromide. Gels were electrophoresed in 0.5X TBE buffer at 1-5V/cm. The

marker DNA used was 1Kb ladder (Bethesda Research Laboratories). The resolved DNA

fragments were visualized and photographed using a 300nm wavelength transilluminator.

Labeling of DNA probes

For restriction mapping both subclones and cosmids were labeled in their entirety.

The subsequent vector band present on the Southern blot provided a useful positive

control and size marker. DNA was labeled using two methods.

1. ^{32}P labeled probes.

50ng of DNA was labeled with ^{32}P-dATP using the random oligolabeling

technique of Feinberg and Vogelstein (1983). Random hexamers were obtained from

Pharmacia. Following the labeling procedure 50μl distilled H_2O was added and the

reaction was passed through a Sephadex spin column. The probes were denatured by

boiling for five minutes and quenched on ice before hybridization to filters. Probes were

not reused.

2. Flourescein-labeled probes.

50ng of DNA was labeled using the ECL labeling system (Amersham) using the manufacturer's specifications.

Transfer of DNA to nylon membranes.

Following electrophoresis the gel was soaked in 0.2M HCl for 10 minutes. The gel was then washed with double distilled water twice and soaked in 1.5M NaCl, 0.5M NaOH for 45 minutes, with three changes of solution. The gel was then rinsed twice in double distilled water and soaked for 30 minutes in 1M Tris, 1.5M NaCl (pH 7.4). A Southern blot was then made from the gel using the bi-directional transfer method as described in Sambrook *et al.* (1989). The nylon membrane used was Genescreen (New England Nuclear) and membranes were soaked in 10X SSC for 30 minutes prior to blotting.

Hybridization of probes to DNA filters.

Prior to hybridization, DNA filters were immersed in hybridization solution - [5X SSPE, 0.3%SDS and 2.5X Denhardt's solution (1X Denhardt's solution being 0.02% BSA, 0.02% Ficoll, 0.02% polyvinyl pyrrolidone)] at 68°C for one hour. Fresh hybridization solution and denatured probe was added at between 0.1 to 0.25 ml/cm^2 and allowed to hybridize at 68°C overnight in a shaking waterbath.

Filters were washed in 0.1X SSC, 0.2%SDS at 68°C three times in a shaking waterbath. In the case of ^{32}P labeled probes, autoradiography was carried out on air-dried filters using Kodak X-OMAT K film exposed at -80°C overnight. For flourescein labeled

probes, hybridized probe was visualized using the ECL detection system (Amersham). This involves the binding of an anti-flourescein antibody conjugated to horse radish peroxidase to the hybridized flourescein labeled probe. H_2O_2 and luminol are then added to the filter and the subsequent reduction of the H_2O_2 coupled to the oxidation of the luminol is catalyzed by the horse radish peroxidase. The oxidation of a luminol molecule causes the release of a photon. Hybridized probes were visualized using Kodak X-OMAT K film at room temperature. Exposure times varied from 1 minute to 30 minutes.

Screening a lambda ZAP cDNA library

A cDNA clone was isolated from a lambda ZAP library (Stratagene) constructed by Barstead and Waterston (1989). The RNA template used in the construction of the cDNA library was derived from a mixed stage population of *C. elegans*.

The cDNA library was screened using methodology outlined in Sambrook *et al.* (1989). After three rounds of screening positive phage were isolated and the pBluescript plasmid was excised according to the Stratagene protocol.

For the screening of the cDNA library the probe used was a subcloned DNA fragment which had been isolated from its vector. This was achieved by running restriction enzyme digested DNA on a 0.7% agarose gel ; 0.5X TBE ; 0.4μg/ml ethidium Bromide. The desired fragment was visualized using a 360nm illuminator and cut from the gel. The gel slice was then spun through silanized glass wool on a desktop centrifuge at 6000 r.p.m. for 15 minutes and the DNA solution in 0.5X TBE was collected. The concentration of excised DNA fragment was estimated by gel electrophoresis.

DNA sequence determination

DNA sequencing was carried out on an Applied Biosystems (ABI) Model 373A automated sequencing machine using protocols and reagents supplied by ABI. The sequencing strategies involved employed the use of both labeled primers (M13 -21 forward and M13 reverse) and the use of labeled di-deoxy terminators. Custom oligonucleotides used for DNA sequencing were designed with the aid of the OLIGO computer program (Rychlik and Rhoads 1989). The custom oligonucleotides were synthesized on a model 391 ABI oligonucleotide synthesizer.

DNA for sequencing was prepared using the alkaline lysis procedure as described previously. No further purification was found necessary other than centrifugation of the DNA solution at 13,000g to pellet cellular debris.

Exonuclease III/S1 nuclease deletions (Henikoff 1987) of cDNA and genomic clones were prepared using the Erase-a-base system (Promega).

DNA sequence analysis

Nucleotide sequences were manipulated and translated using the Eyeball Sequence Editor (ESEE) program (Cabot and Beckenbach 1989). The amino acid sequences predicted using ESEE were used to search the SWISS database and the translated GENBANK database. The searches were carried out using the BLAST algorithm (Altschul et al. 1990). Search computations were carried out at the National Center for Biotechnology Information (NCBI) using the BLAST network service.

Protein sequence analysis

Putative N-glycosylation sites were predicted using the program PROSITE contained within the computer program PCGENE (version 6.25). The algorithm of Rao and Argos (1988) and the algorithm of Eisenberg *et al.* (1984) were both used to establish the presence or absence of transmembranous domains. Computer programs utilizing these algorithms were also present in PCGENE (version 6.25). Hydrophobicity plots were generated using the computer program GREASE (Pearson and Lipman 1988).

RESULTS

Genetic and physical correlation of genes between *par-5* and *let-56*

The tentative rescue of *par-5* by the cosmid C38H7 (D. Shakes, personal communication) and the rescue of *let-56* by the cosmid C11F2 (Clark and Baillie 1992) indicated that these genes were physically separated by approximately 100kb of DNA (figure 2). The genetic map of this region indicated that this 100kb of DNA contained the genes *let-92*, *let-660* and *let-653*. This region also contained the right breakpoints for the deficiencies *sDf19*, *sDf9*, *sDf65*, *sDf61* and *sDf66* (figure 1).

Clark and Baillie (1992) demonstrated by germline transformation that the co-injection of the cosmids F47H9, C29E6 and ZK822 was able to provide a heritable extrachromosomal array capable of rescuing *let-653(s1733)*. A transient rescue of *let-653(s1733)* was also obtained by microinjection of the cosmid C29E6.

The above data proved instrumental in allowing the identification of candidate cosmids for germline transformation rescue of genes within the *par-5-let-56* interval. A summary of germline transformation results is presented in Table 1. The construction of strains and extrachromosomal arrays is outlined in Tables 2 and 3. The correlation of the genetic and physical maps in the *unc-22* region after this study is shown in Figure 3.

The rescue of *let-92*

let-92(s504) was rescued by the sequences present on cosmid B0033. The heritable extrachromosomal array capable of rescuing *let-92(s504)* was denoted *sEx26*. However complete rescue of *let-92* was not achieved. The normal blocking stage of

18

Figure 3

The genetic and physical correlation of genes within the *let-60-unc-22* interval after this study. The clones represented by bold lines are those which have been correlated to the genetic map. C33B8 contains *let-60* (Han and Sternberg 1990), C35H3 contains *dpy-20* (Clark 1990), C38H7 contains *par-5* (D. Shakes, personal communication), B0033 contains *let-92*, C46F3 contains *let-653* (this study), C29E6 contains *let-653*, C11F2 contains *let-56* (Clark and Baillie 1992) and LDM22 contains *unc-22* (Benian *et al.* 1989). *let-660* was positioned to B0033 and up to 2.4 kb to the right of B0033 (this study).

19b
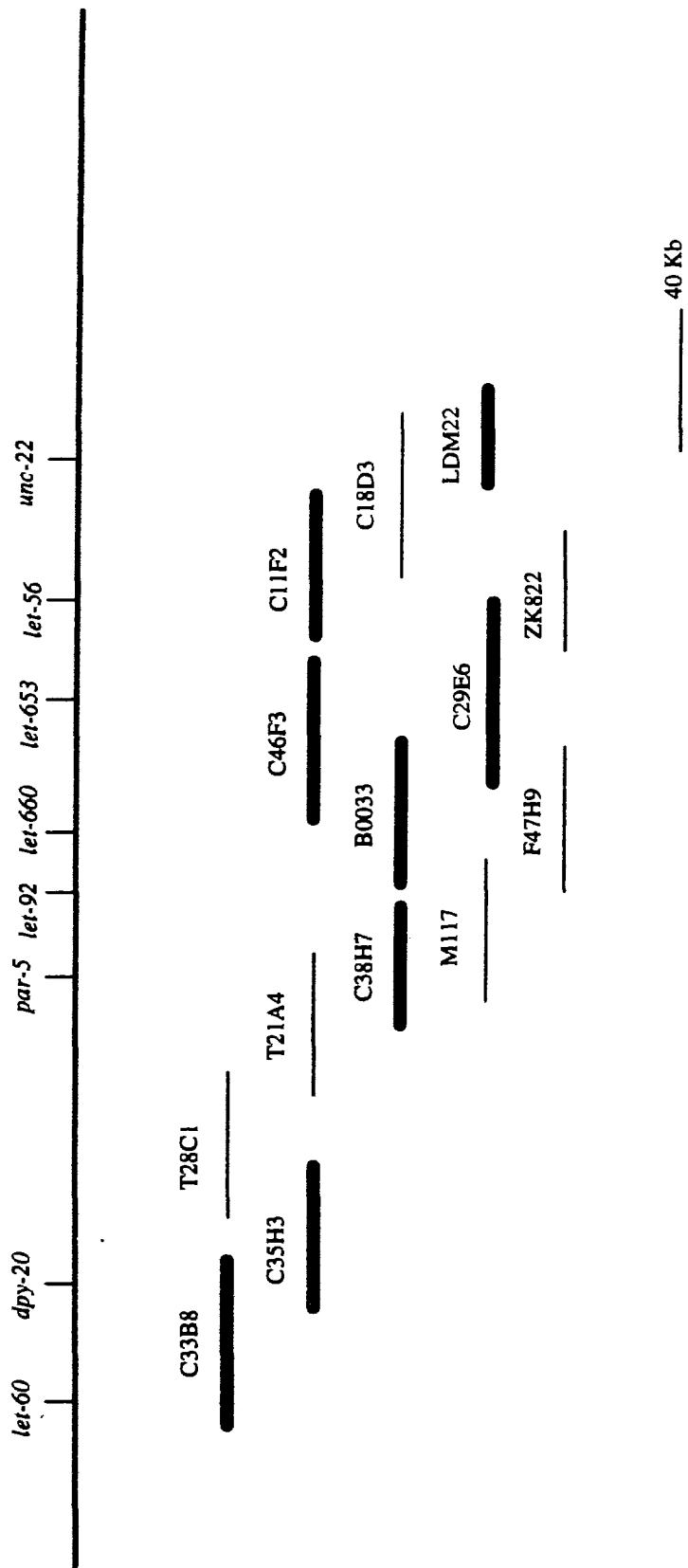
Table 1

Summary of cosmid rescue results

Gene	Allele	Strain	sEx	Cosmid	Rescue
let-653	*s1733*	BC1733	*sEx19*	C46F3	YES
let-653	*s1733*	BC1733	*sEx20*	B0033	NO
let-92	*s504*	BC2903	*sEx26*	B0033	YES[a]
let-92	*s677*	BC1899	*sEx26*	B0033	YES[a]
let-92	*s504*	BC4540	*sEx28*	B0033 C38H7	YES[a]

[a] Complete rescue was not achieved. *let-92* homozygotes attained adulthood but none of their eggs were able to hatch.

Table 2

Strains constructed for cosmid rescue

Strain	Method[a]	Genotype	Cosmid
BC4209	trans.	*let-653(s1733) unc-22(s7) unc-31(e169)/nT1(IV);+/nT1(V);sEx 19*	C46F3
BC4273	trans.	*let-653(s1733) unc-22(s7) unc-31(e169)/nT1(IV);+/nT1(V);sEx 20*	B0033
BC4435	trans.	*let-92(s504) unc-22(s7)/nT1(IV);+/nT1(V);sEx 26*	B0033
BC4440	cross	*+/+; sEx 26*	B0033
BC4441	cross	*+/+; sEx 19*	C46F3
BC4442	trans.	*let-653(s1733) unc-22(s7) unc-31(e169); sEx 19*	C46F3
BC4540	trans	*let-92(s504)unc-22(s7)/nT1(IV);+/nT1(V);sEx28*	B0033 C38H7

[a] Denotes whether the extrachromosomal was introduced into the strain via transformation (trans.) or by introducing the extrachromosomal array via a genetic cross (cross).

21

Table 3

Construction of extrachromosomal arrays for cosmid rescue

Array	Cosmid	Strain injected	Cosmid conc.	pRF4 conc.
sEx19	C46F3	*let-653(s1733) unc-22(s7) unc-31(e169)/nT1(IV);nT1(V)*	10ng/ul	150ng/ul
sEx20	B0033	*let-653(s1733) unc-22(s7) unc-31(e169)/nT1(IV);nT1(V)*	5ng/ul	150ng/ul
sEx26	B0033	*let-92(504) unc-22(s7) unc-31(e169)/nT1(IV);nT1(V)*	5ng/ul	100ng/ul
sEx28	B0033 C38H7	*let-92(s504) unc-22(s7) unc-31(e169)/nT1(IV);nT1(V)*	5ng/ul 5ng/ul	150ng/ul[a]

[a] pCes1943 was used in this experiment instead of pRF4.

let-92 homozygotes is early larval (Rogalski and Baillie 1985). *let-92(s504)* homozygotes bearing *sEx26* achieve adulthood and lay eggs but none of these eggs hatch and they undergo developmental arrest as a bundle of cells.

let-92 is within 0.1 map units of *par-5*, a maternal effect lethal locus. Therefore the above observations could be explained if *let-92(s504)* was actually a small deficiency which covered *par-5* and *let-92*. If B0033 could rescue *let-92* but not *par-5* then a maternal effect lethal phenotype would be observed. To investigate this possibility, the ability of *sEx26* to rescue another allele of *let-92* was studied. *sEx26* was crossed into the strain BC1899 which possesses *let-92(s677)*. The strategy for this cross is outlined in Figure 4. *let-92(s677)* individuals bearing *sEx26* behaved in a similar fashion to *let-92(s504)* homozygotes, whereby the eggs of rescued adults were developmentally arrested. Since both *let-92* alleles behave similarly in the presence of *sEx26* it is unlikely that either of the *let-92* alleles are deficiencies or possess a second site mutation on chromosome IV conferring maternal effect lethality.

In order to determine whether the partially rescued *let-92* individuals lacked viable sperm or a factor that is present in the male germline (Meneely and Herman 1981), rescued *let-92* homozygotes were mated with N2 males . The presence of wild type sperm was unable to restore fertility to *let-92(s504) unc-22(s7); sEx26* individuals.

These results invoke a necessary maternal contribution of the *let-92* gene product which is not supplied by the extrachromosomal array *sEx26*. Thus B0033 may lack the appropriate enhancers or upstream regulatory elements to enable appropriate maternal contribution of the *let-92* gene protein or message. *par-5* has been tentatively assigned to the cosmid C38H7 which is close to the cosmid B0033 (<10kb) and it is conceivable that *let-92* and *par-5* may share transcriptional control elements responsible for maternal contribution. In order to investigate this further I constructed extrachromosomal arrays containing the cosmids C38H7 and B0033. B0033 and C38H7

23

Figure 4

An outline of the method used to cross *sEx26* into a strain bearing *let-92(s677)*.

let-92(s504)unc22(s7)/nT1(IV);+/nT1(V); sEx26   X   N2 males

↓   Pick Rol-6 individuals which do not twitch
in nicotine

+/nT1(IV);+/nT1(V); sEx26

↓   ☿⟩   Allow to self for several generations

+/nT1(IV);+/nT1(V);sEx26 or *+/+;sEx26*   X   *let-92(s677)unc-22(s7)/nT1(IV);+/nT1(V)* males

↓   Pick Rol-6 individuals which twitch
in nicotine

let-92(s677)unc22(s7)/+;sEx26

↓ ☿⟩   Allow to self to study rescued
let-92 homozyg otes

let-92(s677)unc-22(s7);sEx26

24b

do not overlap (Coulson *et al.* 1986). However, C38H7 may contain transcriptional control elements able to allow maternal contribution. If these were close enough to the B0033 sequence within the extrachromosomal array then it might allow the appropriate maternal contribution of the *let-92* gene protein or message. Two stable extrachromosomal arrays containing cosmids B0033 and C38H7 were produced. One extrachromosomal array, denoted *sEx28*, was unable to restore fertility to *let-92(s504)* homozygotes. The other extrachromosomal array was able to restore fertility to *let-92(s504)* homozygotes. The presence of this extrachromosomal array reduced the viability of individuals bearing it and could not be maintained.

The above results suggest the presence of a transcriptional control element on C38H7 capable of allowing a maternal contribution of the *let-92* protein or message. This result does not, however, rule out a simple dosage effect of B0033 within the extrachromosomal array. Thus the ability of an extrachromosomal array to rescue *let-92* fully may be related to the number of copies of B0033 present within the array itself. Evidence supporting this is provided by the fact that extrachromosomal arrays were not produced if B0033 was injected at concentrations of more than 5ng per μl in a total DNA concentration of 100ng per μl. Therefore high copy numbers of B0033 within an extrachromosomal array appear to have a poisonous effect. Thus the extrachromosomal array able to confer fertility to *let-92* homozygotes may have contained more than usual amounts of the cosmid B0033 accounting for both its ability to confer fertility and its inability to be maintained.

It is possible that *par-5* itself represents an upstream regulatory element of *let-92*. However this is unlikely as *par-5* and *let-92* are separated by the right breakpoint of *sDf7* and also because *par-5(it121)* and *let-92(s504)* when in *trans* were found to complement each other.

The rescue of *let-653*

The transient rescue of *let-653(s1733)* by cosmid C29E6 (Clark and Baillie 1992) was corroborated by the stable rescue of *let-653(s1733)* with the cosmid C46F3. C46F3 overlaps extensively with C29E6 (Coulson *et al.* 1986). The heritable extrachromosomal array containing C46F3 was denoted *sEx19*. The cosmid B0033 failed to rescue *let-653(s1733)*.

Phenotypic description of *let-653*

The larval stage at which *let-653(s1733)* homozygotes underwent lethal arrest was estimated to be late L1 or early L2. This was derived from the study of 13 *let-653(s1733)* homozygotes of which 5 were observed to undergo an L1 molt. Growth prior to the lethal arrest was found to be retarded when compared to larval sizes of the background phenotype Unc-22 Unc-31 as determined by Clark (1990) (figure 5).

The life span of *let-653(s1733)* homozygotes ranged from 20 hours to 70 hours after being deposited as an egg and is outlined in Figure 6. Death was deemed to have occurred after the pharynx could no longer be observed pumping. Coincident with the lethal arrest is the appearance of a large vacuole slightly anterior to the lower pharyngeal bulb (figure 7). The size of the vacuole was found to range between 0.02 and 0.17mm (figure 8) and in some cases extended along the entire length of the body (figure 7).

Figure 5

Graph showing the maximum length attained by *let-653(s1733)* homozygotes. 13

homozygotes were examined in this study, of which 5 were observed undergoing a molt.

The larval stage marker was derived from the larval stages of the background genotype

unc-22(s7) unc-31(e169) as determined by Clark (1990).

27b
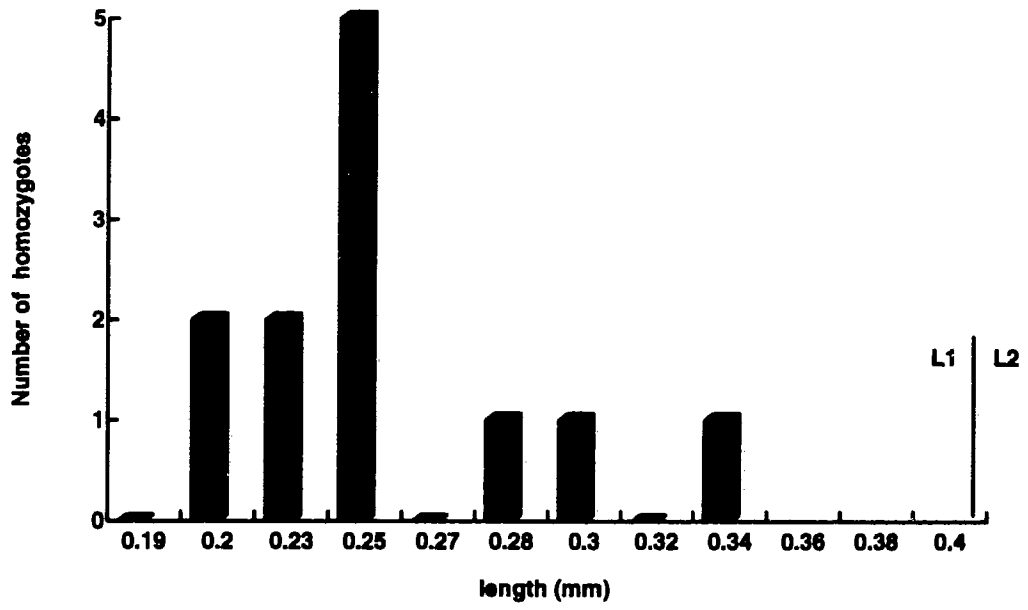
Figure 6

Graph showing the range of longevity displayed by *let-653(s1733)* homozygotes at 20°C. 13 *let-653(s1733)* homozygotes were studied. Death was deemed to have occurred when the pharynx could no longer be observed to be pumping.

Number of homozygotes dying in 10 hour period (y-axis)

Time after egg laying (Hours) (x-axis)
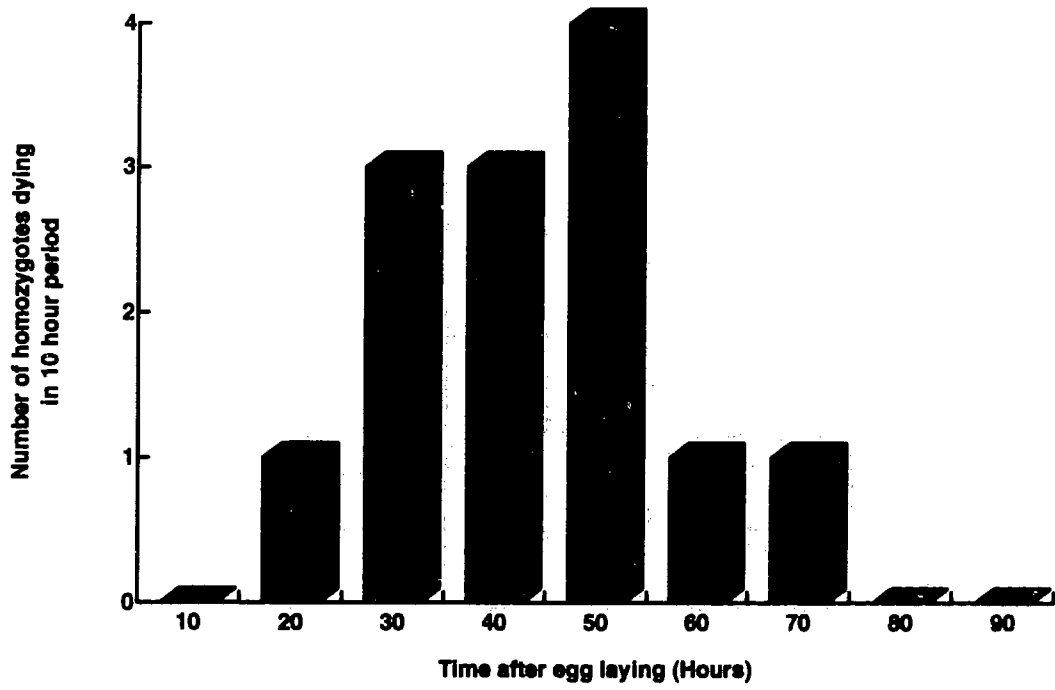
Figure 7

Normarski photographs of the Let-653 phenotype. All photographs were taken at 20°C.
let-653(s1733) homozygotes were derived from the strain BC3261. A: Wild type L1
larva. B: *let-653(s1733)* larva with vacuole just forming. C,D: *let-653(s1733)* larvae
possessing larger vacuoles. E: Close up of vacuole of a *let-653(s1733)* individual. The
vacuole has caused the displacement of the pharynx against the body wall.

F: *let-653(s1733)* larva where the vacuole has extended along almost the entire length of
the body.

29b
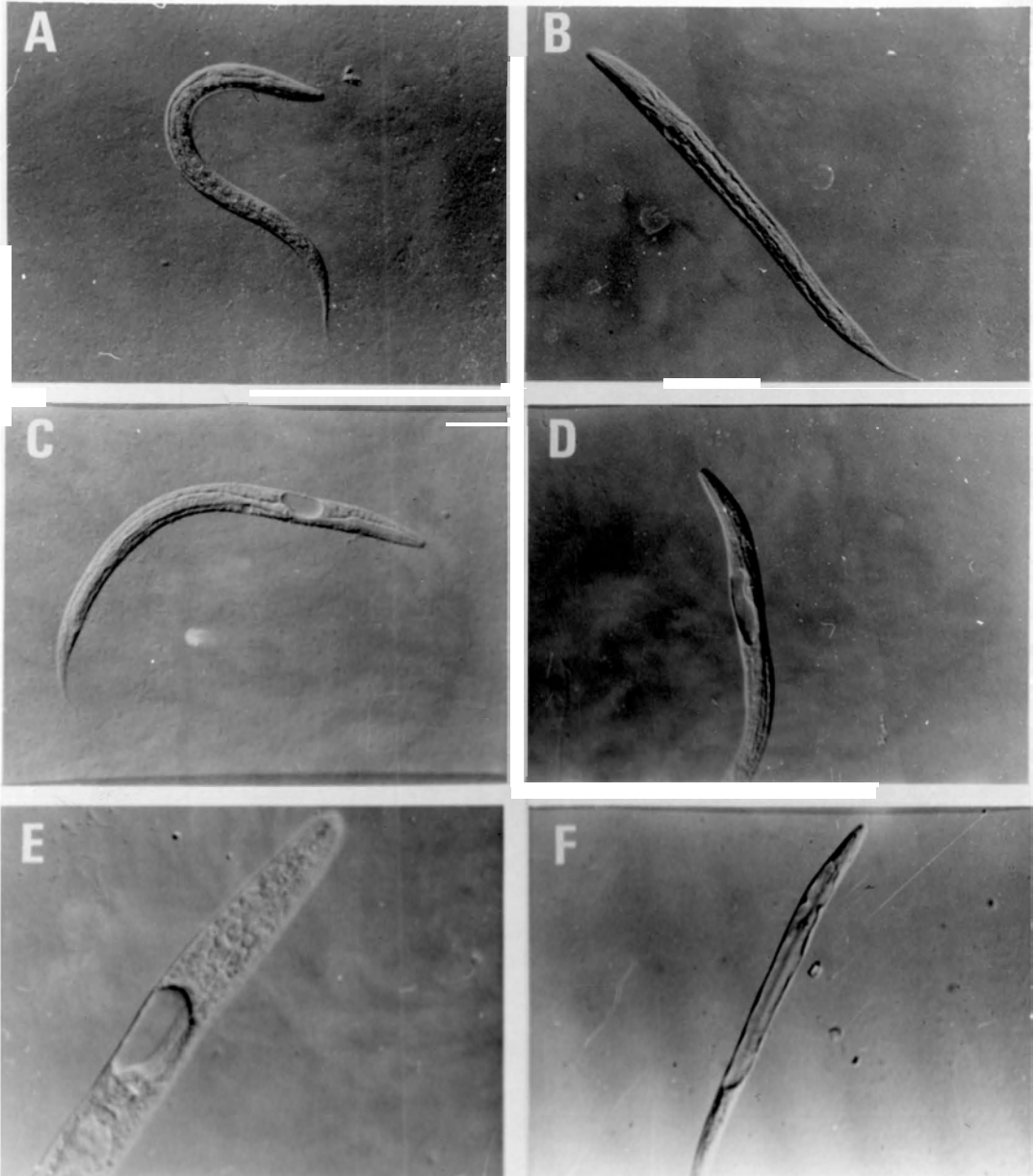
Figure 8

Graph showing the variation in the maximum size of the vacuole observed in

let-653(s1733) homozygotes at 20°C. 13 *let-653(s1733)* homozygotes were studied. The

length of the vacuole was defined as its length in an anterior to posterior direction.

Length of vacuole (mm)

30b
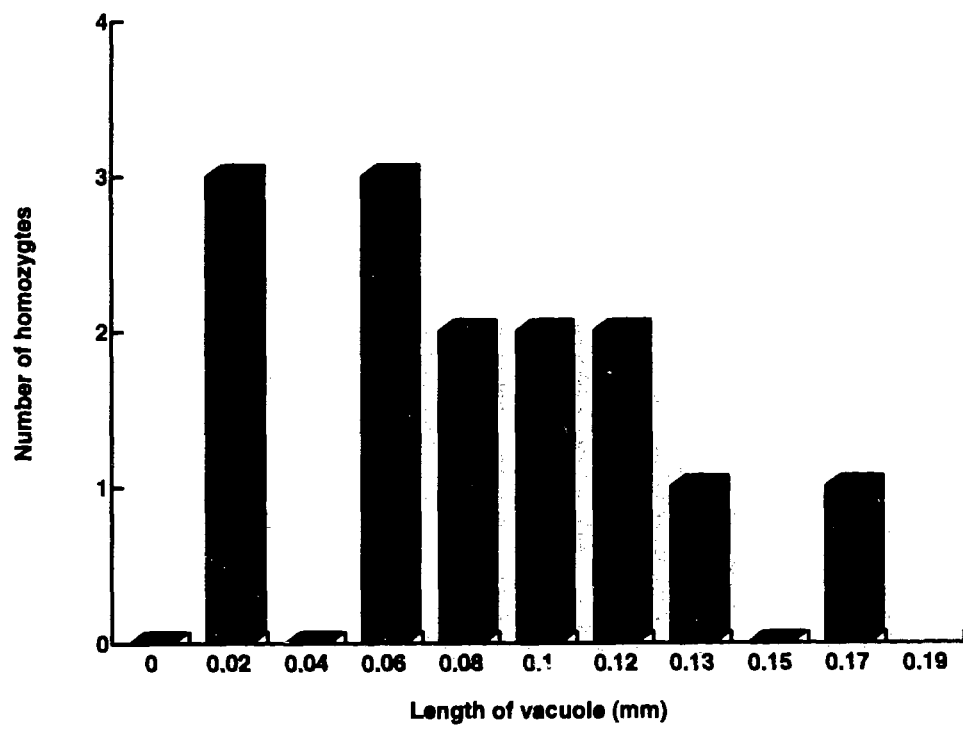
Molecular Characterization of the *let-653* gene

The ability of the cosmid C46F3 to rescue *let-653* and the failure of B0033 to rescue *let-653* positioned the *let-653* gene to the right arm of C46F3.

To establish the extent of overlap between B0033 and C46F3, C46F3 was restriction digested with *Eco*RI and the fragments separated by electrophoresis and Southern blotted onto nylon membrane. The membrane was then probed with ^{32}P labeled B0033. *Eco*RI was the restriction enzyme of choice as this separates the 5.4kb pJB8 vector sequence (Ish-Horowicz and Burke 1981) from the cloned *C. elegans* insert.

A restriction map of the *let-653* candidate region was produced by subcloning *Eco*RI and *Pst*I fragments of C46F3 into the pBluescript (SK⁻) vector. *Eco*RI digests of C46F3 *Eco*RI subclones were separated by electrophoresis, Southern blotted and probed with ^{32}P labeled *Pst*I subclones. For restriction mapping both subclones and cosmids were labeled in their entirety. The subsequent vector band present on the Southern blot providing a useful positive control and size marker. The inferred restriction map of the *let-653* candidate region is shown in Figure 9.

Plasmid rescue of *let-653*

Plasmids from the *let-653* candidate region were used for micro-injection rescue assays for *let-653(s1733)*. The results of the micro-injection assays are outlined in Table 4. Co-injection of pCes1910, pCes1940, pCes1906, pCes1903 and pCes1907 resulted in an Unc twitcher in the F1 which segregated Unc twitchers. This rescued strain was designated *let-653(s1733)unc-22(s7)unc-31(e169);sEx19* (BC4412). To confirm that BC4412 contained an extrachromosomal array BC4412 was crossed with N2 males and Rol-6 progeny were seen within the F1 outcross progeny. This was a necessary

31

Figure 9

A restriction map of the *let-653* candidate region. This map also shows the extent of overlap between the cosmids B0033 and C46F3. Parentheses indicate *Eco*RI fragments which have not been ordered with respect to each other. Plasmid subclones are designated with the prefix pCes.

C46F3

EcoRI — 3.6kb — pCes1905 — (9.0, 6.0, 2.3kb) — EcoRI — 2.4kb pCes1907 — EcoRI EcoRI 1.6kb pCes1906 — EcoRI 3.8kb pCes1910 — EcoRI 5.1kb — EcoRI 4.8kb pCes1903

PstI PstI 1.5kb pCes1933 — PstI 8.0kb pCes1940 — EcoRI PstI 2.5kb pCes2022

B0033

EcoRI — (9.0, 6.0, 2.3kb) — EcoRI

32b
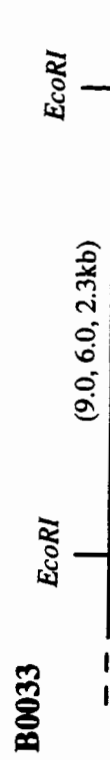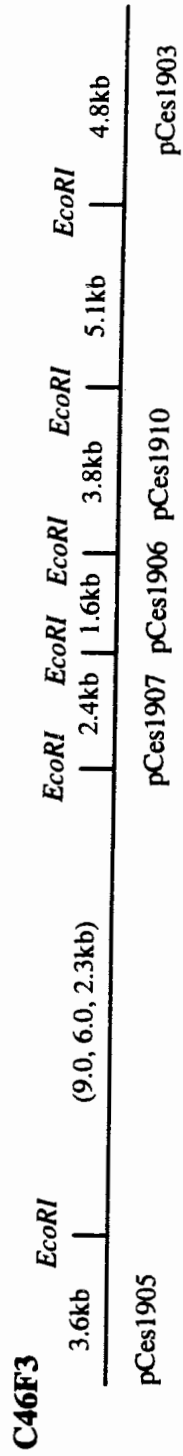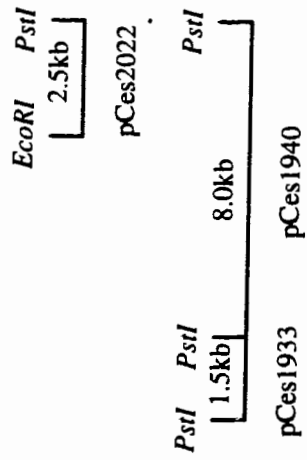
Table 4

Construction of extrachromosomal arrays for plasmid rescue of *let-653(s1733)*.

Array	Plasmid	pCes conc.	pRF4 conc.	Strain[a]	Rescue *let-653*
sEx21	pCes1903	5.0 ng/μl	100ng/μl	BC4273	YES
	pCes1906	5.0 ng/μl			
	pCes1907	5.0 ng/μl			
	pCes1910	5.0ng/μl			
	pCes1940	1.25ng/μl			
sEx22	pCes1940	7.5ng/μl	100ng/μl	BC4423	NO
sEx23	pCes1903	10ng/μl	100ng/μl	BC4424	NO
sEx24	pCes1910	5.0ng/μl	100ng/μl	BC4429	NO
	pCes1907	5.0ng/μl			
sEx27	pCes1906	10ng/μl	100ng/μl[b]	BC4539	NO
sEx43,44,45	pCes1906	60ng/μl	25ng/μl[b]	BC4591	NO
	pCes1940	15ng/μl		BC4590	
				BC4593	
sEx46	pCes1940	15ng/μl	65ng/μl[b]	BC4595	NO
	pCes1907	20ng/μl			
sEx48	pCes1906	30ng/μl	50ng/μl[b]	BC4596	NO
	pCes1907	20ng/μl			
sEx41,42	pCes1906	30ng/μl	40ng/μl[b]	BC4594	YES
	pCes1940	15ng/μl		BC4592	
	pCes1907	20ng/μl			
sEx47	pCes1906	30ng/μl	40ng/μl[b]	BC4597	YES
	pCes1940	15ng/μl			
	pCes1903	20ng/μl			

[a] Denotes strains bearing extrachromosomal arrays. The strain used for the germline transformation was *let-653(s1733) unc-22(s7) unc-31(e169)/nT1(IV); +/nT1(V)* (BC3261).

[b] pCes1943 was used in these experiments instead of pRF4.

confirmation of the *let-653* rescue as the Rol-6 phenotype conferred by the chromosomal array is masked by the Unc-22 Unc-31 phenotype.

Extrachromosomal arrays containing the individual C46F3 subclones present in sEx19 failed to yield a rescue. Rescue of *let-653* could only be achieved by extrachromosomal arrays which contained pCes1906, pCes1940 and pCes1907 (*sEx41* and *sEx42*) or by extrachromosomal arrays which contained pCes1906, pCes1940 and pCes1903 (*sEx47*). Interestingly, extrachromosomal arrays containing only pCes1906 and pCes1940 (*sEx43*, *sEx44* and *sEx45*) were unable to rescue *let-653* (figure 9).

The requirement for both pCes1940 and pCes1906 to rescue *let-653* meant that these plasmid subclones would contain the *let-653* gene. As pCes1906 and pCes1940 represented overlapping subclones it was likely that the *let-653* gene traversed the right breakpoint of pCes1906 and the left breakpoint of pCes1940. An intact *let-653* gene would then be reconstituted through homologous recombination between the two plasmids during the construction of a rescuing extrachromosomal array.

The function of pCes1903 and pCes1907

In order to achieve plasmid rescue of *let-653* the presence of either pCes1903 or pCes1907 was required within the extrachromosomal array. Neither pCes1903 nor pCes1907 possessed any overlap with either pCes1906 or pCes1940 and were unlikely to recombine with their inserted sequences. This means that the action of pCes1903 or pCes1907 does not require being contiguous with the sequences of pCes1906 or pCes1940. This fact and the ability of either pCes1903 or pCes1907 to allow rescue of *let-653* makes it unlikely that these contain any of the coding element of *let-653*. The action of both pCes1903 and pCes1907 is consistent with that of

transcriptional enhancer or control elements required for the appropriate expression of the *let-653* gene.

Isolation of a *let-653* cDNA

Since the *let-653* coding element was thought to extend to the right of pCes1906, pCes1910 was used as a probe to screen a mixed stage *C. elegans* lambda ZAP cDNA library. Approximately 25,000 phage were screened. One positive 3.4kb clone, designated pCes1942, was isolated and was transformed into the rec⁻ host DH5α (Hanahan 1983).

pCes1942 hybridized to both pCes1906 and pCes1910 (data not shown). Since the *let-653* gene was predicted to cross the boundary between pCes1906 and pCes1910, pCes1942 most likely represented at least part of the *let-653* gene transcript. *Eco*RI fragments of pCes1942 were shotgun subcloned into pBluescript(SK⁻). Restriction mapping of pCes1942 was carried out by sequencing from the ends of the subclones and from two ExoIII/SI deletions of pCes1942 denoted pCes1951 and pCes1958. The inferred restriction map is shown in Figure 10.

Sequence analysis of pCes1942

The cDNA was found to be a tandem fusion of two different cDNAs (figure 11). The pCes1942 sequencing strategy is outlined in Figures 12, 13 and 14. Both transcripts were found to correspond with previously identified expressed sequence tags (ESTs) (McCombie, W.R., Kelley, J.M., Aubin, L., Goscoechea, M., Fitzgerald, M.G., Wu, A., Adams, M.D., Dubnick, M., Kerlavage, A.R., Venter, J.C. and Fields, C.A., unpublished results) from a *C. elegans* early embryonic cDNA library (Stratagene).

Figure 10

Restriction map of the cDNA clone pCes1942. E = *Eco*RI. REV and FOR refer to the

orientation of the clone insert with respect to the M13 -20 (FOR) and the M13 reverse

(REV) primers within the pBluescript (SK⁻) vector sequence. pCes1951 and pCes1958

are ExoIII/S1 deletions of pCes1942 created using the enzymes *Sac*II and *Sma*I.

36b
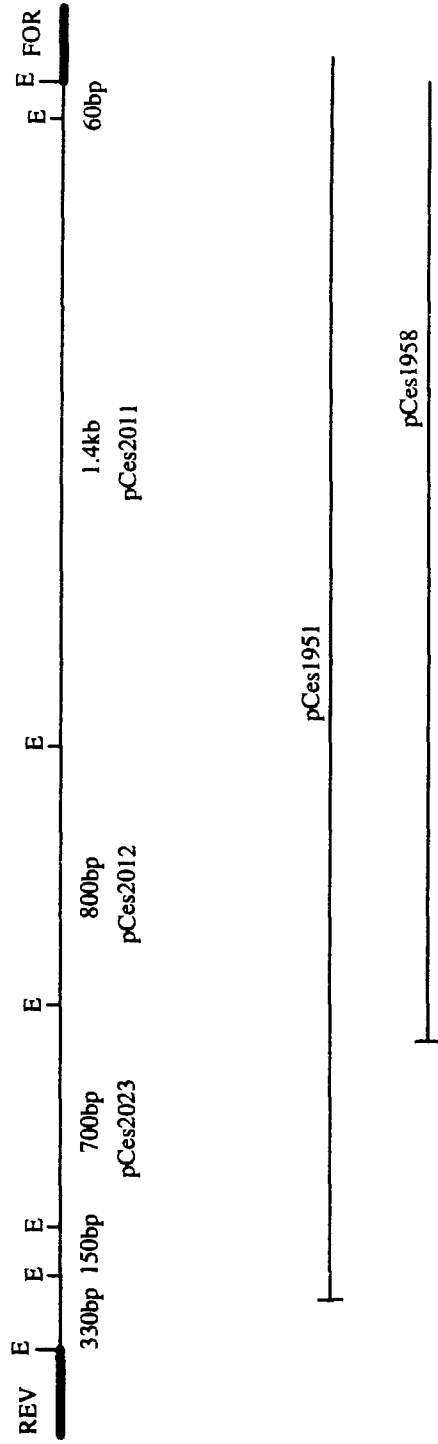
Figure 11

A diagram showing the orientation of the two cDNA transcripts contained within

pCes1942. The DNA sequences shown represent the 3' end of each transcript. The

polyadenylation signal (AATAAA) and poly A tail is underlined in each case. REV and

FOR refer to the orientation of the clone insert with respect to the M13 -20 (FOR) and

the M13 reverse (REV) primers within the pBluescript (SK⁻) vector sequence. E=*Eco*RI.

37b
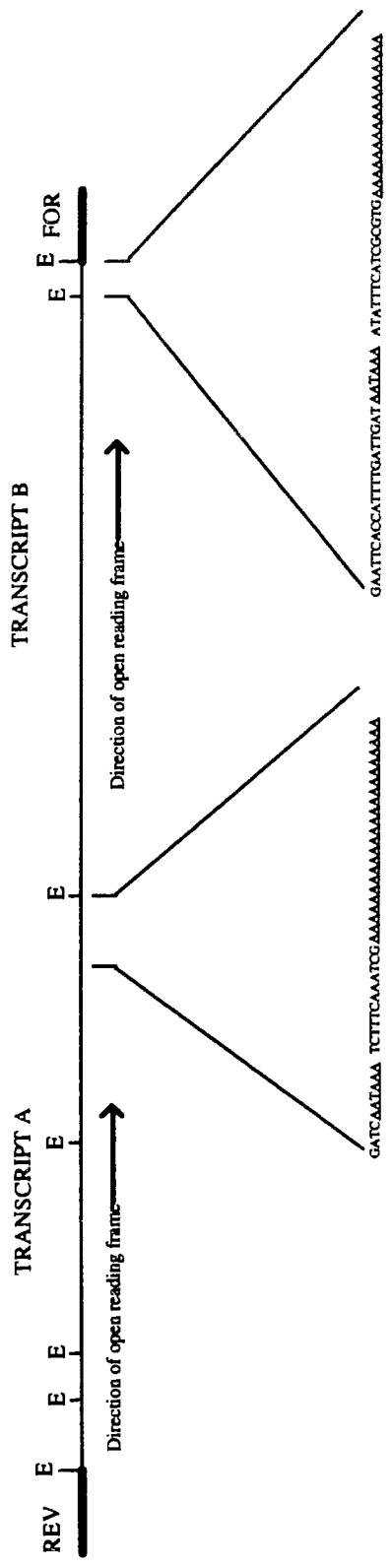
Figure 12

Sequencing strategy for pCes2023. pCes2023 was sequenced using the labeled M13 -20

and M13 reverse primers. Oligonucleotide primers were used to sequence the rest of the

clone using labeled dye terminators. The sequence of the primers used is shown. Base

pair lengths show the position of the primer sequences within the clone. REV and FOR

refer to the orientation of the clone insert with respect to the M13 -20 (FOR) and the M13

reverse (REV) primers within the pBluescript (SK⁻) vector sequence. E=*Eco*RI.

E REV

780bp

5'

ATCTGGTGCAGGCATTCTT G

3'

pCes2023 (780bp)

540bp

220bp

5' GTCAACGGATGTAGAACGAG 3'

FOR E

0bp

38b
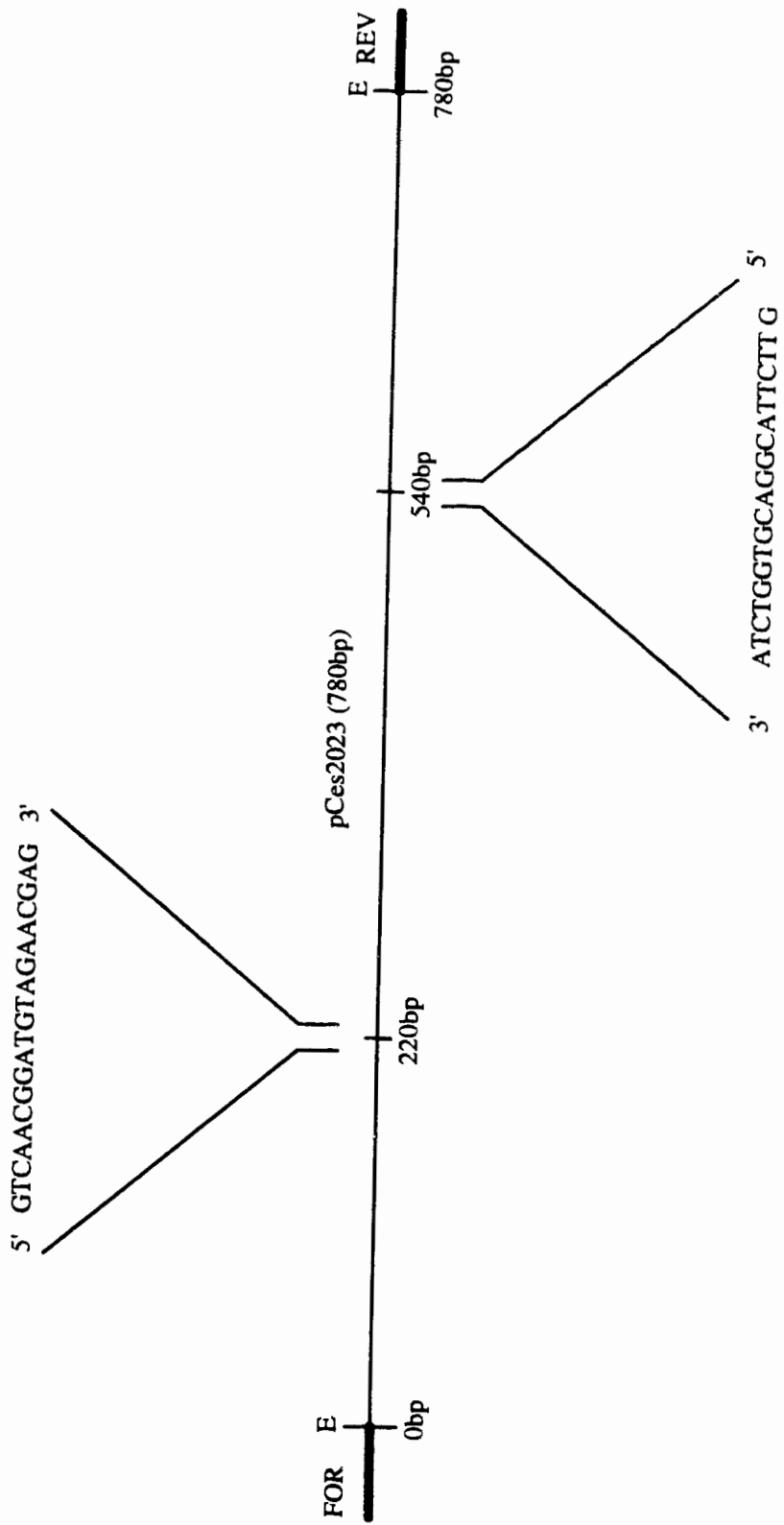
Figure 13

Sequencing strategy for pCes2012. pCes2012 was sequenced using the labeled M13 -20

and M13 reverse primers. An oligonucleotide primer was used to sequence the rest of the

clone using labeled dye terminators. The sequence of the primer used is shown. The base

pair length indicates the position of the primer sequence within the clone. REV and FOR

refer to the orientation of the clone insert with respect to the M13 -20 (FOR) and the M13

reverse (REV) primers within the pBluescript (SK⁻) vector sequence. E=*Eco*RI.

5' ATTGAAGGAGATTGAACGGC 3'

pCes2012 (720bp)

REV E   0bp   262bp   E FOR   720bp

39b
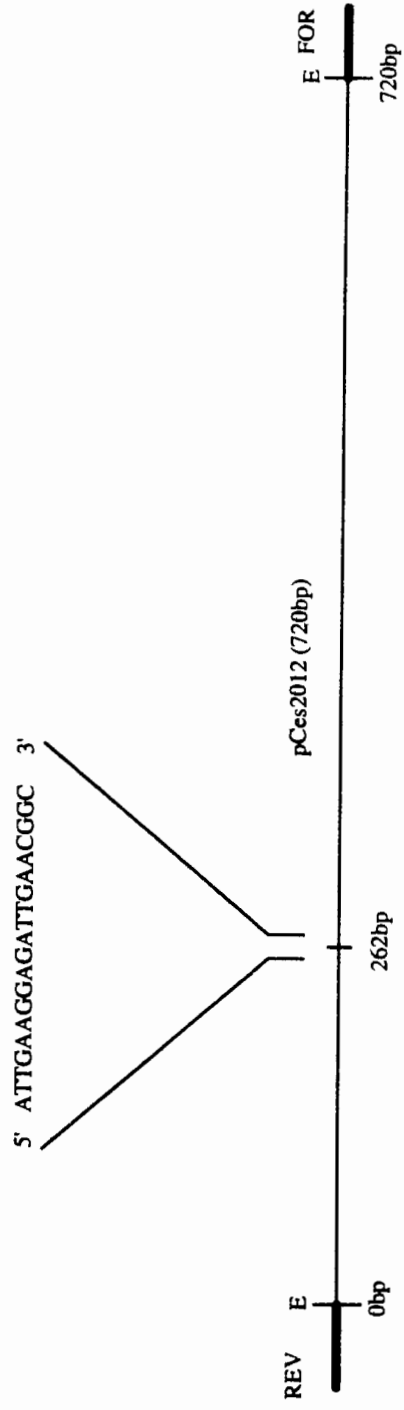
Figure 14

The plasmid clones used in the sequencing of pCes2011. All subclones produced

are ExoIII/SI deletions created using the restriction enzymes *Xho*I and *Kpn*I. REV and

FOR refer to the orientation of the clone insert with respect to the M13 -20 (FOR) and the

M13 reverse (REV) primers within the pBluescript (SK⁻) vector sequence.

FOR

REV

pCes2011 (1.6kb)

pCes2188

pCes2185

pCes2123

pCes2128

pCes2126

pCes2124

pCes2159

pCes2187

pCes2098

pCes2190

pCes2158

pCes2157

pCes2100

40b

Figure 15

Nucleotide sequence and derived amino acid sequence of the 1270 nucleotide transcript

A. The amino acid sequence indicated is based upon the most extensive predicted open

reading frame for this sequence. The polyadenylation signal (AATAAA) and poly A tail

are shown underlined.

```
CGACAAGCCTGGAGACTTCTTTCTGGATATCTTCCAACAAATGCAGAACGTCGTGAAGTA  60
 R  Q  A  W  R  L  L  S  G  Y  L  P  T  N  A  E  R  R  E  V   20

ACTCTTCAATGTAAAAGAGACGAATATTGGCATTATGTAGAACAATATTTCCATTCCCGG 120
 T  L  Q  C  K  R  D  E  Y  W  H  Y  V  E  Q  Y  F  H  S  R   40

TTTGATGATCAGAATGCTGATACGTTCAGACAGATTAATATTGATATTCCAAGAATGTGT 180
 F  D  D  Q  N  A  D  T  F  R  Q  I  N  I  D  I  P  R  M  C   60

CCATTGATTCCGTTGTTCCAGCAGAAAATGGTTCAGGAGATGTTCGAGAGAATACTCTAC 240
 P  L  I  P  L  F  Q  Q  K  M  V  Q  E  M  F  E  R  I  L  Y   80

ATTTGGGCTATCCGTCATCCAGCCAGTGGATACGTTCAAGGAATCAACGATCTCGTTACC 300
 I  W  A  I  R  H  P  A  S  G  Y  V  Q  G  I  N  D  L  V  T   100

CCGTTCTTTGTCGTTTTTCTCTCAGAATTCATTCCACAAGATGTCGAAGTCGGTTCTTTT 360
 P  F  F  V  V  F  L  S  E  F  I  P  Q  D  V  E  V  G  S  F   120

GACGTTTCTCAACTTCCTTTGGAGCAATGCCAGCTCATCGAAGCAGATTCATTTTGGTGT 420
 D  V  S  Q  L  P  L  E  Q  C  Q  L  I  E  A  D  S  F  W  C   140

GTTTCGTCGCTACTCGATTCTATCCAGGATAACTACACATTTTCACAACCCGGAATTCAA 480
 V  S  S  L  L  D  S  I  Q  D  N  Y  T  F  S  Q  P  G  I  Q   160

CGAAAGGTGCTCCAACTTCGCCATTTGATGAGTCGAGTTGATCGGCCTCTACACAAGCAC 540
 R  K  V  L  Q  L  R  H  L  M  S  R  V  D  R  P  L  H  K  H   180

TTGGAATCCAATGGGATTGAATATTTGCAATTCGCATTCCGTTGGATGAATAATCTTTTG 600
 L  E  S  N  G  I  E  Y  L  Q  F  A  F  R  W  M  N  N  L  L   200

ATGCGAGAAATTCCGCTCAGAGCGACTATCAGGCTATGGGACACTTATCTGAGTGAGCCC 660
 M  R  E  I  P  L  R  A  T  I  R  L  W  D  T  Y  L  S  E  P   220

GACGGTTTCATGCAGTTTCATAACTATGTATGCGCTGCGTTCTTACGGACGTGGTCTAAG 720
 D  G  F  M  Q  F  H  N  Y  V  C  A  A  F  L  R  T  W  S  K   240

CAATTGCAAGCCGAAAAAGATTTCCAGGGAGTAATGATTCTTCTGCAAAACTTGCCAACT 780
 Q  L  Q  A  E  K  D  F  Q  G  V  M  I  L  L  Q  N  L  P  T   260

CAATTCATGGGGGTGATCCGTGAGATCTGCGGACTAACTGCAGATGCCTTCTCTCTTCAA 840
 Q  F  M  G  V  I  R  E  I  C  G  L  T  A  D  A  F  S  L  Q   280

TCCGTGTTTGACGGTGCACGCCGNCATCTTTCAGCGCAGGCCGCCTCTCCATAATTACTC 900
 S  V  F  D  G  A  R  R  H  L  S  A  Q  A  A  S  P  *

AAAAATCGATAACCGTGTATGTCTTGTGAATCTGTGATGTTCTATTATTGAGGCTAATCG 960

TTCTTTTCAATATTATTATATTGGGTCCTATTGTGTGTGTGTTTGTCTGTACATTTACTC 1020

GTTCTACATCCGTTGACTAGGTAGGTCTTACTCACTCATTTTCAAATAAGGATTACTTAC 1080

AGAGAGTTCATCCTTTTAACCTACTACTAGTGCCTTTCTATTTTCCTCAATTCCCTTTAA 1140

CTGTAAAATAGTTACTTTTTTGTCCAGAAATCAATTTATCGCCTCCCCTCTCGAGTTTTT 1200

CTTCCACCATTGTTCATCTTTGATCCAATAAATCTTTCAAATCGAAAAAAACAAAAAAAA 1260

AAAAAAAAAA 1270
```

41b

The shorter cDNA, denoted transcript A, 1270bp in length was found to match identically with the EST wEST02625 (Genbank accession number T01904). Transcript A contained a poly A tail 12bp upstream of which was found a polyadenylation signal (AATAAA). The sequence of transcript A and its predicted open reading frame is presented in Figure 15. A database search of the predicted transcript A protein against both the Swiss Prot and translated Genbank databases revealed no tangible similarity with any previously characterized protein or gene.

The second transcript, denoted transcript B, was 2208bp in length. Transcript B corresponded to the EST wEST02977 (Genbank accession number T02256). Transcript B contained a poly A tail 15bp upstream of which was a polyadenylation signal (AATAAA). Transcript B possessed a single extensive open reading frame terminating in a TAA stop codon 169bp upstream of the polyadenylation signal. The sequence of transcript B and its predicted open reading frame is presented in Figure 16.

The genomic sequence of the *let-653* gene

Genomic subclones pCes1906 and pCes1910 derived from C46F3 were sequenced (figures 17 and 18). The sequence data generated showed that transcript B represented the *let-653* transcript. Sequence analysis of pCes2022 showed that this contained the 60bp of the 3' end of the *let-653* gene (figure 19).

Transcript B is unlikely to represent a full length cDNA for two reasons. First, the most 5' ATG present in the transcript is not in frame with the predicted open reading frame. Second, within the genomic sequence 5' to the cDNA the open reading frame is preserved for 21bp at which point a putative intron is encountered. This intron, 53bp in length contains stop codons in all three reading frames and possesses strongly conserved 3' and 5' splice sites. This is the most common length for introns in *C. elegans*

42

Figure 16

Nucleotide sequence and derived amino acid sequence of the 2208 nucleotide transcript

B. The amino acid sequence indicated is based upon the most extensive predicted open

reading frame for this sequence. The polyadenylation signal (AATAAA) and poly A tail

are shown underlined.

43a

```
CATTTTATGTTCGATGGCCTCGAGTTCGTCTCAACTTCACGGCAGTTGCTGAAGCGAGAC  60
      F   Y   V   R   W   P   R   V   R   L   N   F   T   A   V   A   E   A   R      19

TTTCATTAAAAGGGTGTCAATCCGCATGTTCACTTGGAGAAGATCCAGTCAGTCCAGGAA  120
  L   S   L   K   G   C   Q   S   A   C   S   L   G   E   D   P   V   S   P   G      39

AACAATTAGCGTGTGCTGCAGTGAATCATCAAGCATCTCCAGATGGCTTTTCACATCTTT  180
  K   Q   L   A   C   A   A   V   N   H   Q   A   S   P   D   G   F   S   H   L      59

GCGCCGTTTTCCAGCCGCATCAACTACAAAATGTTGACGGTTATGTTGAGGCGGATGATC  240
  C   A   V   F   Q   P   H   Q   L   Q   N   V   D   G   Y   V   E   A   D   D      79

GATTCACATTCTACTGGAAATATTGCCTTCCTTCCACTCGAAAATGTTCTGGAGAATACG  300
  R   F   T   F   Y   W   K   Y   C   L   P   S   T   R   K   C   S   G   E   Y      99

CATTCACATATCTTTCGGATCGTTATATGGATCCCCAAGTCAGTTATAAAATGGACAACG  360
  A   F   T   Y   L   S   D   R   Y   M   D   P   Q   V   S   Y   K   M   D   N      119

ACAGCAAATTTAGAAGAGTGTCTTTCGGATTGTTTGGATGGAAAAAATCATTTGAATGCC  420
  D   S   K   F   R   R   V   S   F   G   L   F   G   W   K   K   S   F   E   C      139

GTTCAATCTCCTTCAATAGGACAGACGGAGGATGTCATATGTCTAAGGATTCACAAATTT  480
  R   S   I   S   F   N   R   T   D   G   G   C   H   M   S   K   D   S   Q   I      159

CACGACCGGAAGCCATTCGATTGAACAATAACCCCAACTATCGAATTGATTATTACGAGA  540
  S   R   P   E   A   I   R   L   N   N   N   P   N   Y   R   I   D   Y   Y   E      179

ATAACTGTTACAATTTATCCGAATCATTCACATTCAAACATGAATGTCGTGATAATGGAA  600
  N   N   C   Y   N   L   S   E   S   F   T   F   K   H   E   C   R   D   N   G      199

TCTCAGTCAGCGTCAAGTCTCGTCTTCCTTATACTGGAGCCATTTATGGACTTTATGACT  660
  I   S   V   S   V   K   S   R   L   P   Y   T   G   A   I   Y   G   L   Y   D      219

TCTTCACTTGTCGGACTGAACCAAAAGAAGCCACTGAATTCGATCACTTTTTCCCATATC  720
  F   F   T   C   R   T   E   P   K   E   A   T   E   F   D   H   F   F   P   Y      239

AGACTGTCAGCAAGAACTGTTCGGATTCTATTAAGTACAAGGGAAACGAAATGGTACTTG  780
  Q   T   V   S   K   N   C   S   D   S   I   K   Y   K   G   N   E   M   V   L      259

AAGTTGTCTTATCTACTGATGGAATAGAGCCACTTTATTTCATTACTCCTGAAGATTTGA  840
  E   V   V   L   S   T   D   G   I   E   P   L   Y   F   I   T   P   E   D   L      279

CATACCAGGCAAAATGTCCAATTAGTGGTGTAATAGCAAAGGATCCCGCTAACATAAAAT  900
  T   Y   Q   A   K   C   P   I   S   G   V   I   A   K   D   P   A   N   I   K      299

CCTCGGCTCATTTGGATAACAGAAACAAGGCAATGGAAGCATCAGCGCATGCATTGTTTG  960
  S   S   A   H   L   D   N   R   N   K   A   M   E   A   S   A   H   A   L   F      319

AGTTACTCTCGAAAACTGGAGATGATGAAGCCCTTCAAAACACATTCCCACTTCCATTGA  1020
  E   L   L   S   K   T   G   D   D   E   A   L   Q   N   T   F   P   L   P   L      339

CTACCACAACAGAAGCAATAAGTGATGTGCCAACTACTACCGTACAAACTTCAACTACCG  1080
  T   T   T   T   E   A   I   S   D   V   P   T   T   T   V   Q   T   S   T   T      359
```
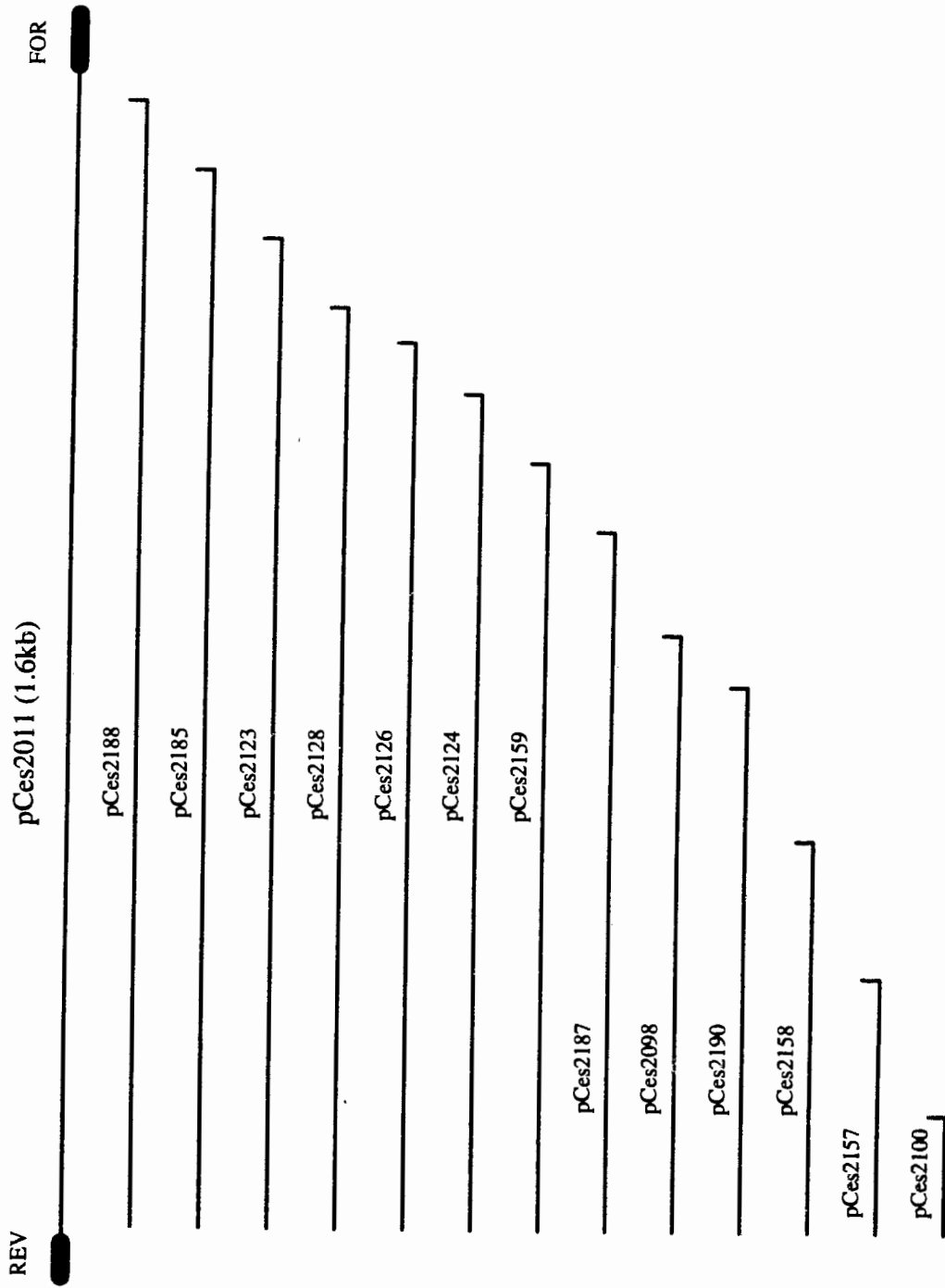
```
TACCAACCACACCATCCAAAACAACTGCTACCACTACAACAACTCCAAAGCCAACCACTA 1140
V  P  T  T  P  S  K  T  T  A  T  T  T  T  T  P  K  P  T  T    379

CAGAGACAGCTACAACTTCTTCATCAACAACCACAGTAACCACACAAAAACCTACAACCG 1200
T  E  T  A  T  T  S  S  S  T  T  T  V  T  T  Q  K  P  T  T    399

TTACGTCAACAACAACATTACCATCCACAACGGCATCTACAACAACTAAAACTACTACAA 1260
V  T  S  T  T  T  L  P  S  T  T  A  S  T  T  T  K  T  T  T    419

GTACGCCCACGTCACCACAAACAACGACCACTCATGTAGGAGCTCCCGCATCTTCCGTTG 1320
S  T  P  T  S  P  Q  T  T  T  T  H  V  G  A  P  A  S  S  V    439

CATCGGTGGTCATGATGGCTCAACACTTGCCGGAAAACCAAAAGTTCCAGTGATTTTCG 1380
A  S  V  A  H  D  G  S  T  L  A  G  K  P  K  V  P  V  I  F    459

ATATTTTCCACAACGGACAACCAGTTGAAGCCGTCGTTGTTGGAACAAAAATCAGTCTAT 1440
D  I  F  H  N  G  Q  P  V  E  A  V  V  V  G  T  K  I  S  L    479

CATTCCGTCCACACTATCCAATTCCACCGGAGTACGTAGACGTTCGTGGTTGCCAAGTTG 1500
S  F  R  P  H  Y  P  I  P  P  E  Y  V  D  V  R  G  C  Q  V    499

AGCCAATTGATCCAAAATATGAATGGGAGCACGAGCCACTGTTCATCATAAGAGATGGAT 1560
E  P  I  D  P  K  Y  E  W  E  H  E  P  L  F  I  I  R  D  G    519

GTCCTGCTGACGGTGTTGGTCTCGTCTGCCCACCAACACATTCTGAGTTTGGAGCAAAAG 1620
C  P  A  D  G  V  G  L  V  C  P  P  T  H  S  E  F  G  A  K    539

TTTCCGTTGAGGCATTCAGATATCAAACAACTGGACAAGTACAGTACTCCTGTCTCGTGA 1680
V  S  V  E  A  F  R  Y  Q  T  T  G  Q  V  Q  Y  S  C  L  V    559

GAATATGTCCATTTGCTCCGTGTCCGAAGAACACATGCGACGACGTGGAAGGCTGTGACA 1740
R  I  C  P  F  A  P  C  P  K  N  T  C  D  D  V  E  G  C  D    579

GTAGCTACATGCATCGATACAGGCGAGAATTATCACTTGAAGACATCAAAAAAGCATTGG 1800
S  S  Y  M  H  R  Y  R  R  E  L  S  L  E  D  I  K  K  A  L    599

AAGCAAACCCAGAACTTGCTTCTCAGTTCGGAATTTCACCGTCCGCGTTTGCAAGAAATC 1860
E  A  N  P  E  L  A  S  Q  F  G  I  S  P  S  A  F  A  R  N    619

CGTCGAAATCTAAAAATTTCACAAGCGTTGTTGAAGAACAACAAAGAATCGCACTCGGCG 1920
P  S  K  S  N  F  T  S  V  V  E  E  Q  Q  R  I  A  L  G    639

GAGATTATCTTGTTCGGCGTCGATTGATTGTTGTCAACTCTGAAGATCAATTGAGATATT 1980
G  D  Y  L  V  R  R  R  L  I  V  V  N  S  E  D  Q  L  R  Y    659

ACGTTCGAACTGGAAACATCTGAAAATACACACAAAAAATGTCTTCATGTCAGCTTTATA 2040
Y  V  R  T  G  N  I  *

ACTTTGTTTTTATTCTTCTTTCATTCAATTATCCAAACGCCACACGAGCGGGTCTCACAT 2100

GGAAAACTCTTGGAGTTATTCTTCTTATGATTTAAGTTAAATTTTTTGAATTCACCATTT 2160

TGATTGATAATAAAATATTTCATCGCGTGAAAAAAAAAAAAAAAAAAAAAA        2208
```

Figure 17

The plasmid clones used in the sequencing of pCes1906. pCes2014 was created

by restriction enzyme digestion of pCes1906 with *Pst*I and subsequent religation. All

other clones are ExoIII/SI deletions created using the restriction enzymes *Xho*I and *Kpn*I.

REV and FOR refer to the orientation of the clone insert with respect to the M13 -20

(FOR) and the M13 reverse (REV) primers within the pBluescript (SK⁻) vector sequence.

45b
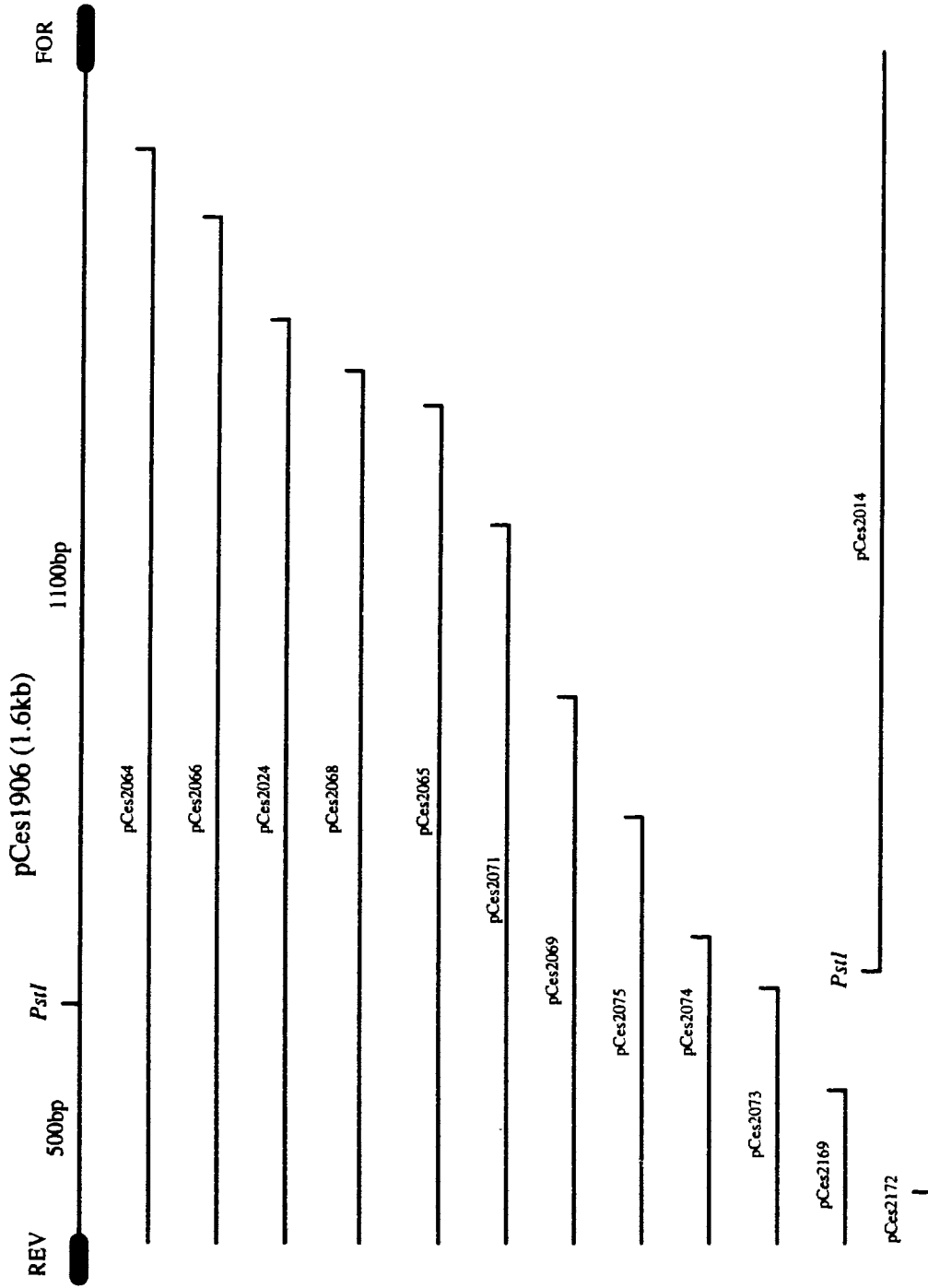
Figure 18

The plasmid clones used in the sequencing of pCes1910. pCes2017, 2019, 2030,

2016, 2025, 2020 and 2027 are subclones of pCes1910 created using the restriction

enzymes indicated. All other clones are ExoIII/SI deletions created using the restriction

enzymes *Xho*I and *Kpn*I. REV and FOR refer to the orientation of the clone insert with

respect to the M13 -20 (FOR) and the M13 reverse (REV) primers within the pBluescript

(SK⁻) vector sequence.

pCes1910 (3.8kb)

SacI

REV   XbaI   ClaI   EcoRV   HindIII   BamHI   FOR

pCes2179

pCes2120

pCes2114

pCes2109

pCes2112

pCes2183

pCes2104

pCes2117

pCes2173

pCes2177

pCes2119

pCes2017   BamHI

pCes2019   EcoRV

pCes2107

pCes2105   SacI

pCes2030

pCes2110

pCes2182

pCes2016   ClaI

pCes2108

ClaI   pCes2025   BamHI

pCes2122

pCes2111

pCes2116

pCes2181   XbaI

pCes2020

pCes2121   XbaI   pCes2027   HindIII

46b

Figure 19

Diagram showing the presence of the 3' end of the *let-653* gene within pCes2022. The
polyadenylation signal (AATAAA) is underlined. Genomic sequence matching the
cDNA sequence is presented in uppercase. REV and FOR refer to the orientation of the
clone insert with respect to the M13 -20 (FOR) and the M13 reverse (REV) primers
within the pBluescript (SK⁻) vector sequence.

47a

FOR
EcoRI
pCes2022 (2.5kb)
PstI
REV

GAATTCACCATTTTGATTGATAAATAATATTTCATCGCttgatattattttgagtttt

47b

(Blumenthal and Thomas 1988). Upstream of the intron the open reading frame is preserved and an in-frame ATG is encountered. Immediately upstream of the ATG a stop codon is present. It is most likely that this ATG represents the initiating methionine of the *let-653* gene.

The *let-653* genomic sequence possesses a putative *trans*-splice acceptor site immediately upstream of the *let-653* potential initiating methionine (Krause and Hirsh 1987). Therefore, it is possible that the *let-653* transcript represents a *trans*-spliced message and that pCes1903 and pCes1907 contain promoter sequences able to transcribe the *let-653* message.

The 5' splice site of intron 6 contains GC in the place of the usual GT. This variance on the consensus sequence has been reported many times previously (Jackson 1991), however this is the first time that this variance has been reported in *C. elegans* (T. Blumenthal, personal communication).

The *let-653* gene was found to contain 10 exons and 9 introns. The entire sequence of the *let-653* gene and the predicted protein sequence is shown in Figure 20. A schematic representation of the *let-653* gene is shown in Figure 21.

The *let-653* protein

The predicted Let-653 protein was found to be 694 amino acids in length and the unmodified protein was predicted to have a weight of 76,965 daltons. The Let-653 protein shows similarity with the family of glycoproteins called mucins. Mucin genes have been studied extensively (for review see Strous and Dekkler 1992). Mucin genes do not show sequence homology with each other at the nucleotide nor amino acid sequence level but share a number of characteristic features. These features include a central threonine and serine rich region which is subject to heavy O-glycosylation. The regions flanking the threonine/serine rich region are cysteine rich and are subject to N-glycosylation. The number of N-glycosylation sites in this region varies between different

48

mucin proteins. The MUC-1, MUC-2 and MUC-3 human mucin genes possess 1, 2 and 2 N-glycosylation sites respectfully; the porcine submaxillary mucin possesses 7 sites and the rat intestinal mucin possesses 13 sites. As most mucins are secreted many possess a hydrophobic amino terminus which is later cleaved off. The hydrophobic terminus being typical of a classical signal sequence for secreted proteins (Crowley *et al.* 1983; Blobel and Dobberstein 1975).

The Let-653 protein was found to possess a threonine/serine rich domain (Exon 7) (figure 22), this region being 46% threonine and 13% serine. Flanking this region were two cysteine rich regions. The amino end contained 14 cysteine residues and the carboxylic end contained 7 cysteine residues. These regions were found to contain six putative N-glycosylation sites (figure 20). Consistent with other mucins none of these predicted N-glycosyaltion sites were found to be present within the threonine/serine rich region.

The hydropathy profile of the Let-653 protein revealed a hydrophobic amino terminus which suggests that Let-653 protein is, like many mucin proteins, secreted (figure 23). The hydropathy profiles shown in Figure 23 also shows that the structure of the Let-653 may be more related to the human MUC-1 protein than the *Drosophila melanogaster* larval glue protein, which like all *Drosophila* glue proteins belongs to the mucin family (Meyerowitz *et al.* 1987; Swida *et al.* 1990). Although some mucin proteins have been determined to be transmembranous (Gendler *et al.* 1990; Ligtenberg *et al.* 1990; Killen *et al.* 1987) the Let-653 protein was not found to possess predicted transmembranous domains.

Inconsistent with most other mucin proteins, however, the threonine/serine rich region of the Let-653 protein does not consist of tandem repeats. However other mucin like proteins have also been reported where the threonine/serine rich region is not made up of tandem repeats. The canine tracheobronchial mucin (Verma and Davidson

Figure 20

The genomic sequence of the *let-653* coding region. Exons are denoted in uppercase letters. Introns are denoted in lower case letters. The predicted 5' coding element and intron sequence is underlined. Putative N-glycosylation sites are denoted with (Cho). Stop codons are represented by an asterisk. The predicted protein sequence is shown for each exon. The single TATA sequence 5' upstream of the initiating methionine is marked in bold type. Dashes represent gaps in intron sequence data. Undetermined nucleotides are designated by 'n'. Intron splice sites which do not conform to the intron splice consensus (Blumenthal and Thomas 1988) are shown boxed. The 5' splice site of intron 6 has GC in place of the usual GT. This variant consensus has been reported in 25 other genes, but not previously in *C. elegans* (Jackson 1991).

```
gaattcgaagtgttttctcccaaatgatgtacctaaacttgttatctaaccacttttctcacttattcgtgcttcgg   80

aggagctttctcttccgctttatacctcgttctccccattcaaatgacccgaaaacagtctcgtctgttcatatt   160

cttcccttatttgttccgtatccaccaatctcaattctcatttcttcctcctaatctgctctctcttcag   240

taaATGGCGACATCCACTAATTTCTCTACTATTGCTAATAGCATTCTACTCGTCAGAAGgtgattcttgaaagtt   320
   * M   R   H   P   L   I   S   L   L   L   I   A   F   Y   S   T   S   S   E

gaacttttcaattagaattttttggttttttccagCATTCGTTCCAAAAATGTAACTCATTTTATGTTCGATGGCCTCG   400
                                      A   F   V   P   K   C   N   S   F   Y   V   R   W   P   R

AGTTCGTCTCCAACTTCACGGCAGTTGCTGAAGCCGAGACTTTCATTAAAGGGGTGTCAATCCGGCATGTTCACTTGGAGAAG   480
V   R   L   N   F   T   A   V   A   E   A   R   L   S   L   K   G   C   Q   S   A   C   S   L   G   E
    (Cho)

ATCCAGTCAGTCCAGGAAAACAATTAGAGTGTGCTGCAGTGAATCATCAAGCATCTCCAGATGGCTTTTCACATCTTTGC   560
D   P   V   S   P   G   K   Q   L   A   C   A   A   V   N   H   Q   A   S   P   D   G   F   S   H   L   C

GCCGTTTTCCAGCCGCATCAACTACAACTACAAAATGTTGACGGTTATGTTGAGGCGGATGATGATCGATTCACATTCTACTGGAAATA   640
A   V   F   Q   P   H   Q   L   Q   N   V   D   G   Y   V   E   A   D   D   D   R   F   T   F   Y   W   K   Y

TTGCCTTCCTTgtgagttttgttcaataaggagattaagttattcattccaaattcaaaattaaatacagttgaagct   720
C   L   P

cttaaatgtcattaaaaaataaaaatttgtaatttttacaaaatttgtaggcagacatgtgttggtgtcca   800

gtattttgcgcctatgtgtgggagctttctaaagtagaagaaagtctcataaaatattcctcaattctggtttc   880

agagctatgtatgtatatntagttatcattttatctatttttaatttttttattcttgtacttcaaaatcctgatttctgaatgttttttt   960

aatgtgaaatccattgaagccattttttaattttttattttgtacttcaaaatcctgatttctgaatgttttttt   1040

ttatatttataagtcaaactgtttttttttcagCCACTCGAAAAATGTTCTGGAGAATACGCGATTCACATATCTTTCGGAT   1120
                                   S   T   R   K   C   S   G   E   Y   A   F   T   Y   L   S   D

CGTTATATGGATCCCCAAGTCAGTCAGTTATAAAAATGGACAACAACAGCAAATTTAGAAGAGTGTCTTTCGAATTGTTTGGATG   1200
R   Y   M   D   P   Q   V   S   Y   K   M   D   N   D   S   K   F   R   R   V   S   F   G   L   F   G   W
                                   (Cho)

GAAAAAATCATTTGAATGCCGTTCAATCTCCTTCAATAGGAGATCTCTTCAATAGGAGATGTCATATGTCTAAGGATTCACAAATTT   1280
K   K   S   F   E   C   R   S   I   S   F   N   R   T   D   G   G   C   H   M   S   K   D   S   Q   I
```

50b

CACGACCGGAAGCCATTCGATTGAACAATAACCCAACTATCGAATTGATTATTACGAGAATAAACTGTTACAATTgtaag 1360
S   R   P   E   A   I   R   L   N   N   N   P   N   Y   R   I   D   Y   Y   E   N   N   C   Y   N
(Cho)

tgaattggtaaagtaaagtattcattgaagtattccagTATCCGAATCATTCACATTCAAACATGAATGTCGTGATAA 1440
L   S   E   S   F   T   F   K   H   E   C   R   D   N

TGGAATCTCAGTCAGCGTCAAGTCTCGTTCCTTATACTGGAGCCATTATGGACTTTATGGACTTCTTCACTTGTCGGA 1520
G   I   S   V   S   V   K   S   R   L   P   Y   T   G   A   I   Y   G   L   Y   D   F   F   T   C   R

CTGAACCAAAAGAAGCCACTGAATTCGATCACTTTTCCCATATCAGACTGTCAGCAGAACTGTTCGGATTCTATTAAG 1600
T   E   P   K   E   A   T   E   F   D   H   F   F   P   Y   Q   T   V   S   K   N   C   S   D   S   I   K
(Cho)

TACAAGgtaagatcagaaagtgtccaaacacataattgccaaaaatggttaaaaaattacaactatgatcaaaca 1680
Y   K

tttttgaaaagatttcttggtctgaaaaaaccaatttgtttgtgccaatatcgtaggttatgttttgtgtcaaa 1760

ctatcattatttttgcatcaagtcattagggcggctaataagttggacgtattngtcaatttacgttttcctcaaat 1840

ttnatatggtgacttatagtaggcagtgagtatataagcntatcagttcagctcatattttngccagcttcttggtact 1920

ttgatctgaaaaattcaaaatttttgaaattttttgcgatttttgggttaggcgtcagagtttgatattagt 2000

atttaaatcaattaagtttgtttgttgtagattacaaaaattgaatcccgacttgattggcccacacccaaaactttttatag 2080

ccaatcaaagtcaaaagtgaagtgtactaaaaatttgtactaaaaatttgtatcttcctgaaatttcaaactaaaaagtgatacgttta 2160

tccaagtaaacatttataaccaattggtctgtgttccaaacctcctttcgaatttatgagttcaacttttacaagatc 2240

tgggatgttcaaaagaccgttatgctcatttcccaaattatttttccaagaagcatcaaatattataaaattatgtt 2320

gtttctactgtattatctttgtttttcatatggttcaagtgttccaagtgttgtctaccaaggaaaaagatttgaacttttgatcaaaa 2400

aacttaagttttaggaagttagtttgtacaaaaaagcaaaaattagtgctcacgtgaaagaaacaggcttcaatttat 2480

gatatttgatgctttttgtaaaaaataatttgaaaaataagcataacggactttgaacatcccagatcttgtaaaagt 2560

tggaactcataaatcccaacggagcctggaactgaccagacttgttataaatgaattacttggatgacgtatccatt 2640

tttagtttgaaatgttaagaaaagattgaatctgaaaattacaaatttttagtacacctcacttttgactttgattg 2720

gctataaagtttgggtggtgtccaatcgagtcaggattcaatttttgtaatctacaaacaaactaattgatttaga 2800

tattaacatcaaactgtaacgcctaactcaaaaaatcgcaaaaaaaaatttcgaaaaaaatttgaatttttcagatc 2880

aaagtaccgagaagctggctgaaatagctaaaactaataagctgtatatacttactgcatattataagtcaccatata 2960

aaattaaggaaaaccgtaaattgacgaaatacgtcaaaacttattgccgccctaatatattctcatgatatttccaa 3040

aatacaaatttccagGGAAACGAAATGGTACTTGAAGTTGTCTTATCTACTGATGGAATAGAGCCACTTTATTCATTAC 3120
        G N E M V L E V V L S T D G I E P L Y F I T

TCCTGAAGATTTGACATACCAGCCAAAATGTCCAATTAGTGGTGTAATAGCAAAGGATCCCGCTAACAAAATCCTCGG 3200
  P E D L T Y Q A K C P I S G V I A K D P A N I K S S

CTCATTTGGATAACAGgtaaaacttgctagaattattattagttgtttattcattaaattttcactcctaata 3280
A H L D N R

cagaattagttgtttgaatgattgttgatagttagttaaatatgttagAAACAAGGCAATGAAGCATCAGCGCATGCAT 3360
          N K A M E A S A H A

TGTTTGAGTTACTCTCGAAAACTGGAGATGATGAAGCCCTCAAAACACATTCCCACTTCCATTGACTACCACAACAGAA 3440
L F E L L S K T G D D E A L Q N T F P L P T T T T E

GTAATAAGgcaagttagactacaaccaaggagaaccatcaactac----tccaatatttaaacgttttttgtgttcag 3515
V I S

TGATGTGCCAACTACTACCGTACAAACTTCAACTACCGTACCAACCACACCATCCAAAACAACTGCTACCACTACAACAA 3595
D V P T T T V Q T S T T V P T T P S K T T A T T T

CTCCAAAGCCAACCACTACAGAGACAGCTACAACTTCTTCATCAACAACCACAGTAACCACAAAAACCTACAACCGTT 3675
T P K P T T E T A T T S S S T T V T T Q K P T T V

ACGTCAACAACAACATTACCATCCACAACGGCATCTACAACGCTAAAACTACTACAAGTACGCCCACGTCACCACAAAC 3755
T S T T L P S T T A S T T K T T S T P T T S P Q T

AACGACCACTCATGTAGGAGCTCCCGCATCTTCCGTTGCATCGGTGGCTCATGGTGGtaagtattttgtcaa 3835
T T T H V G A P A S S V A S V A H D G

accaagtatattacaacaaggaaaagttatgaaaattcatttttattgtttcaaaaacggcaagctcaacatttn 3915

gttctctaaaacttcaaacttccaatttccagttctccacgagcggcttttttaataaaaactattttaaangt 3995

tatcaaatttctgcaaaatctcaatttaaaaaagtataattcagCTCAACACTTGCCGGAAAACCAAAAGTTCCAGT 4075
                                            S T L A G K P K V P V

GATTTTCGATATTTTCCACAACGGACAACCAGTTGAAGCCGTCGTTGTTGGAACAAAAATCAGTCTATCATTCCGTCCAC 4155
I F D I F H N G Q P V E A V V V G T K I S L S F R P

ACTATCCAATTCCACCGGAGTACGTAGACGTTCGTGGTTGCCAAGTTGAGCCAATTGATCCAAAAATATGAATGGGAGCAC 4235
H Y P I P P E Y V D V R G C Q V E P I D P K Y E W E H

GAGCCACTGTTCATCATAAGAGATGGATGTCCTGCTGACGGTGTTGGTCTCGTCTGCCCACCAACACATTCTGAGTTTGG 4315
E P L F I I R D G C P A D G V L V C P P T H S E F G

AGCAAAAGTTTCCGTTGAGCCATTCAGATATCAAACAACTGGACAAGTACAGTACTCCTGTCTCGTGAGAATATGTCCAT 4395
A K V S V E A F R Y Q T T G Q V Q Y S C L V R I C P

TTGCTCCGTGTCCGAAGgtggttgagcttgatctagtgtctagaaaataatcctttgtcaaacagatggtcgcaagt 4475
F A P C P K

gatctgaattttaaaagttcaaacaaattcagAACACATGCGACGACGTGGAAGGCTGTGACAGTAGCTACACATGCATC 4555
                                N T C D D V E G C D S S Y M H

GATACAGGCGAGAATTATCACTTGAGGACATCAAAAAAGCATTGGAAGCAAACCCAGAACTTGCTTCTCAGTTCGGAATT 4635
R Y R E L S L E D I K K A L E A N P E L A S Q F G I

TCACCGTCCGCGTTGCAAGAAATCCGTCGAAATCTAAAAATTCACAAGCGTTGTTGTggttggttcaaattagtttatt 4715
S P S A F A R N P S K N F T S V V
                                  (Cho)

tccaatccatattaataagttcagAAGAACAACAAAGAATCGCACTCGGCGGCGAGATTATCTTGTTCGGCGTCGATTGA 4795
                        E E Q R I A L G G D Y L V R R R L

TTGTTGTCAACTCTGAAGATCAATTGAGATATTACGTTCGAACTGGAAACATCTGAAAAATACACACAAAAAATGTCTTCA 4875
I V V N S E D Q L R Y Y V R T G N I *

TGTCAGCTTTATAACTTTGTTTTTATTCTTCTTTCATTCAATTATCCAAACGCCACGAGCGGTCTCACATGGAAAAC 4955

TCTTGGAGTTATTCTTCTTATGATTTAAGTTAAATTTTTTGAATTCACCATTTTGATTGATAATAAAATATTTCATCGCt 5035

tgatattattttttgagttt 5054

Figure 21

Diagram of the *let-653* gene. The 10 exons are shown as filled boxes. The arrow

indicates the direction of transcription.

54b

Figure 22

Graph showing the local base concentration of the amino acids serine and threonine for

the Let-653 protein. The graph shows the relative abundance of serine and threonine

within intervals of 19 amino acids. The graph was produced using the program PCGENE

(ver. 6.25). The X-axis represents the amino acid position. The Y-axis represents the

relative abundance.

55b
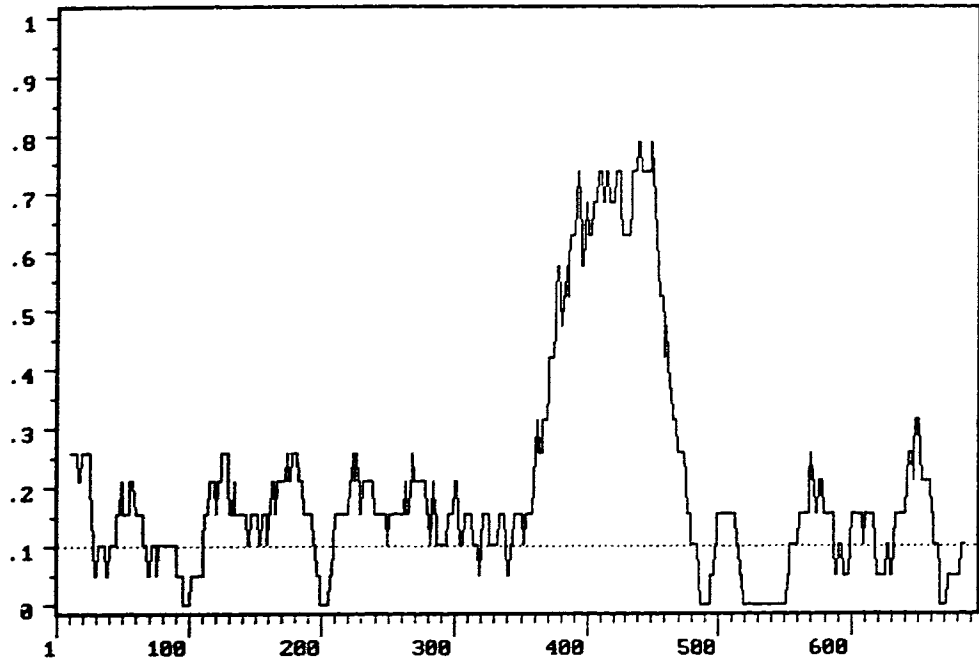
Figure 23

Hydropathy profiles calculated by the algorithm of Kyte and Doolittle (1982) using the

program GREASE (Pearson and Lipman 1988). A window of 19 amino acids was used in

the calculation. Hydropathy profiles of the Let-653 protein, the human MUC-1 mucin

protein (Gendler *et al.* 1990) and the larval glue protein of *Drosophila virilis* (Swida *et al.*

1990). The MUC-1 protein possesses a 31 amino acid transmembranous domain.

CH₃ ... Let-653 Protein ... 685

CH₃ ... MUC-1 mucin ... 615

CH₃ ... Larval glue protein ... 223
LGP-1
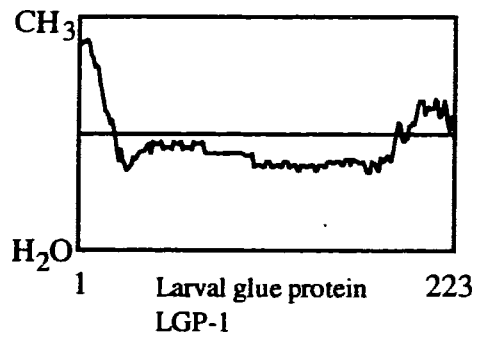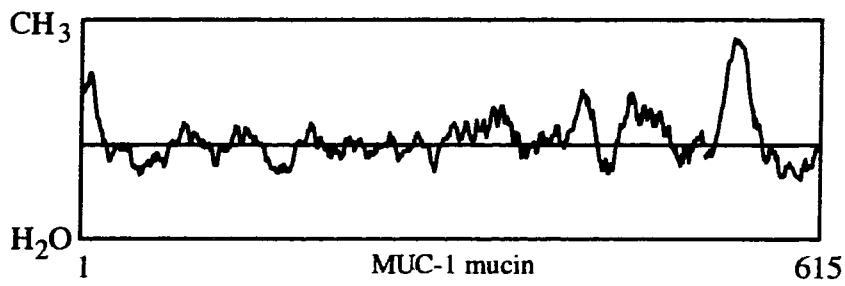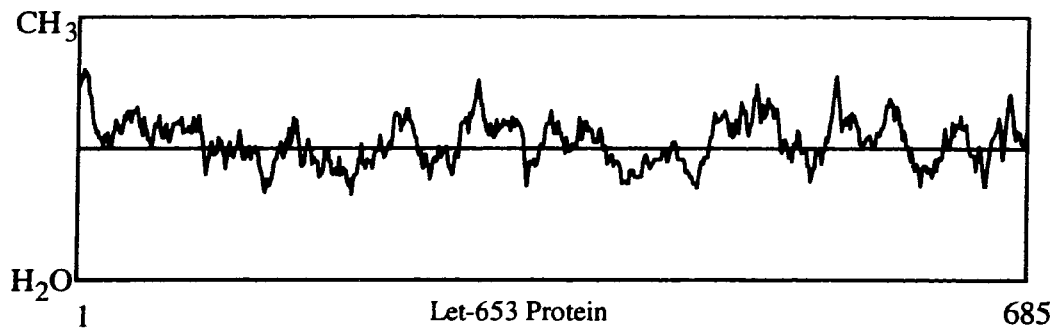
1993) and the frog integumentary mucin C.1 (FIM-C.1) of *Xenopus laevis* (Hauser and Hoffmann 1992) both possess threonine/serine rich regions which do not consist of tandem repeats.

Some mucin-like proteins contain P-domains within the cysteine rich regions. These P-domains consist of highly conserved intragenic repeated cysteine rich modules which are thought to possess growth modulating properties (Hauser and Hoffmann 1992; Thim 1989). The Let-653 protein shows no evidence of repeated sequences within the cysteine rich region and therefore shows no evidence of possessing P-domains.

The physical positioning of *let-660*

In an attempt to rescue *let-660*, *let-660(s1996) Unc-22(s7) unc-31(e169)/nT1(IV);nT1(V)* (BC3506) males were mated to *+/+; sEx26* (BC4440) hermaphrodites. Six *let-660(s1996) unc22(s7) unc-31(e169)/+; sEx26* hermaphrodites were allowed to self. One adult Unc twitcher was isolated. This adult Unc twitcher displayed a leaky adult sterile phenotype and founded the strain BC4465. To determine whether BC4465 represented a rescued strain of *let-660* it was necessary to show the presence of *sEx26* in this strain. This was necessary as the Rol-6 phenotype (which indicates the presence of the extrachromosomal array) is masked within an *unc-22 unc-31* background. To establish whether the extrachromosomal array was present within BC4465 hermaphrodites were mated to *let-660(s1996) unc-22(s7) unc-31(e169) /nT1(IV);nT1(V)* (BC3506) males. In the F1 progeny of this cross no Rol-6 progeny were seen and when the *let-660(s1996) unc-22(s7) unc-31(e169)//nT1(IV);nT1(V)* progeny were allowed to self adult Unc twitchers were seen. The most likely explanation for this observation was that the *let-660(s1996)* bearing chromosome possessed a second mutation in an essential gene which was eliminated through recombination. This

explanation was also corroborated by the fact that when *let-660* is in trans to *sDf7* (figure 1) an adult sterile phenotype is observed (M. Marra, personal communication). This indicated that the second essential gene on the *let-660(s1996)* bearing chromosome lay outside the region covered by the deficiency *sDf7*. These results can be interpreted to mean that *let-660* can now be assigned a leaky adult sterile or egg laying deficient phenotype. Since *let-660(s1996)* homozygotes do show a degree of fertility rescue of *let-660* would be complicated.

Even in the absence of a rescue of *let-660*, this gene can be assigned to the physical map to a resolution of <50kb. This assignment to the physical map is based upon the physical positioning of the *let-660* flanking markers *let-92* and *let-653*. *let-92* is assigned to the cosmid B0033. The *let-653* coding element is assigned to the C46F3 subclones pCes1906 and pCes1940. This means that *let-660* lies within the region encompassed by the cosmid B0033 and the 2.4kb plasmid pCes1907 derived form the cosmid C46F3 (figure 9).

DISCUSSION

In this thesis I have described the positioning of genes *let-92* and *let-660* onto the *C. elegans* physical map. I presented evidence for a maternal requirement for the *let-92* gene product. I have also presented the complete coding sequences of the *let-653* gene and identified sequences containing transcriptional control elements or enhancers for the *let-653* gene.

The major strategy employed in this study was germline transformation using DNA from the physical map. This approach took advantage of the fact that the physical map in the region of study was represented by overlapping cosmids. The DNA from these cosmids can be easily prepared and injected. The success of such an approach is also helped by the size of the cosmid inserts (approximately 40kb) which means that any cosmid has the potential to contain several intact genes. This study showed that plasmid subclones from a rescuing cosmid can also be used for germline transformation rescue even if the complete gene is not present on a single plasmid. In this case as long as overlapping plasmids are present homologous recombination can reform the intact gene during the construction of the extrachromosomal array. This approach of injecting overlapping plasmid subclones has also been successfully utilized in the rescue of *let-56* (S. Jones and M. Marra, unpublished results).

The physical positioning of *let-92*

Extrachromosomal arrays bearing the cosmid B0033 altered the homozygous phenotype of *let-92* animals from a L2 lethal arrest to a maternal effect lethal phenotype. Therefore complete rescue was not achieved. This result, however, does give

an insight into the stages at which the *let-92* gene product is required and suggests that a maternal contribution of the *let-92* gene product is essential for development.

There are two major reasons why B0033 may not be able to rescue *let-92* fully. First, very little is known about the behavior and expression of genes present within an extrachromosomal array. Extrachromosomal arrays may express genes adequately within the somatic cells but poorly within the germline. The overall level of expression of genes in extrachromosomal arrays may also be altered. Therefore, the ability of an extrachromosomal array to rescue a gene may depend upon the number of copies of the gene that are present. In the case of *let-92* the levels of expression may be sufficient to prevent L2 arrest but insufficient to restore fertility within the germline. Secondly, B0033 may not fully rescue *let-92* because of the absence of transcriptional control element or enhancers which are not present on B0033 which are required for the maternal contribution of the *let-92* gene product or message.

The ability of an extrachromosomal array containing B0033 and C38H7 to confer fertility to *let-92* homozygotes infers that a missing transcriptional control element is responsible for the incomplete rescue of *let-92*. C38H7 is thought to contain the maternal effect lethal gene *par-5* (D. Shakes, personal communication). Therefore C38H7 contains at least one transcriptional control element or enhancer that allows germline or maternal expression of the *par-5* gene. However, due to fact that the extrachromosomal array capable of rescuing *let-92* fully could not be maintained this result should be considered at best preliminary.

A rescuing extrachromosomal array converting a larval lethal arrest to a maternal effect lethal phenotype has been previously reported in *C. elegans*. McKay (1993) reported the incomplete rescue of *let-378*. In this case an extrachromosomal array was able to alter the homozygous phenotype of *let-378* from that of a L1/L2 lethal arrest to a maternal effect lethal phenotype whereby eggs were laid but few were able to hatch.

It is not known at present what proportion of essential genes within *C. elegans* are required during development and whose products are also contributed maternally. For essential genes whose products are required at various times during development it is difficult to detect maternal effects through genetic analysis. This is because mutants in essential genes will arrest before adulthood and do not produce progeny. In a study of essential genes in *Drosophila melanogaster* conducted by Eberl *et al.* (1992) the zygotic lethal phenotypes of essential genes within polytene section 17 were examined and their maternal requirements were analyzed in homozygous germline clones using the dominant female sterile technique (Perrimon and Gans 1983). Of 17 essential gene loci studied, 12 were found to be expressed within the germline and display maternal effects. From this data it is not unreasonable to predict that many essential genes in *C. elegans* are also maternally contributed. The maternal requirement for many essential genes would have implications in whether other essential genes can be completely rescued through germline transformation. If the transcriptional control element or enhancer responsible for germline expression is a large distance form the gene or genes it controls then it may not be present within the cosmid sequences injected.

The use of a marker construct containing *rol-6(su1006)* (Kramer *et al.* 1990) proved in this study to be a very useful positive control in the germline transformation experiments. Germline transformation experiments carried out by Clark (1990) without the use of a positive marker failed to rescue *let-92* with the injection of 100ng/µl of the cosmid B0033. The use of a marker construct showed that extrachromosomal arrays were not isolated if B0033 was injected at concentrations greater that 5ng in a total of 100ng DNA per µl. This suggests that a high copy number of B0033 within an extrachromosomal array is lethal. This lethality may be due to the overexpression of a gene that is present within B0033.

par-5, a maternal effect lethal gene involved in the establishment or maintenance of cleavage patterns of the embryo (D. Shakes, personal communication) is adjacent to *let-92*. As both genes display a maternal contribution of their gene products it is possible that *par-5* and *let-92* share a common transcriptional control or enhancer element for this maternal contribution. As both *par-5* and *let-92* are maternally contributed and are in close proximity it is also possible that *par-5* and *let-92* may possess similar functions. However, a precise characterization of the arrested eggs of rescued *let-92* homozygotes has shown that the early embryonic cleavage patterns bear no similarity to the characteristic cleavage patterns of *par-5* homozygotes (D. Shakes, personal communication).

The phenotype of *let-653*

The phenotype of *let-653* is a L1/L2 lethal arrest coincident with the appearance of a vacuole slightly anterior to the lower pharyngeal bulb. The presence and position of this vacuole suggest a dysfunction of the excretory/secretory apparatus (J. Sulston and D. Riddle, personal communication). The secretory/excretory apparatus consists of a binucleate gland cell, a large excretory cell which is connected to the outside of the nematode via the excretory duct and the excretory pore (figure 24). Functions which have been attributed to the nematode excretory/secretory systems include excretion of metabolic waste, osmoregulation (Weinstein 1952; Croll *et al.* 1972; Nelson and Riddle 1984) and secretion of peptidases involved in molting (Singh and Sulston 1978). Water is thought to be continually entering the worm through the cuticle and hypodermis.

Figure 24

A representation of the secretory/excretory apparatus of *C. elegans*. Adapted from Nelson

et al. (1983).

PHARYNX

EXCRETORY
CANALS

EXCRETORY DUCT

EXCRETORY CELL
NUCLEUS

EXCRETORY GLANDS

63b
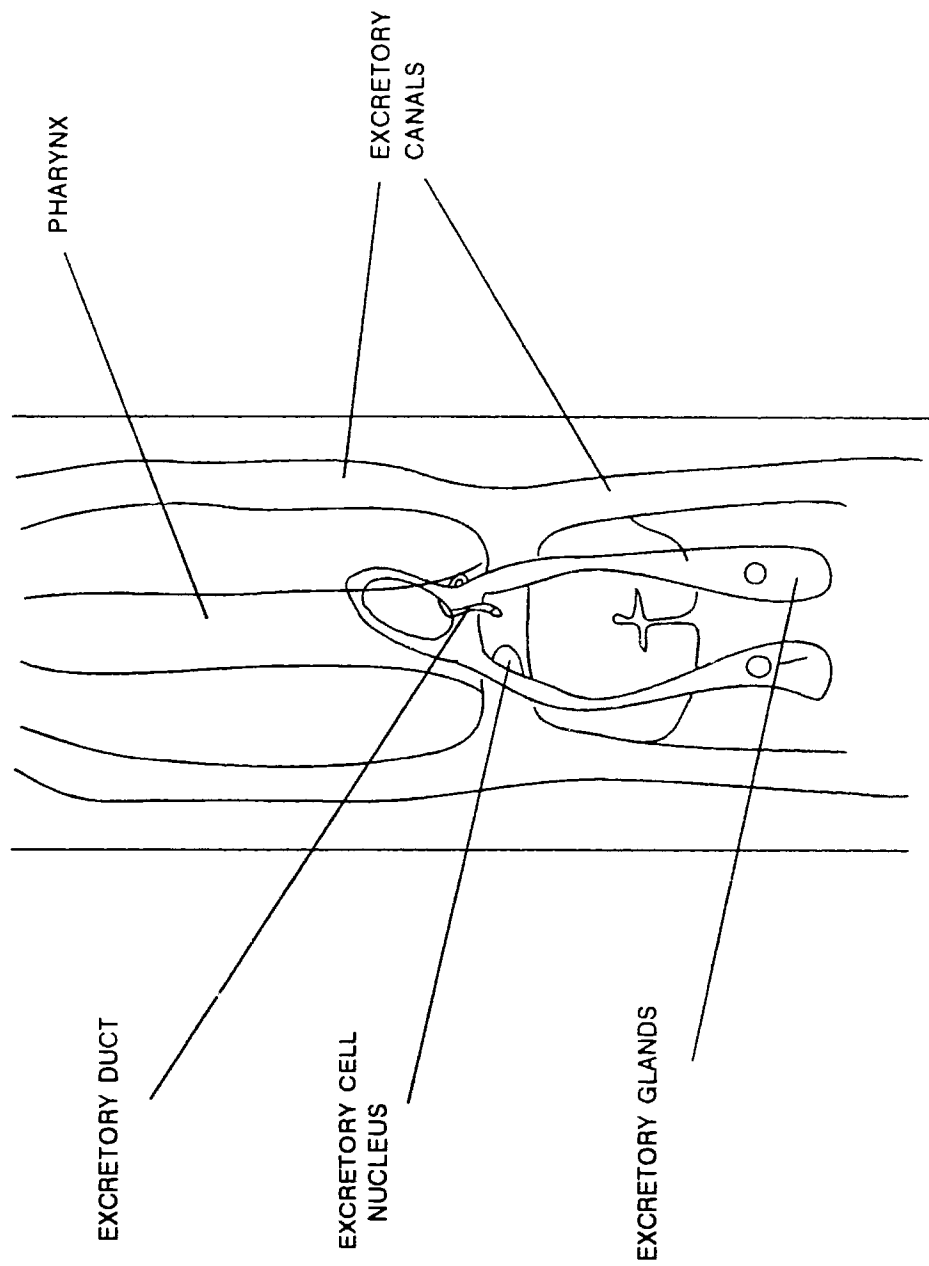
The internal hydrostatic pressure and osmotic potential within the nematode may be regulated by the flow of water through the secretory/excretory sinuses. Therefore a vacuole in this region is consistent with *let-653* homozygotes possessing dysfunctional excretory/secretory apparatus and/or their inability to osmoregulate adequately. This inference is corroborated with laser ablation experiments carried out by Nelson and Riddle (1984). In these experiments it was shown that laser ablation of the gland cell seemed to confer no obvious developmental or behavioral effects, however, the ablation of the excretory cell, its pore or its duct resulted in the animals filling with fluid and death occurring within 12-24 hours. The first indication of this fluid accumulation was the swelling of the excretory cell, followed by an increase in this swelling and later large zones of fluid filling the hypodermis. This is entirely consistent with the phenotype of *let-653(s1733)* where, in a small proportion of homozygotes, the vacuole covers large zones of the hypodermis (figure 7F). Further evidence to suggest dysfunction of *let-653* secretory/excretory apparatus is displayed in *lag-1*, *lag-2* and *lin-12;glp-1* mutants where the excretory cell and excretory duct is absent (Lambie and Kimble, 1991) and a small vacuole or swelling is observed at the normal location of the excretory pore (E. Lambie personal communication). Interestingly individuals possessing *let-60* and *lin-3* loss of function alleles also display a similar vacuole (M. Marra, personal communication).

The *let-653* gene encodes a mucin-like protein

Mucins are normally secreted as mucus in order to present a selective barrier between living cells and their environment. This protective function of mucin is provided by its viscous and rheological properties. These properties of mucin are determined to a great extent by their ability to form intermolecular disulfide bridges to form oligomeric mucin and by the presence of amino acid sequences containing

64

attachment sites for O-linked oligosacharrides. Mucins typically show heavy O-linked glycosylation whereby 50% of the dry weight of many mucins consist of carbohydrate chains.

Mucins have been invoked in the pathology of many diseases. In cystic fibrosis the rheological properties of mucin change as it undergoes an increased level of glycosylation (Cheng *et al.* 1989). The link between the defective Cl⁻ transport activity in cystic fibrosis and this increase in glycosylation is not understood (Anderson *et al.* 1991). In many human carcinomas production of the mucin protein MUC-1 is increased and the glycosylation pattern is altered (Zotter *et al.* 1988). Mucins have also been implicated in the metastatic potential of carcinoma cells. In this case the protective properties of mucin may be reversed as they are thought to interfere with the immune surveillance by T-lymphocytes by causing steric hindrance of surface antigen presentation.

To propose a function for the Let-653 protein in *C. elegans* would be speculative at this point. However some correlation between the protein and phenotype can be made. Mucins are secreted in order to protect regions where cells come into contact with the outside environment. The excretory duct and excretory cell represent a scenario where the protective properties of mucin could be utilized. The mucin-like Let-653 protein within the secretory/excretory apparatus could provide a protective coating to prevent the influx of outside water into the hypertonic excretory/secretory apparatus. Dysfunction or absence of this mucin-like protein may allow the influx of water into the excretory/secretory apparatus resulting in the characteristic vacuole of the *let-653* homozygote and the subsequent breakdown of osmoregulation within the nematode.

The molecular function of the cysteine rich regions of mucin has not yet been fully established. It is very likely that these represent functional domains which play a role in the oligomerization of mucins by the formation of disulfide bridges (Dekker *et al.* 1991). If both the cysteine rich regions represent functional domains of the Let-653

protein then it is conceivable that complementing alleles of *let-653* could be isolated. This has been demonstrated by Marco Marra who has been able to isolate complementing alleles of *let-653*. Even if the role of the cysteine rich regions is purely for oligomerization a mechanism for complementing alleles of *let-653* can be presented. If the oligomerization of mucin monomers takes place only between heterologous cysteine rich domains, then complementing alleles whereby a different cysteine rich domain is mutated in each case would be able to form dimers. In contrast, in an individual where both alleles possess a mutation in the same cysteine rich domain no dimerization could take place. Studies have shown that human cervical, human tracheo-bronchial and rat gastric mucins consists of populations which are 1, 2 and 3 times the length of the monomeric subunit (Carlstedt and Sheenan 1989; Thornton *et al.* 1990). Therefore even if complementing alleles are able only to make mucin dimers the essential function of the mucin may be restored.

The control of *let-653* expression.

The requirement of an upstream regulatory element for appropriate transcription of *let-653* would not be unique in *C. elegans*. The necessity for the presence of upstream regulatory elements has been shown for several *C. elegans* genes. Okkema *et al.* (1993) demonstrated the existence of transcriptional enhancers for the genes *myo-1*, *myo-2*, *myo-3* and *unc-54*. In the case of *myo-3* an enhancer element was found to exist between 1 and 1.5kb upstream of the initiating methionine. In the gene *mec-3* two transcriptional enhancer elements exist within the 400bp of sequence upstream of the initiating methionine (Way *et al.* 1991). In the vitellogenin gene of *C. elegans*, *vit-2*, 7 distinct transcriptional enhancer sequences have been elucidated within the 250bp upstream of the promoter (MacMorris *et al.* 1992). An analysis of the upstream elements

66

of the *C. elegans* gut esterase gene, *ges-1*, showed that the 1.3kb of DNA upstream of the initiating methionine contained 3 transcriptional activator elements and 4 transcriptional repressor elements (Aamodt *et al.* 1991).

The rescue of *let-653* required either pCes1907 or pCes1903 plasmid sequences to be present within the extrachromosomal array along with pCes1940 and pCes1906. This result predicts that pCes1907 and pCes1903 both possess transcriptional control elements with an interchangeable or redundant function. In contrast to the examples of transcriptional enhancers above, pCes1907 and pCes1903 were not contiguous with the *let-653* coding element when injected. This means that there may be a great deal of variability in the distance between the transcriptional control element and the *let-653* promoter within the extrachromosomal array. This is a feature consistent with the behavior of transcriptional control elements or enhancers.

The ability of both pCes1903 and pCes1907 to control *let-653* gene expression is interesting and deserves discussion. One possible explanation is that *let-653* possesses both upstream and downstream regulatory or control elements which display some redundancy. It is also possible that one of the constructs, pCes1903 or pCes1907, would not normally play a role in *let-653* expression. Instead it may normally control the expression of an adjacent or nearby gene but within the extrachromosomal array it is able to adequately control expression of *let-653* to allow rescue. This latter explanation is supported by a study by Okkema *et al.* (1993). In this work a screen was carried out for enhancers of the *C. elegans myo-2* gene. In all 18 DNA fragments randomly derived from the genome were studied representing a total of 62kb of genomic DNA. 6 of the 18 fragments tested functioned as enhancers and all of the expression was observed in muscle cells. However when these enhancers were used to express the *glp-1* gene expression in non-muscle cells was observed. These results indicate that genomic DNA is relatively abundant in enhancer sequences and that much of the specificity of expression

can remain at the promoter. From these results we can predict that many enhancers in *C. elegans* can be interchanged with relatively little effect upon the expression patterns of their target genes.

In this study ESTs have proven to be useful in determining an embryonic expression for the *let-653* transcript. The fact that *let-653* is known to be expressed embryonically but the blocking stage of *let-653(s1733)* homozygotes is L1/L2 suggests one of two things. Either the *let-653* gene product is not essential for development and has no phenotype until the blocking stage is reached. or a maternal contribution of *let-653* gene product is sufficient for development until the blocking stage is reached.

Genomic organization within the *unc-22* region

Prasad and Baillie (1989) suggested that the *unc-22* region may be transcriptionally synchronized. This inference was based upon some of the coding elements in this region showing a maximal expression at the L2 stage. This present study suggests another way in which we may describe the organization of the *unc-22* region, since it may contain those genes for which a maternal contribution is required. *par-5* and *let-92*, which are adjacent, both show a maternal contribution of their products. Due to the disparity between the *let-653* embryonic expression and the L1/L2 blocking stage this presents evidence that the *let-653* gene product may also be maternally contributed. It is possible that many of the essential genes within the *unc-22* region may also display a maternal contribution which classical genetic analysis would be unable to detect.

CONCLUSION

The genetic and physical correlation of the genes *let-653*. *let-660* and

let-92 has been achieved. The order of these genes on the physical map has been shown

to be consistent with the order predicted from the genetic map. The incomplete rescue of

let-92 has demonstrated that extrachromosomal arrays can be used to predict a maternal

contribution for a gene product.

The plasmid rescue of *let-653* has shown that homologous recombination

can be exploited in the construction of extrachromosomal arrays to reconstitute complete

genes from overlapping sequences. In order for the plasmid rescue of *let-653* to be

achieved sequences either 5' or 3' to the *let-653* gene were required to be present within

the extrachromosomal array. It is likely that these sequences contain transcriptional

control elements which are capable of allowing the expression of the *let-653* gene. The

observation that two sequences have the ability to allow *let-653* expression can be

interpreted in two ways. Either the *let-653* gene has two transcriptional control sites

which display some redundancy or that a transcriptional control element from another

gene is able to allow the expression of *let-653* at levels adequate to allow rescue.

The *let-653* gene has been found to encode a mucin-like protein. The

Let-653 protein displays the characteristic features of mucins, possessing a

serine/threonine rich domain, two cysteine rich domains and a hydrophobic leader

sequence.

let-653 homozygotes have been determined to undergo a lethal arrest within

the L1 or L2 larval stages. Concurrent with this lethal arrest is the appearance of a

vacuole anterior to the lower pharyngeal bulb. This vacuole is thought to be caused by the

dysfunction of the secretory/excretory apparatus. This dysfunction may be caused by the

absence of functional mucin within the excretory/secretory sinuses which allows the influx of water into the nematode and interferes with its ability to osmoregulate.

REFERENCES

Aamodt, E.J., Chung, M.A. and McGee, J. (1991). Spatial control of a gut-specific esterase during *Caenorhabditis elegans* development. Science 252: 579-582.

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R., Kerlavage, A.R., McCombie, W.R. and Venter J.C. (1991). Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project. Science 252: 1651-1656.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215(3): 403-410.

Anderson, M.P., Rich, D.P., Gregory, R.J., Smith, A.E. and Welsh, M.J. (1991). Generation of cAMP-activated chloride currents by expression of CFTR. Science 251: 679.

Barstead, R.J. and Waterston, R.H. (1989). The Basal component of the nematode dense-body is vinculin. J. Biol. Chem. 114: 715-724.

Benian, G.M., Kiff, J.E., Neckelmann, N., Moerman, D.G. and Waterston, R.H. (1989). Sequence of an unusually large protein implicated in the regulation of myosin activety in *Caenorhabditis elegans*. Nature 342: 45-50.

Blobel, G. and Dobberstein, B. (1975). Transfer of proteins across membranes. J. Cell. Biol. 67: 835-851.

Blumental, T. and Thomas, J. (1988). *Cis* and *trans* mRNA splicing in C. elegans. Trends in Genetics 4: 305-309.

Brenner, S. (1974) The genetics of *Caenorhabditis elegans*. Genetics 115: 71-94.

Burke, D.T., Carle, G.F., and Olson, M.V. (1987). Cloning large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236: 806-812.

Cabot, E.L. and Beckenbach, A.T. (1989). Simultaneous editing of multiple nucleic acid sequences with ESEE. Comp. Appl. Biol. Sci. 5: 233-234.

Carlstedt, I. and Sheehan, J.K. (1989). Structure and macromolecular properties of cervical mucus glycoproteins, in Mucus and Related Topics, Chantler, E. and Ratcliffe, N.A., Eds., University of Cambridge; Symp. Soc. Exp. Biol. 43: 289.

Charest, D.L., Clark, D.V., Green, M.E. and Baillie D.L. (1990). Genetic and fine structure analysis of *unc-26(IV)* and adjacent regions in *Caenorhabditis elegans*. Mol. Gen. Genet. 221: 459-465.

Charron, J., Malynn, B.A., Fisher, P., Stewart, V., Jeanotte, L., Goff, S.P., Robertston, E.J. and Alt, F.W. (1992). Embryonic lethality in mice homozygous for a targeted disruption of the N-myc gene. Genes and Dev. 6(12): 2248-2257.

Cheng, P.W., Boat, T.F., Cranfill, K., Yankaskas, J.R. and Boucher (1989). Increased sulfation of glycoconjugates by cultured nasal epithelial cells from patients with cystic fibrosis. J. Clin. Invest. 84: 68.

Clark, D.V. (1990). The *unc-22(IV)* region of *Caenorhabditis elegans*: genetic analysis and molecular mapping. PhD Thesis, Simon Fraser University, Burnaby, B.C., Canada.

Clark, D. V. and Baillie, D. L. (1992) Genetic analysis and complementation by germline transformation of lethal mutations in the *unc-22(IV)* region of *Caenorhabditis elegans*. Mol. Gen. Genet. 232: 97-105.

Clark, D.V., Rogalski, T.M., Donati, L.M. and Baillie, D.L. (1988). The *unc-22(IV)* region of *Caenorhabditis elegans*: genetic analysis of lethal mutations. Genetics 119: 345-353.

Coulondre, C. and Miller, J.H. (1977). Genetic studies of the lac repressor; IV. Mutagenic specificty in the lacI gene of *E. Coli*. J. Mol. Biol. 117: 577-606.

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

Coulson, A., Waterson, R., Kiff, H., Sulston, J.E. and Kohara, Y. (1988). Genomic linking with yeast artificial chromosomes. Nature 335: 184-186.

Coulson, A., Kozono, Y., Lutterbach, B., Shownkeen, R, Sulston, J. and Waterston, R. (1991). YACs and the *C. elegans* genome. BioEssays 13: 413-417.

Croll, N.A., Slater, L. and Smith J.M. (1972). *Ancylostoma tubaeforme*: Osmoregulatory ampulla of larvae. Exp. Parasitol. 33: 356-360.

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

Dekker, J., Aelmans, P.H. and Strous, G.J. (1991). The oligomeric structure of rat and human gastric mucins. Biochem. J. 277: 423.

Diehl, B.E. and Pringle, J. (1991). Molecular analysis of *Saccharomyces cerevisiae* chromosome I: identification of transcribed regions and demonstration that some encode essential functions. Genetics 127: 287-298.

Doudney, C.O. (1976). Mutation in ultraviolet light - damaged micro-organisms in S.Y. Wang (Ed.), Photochemistry and Photobiology of nucleic acids; Vol II, Academic press.

Eberl, D.F., Perkins, L.A., Engelstein, M., Hilliker, A.J. and Perrimon, N. (1992). Genetic and developmental analysis of polytene section 17 of the X chromosome of *Drosophila melanogaster*. Genetics 130: 569-583.

Eisenberg, D., Schwarz, E., Komarony, M. and Wall, R. (1984). Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J. Mol. Biol. 179: 125-142.

Feinberg, A. and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activety. Anal. Biochem. 132: 6-13.

Ferguson, E. L. and Horvitz, H. R. (1985). Identification and characterization of 22 genes that effect the vuval lineages of the nematode *Caenorhabditis elegans*. Genetics 110: 17-72.

Gendler, S.J., Lancaster, C.A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E. and Wilson, D. (1990). Molecular cloning and expression of Human tumor-associated polymorphic epithelial mucin. J. Biol. Chem. 265: 15286-15293.

Han, M. and Sternberg, P.W. (1990). *let-60*, a gene that specifies cell fates using *C. elegans* vulval induction, encodes a *ras* protein. Cell 63: 921-931.

Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. 166: 557

Hartl, D.L., Ajioka, J.W., Cai, H., Lohe A.R., Lozovskaya, E.R., Smoller, D.A. and Duncan I.W. (1992). Towards a *Drosophila* genome map. Trends in genetics 8. 70-75.

Hauser, F. and Werner, H. (1992). P-domains as shuffled cysteine-rich modules in integumentary mucin C.1 (FIM-C.1) from *Xenopus laevis*. J. Biol. Chem. 267: 24620-24624.

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

Horvitz, H. R., Brenner, S., Hodgkin, J. and Herman, R. K. (1979). A uniform genetic nomenclature for the nematode *Caenorhabditis elegans*. Mol. Gen. Genet. 175: 129-133.

Howell, A.M., Gilmour, S.G., Mancebo, R.A. and Rose, A.M. (1987). Genetic analysis of a large autosomal region of chromosome I in *Caenorhabditis elegans*. Genet. Res. Camb. 49:207-213.

Howell, A.M. and Rose A.M. (1990). Essential genes in the *hDf6* region of chromosome I in *Caenorhabditis elegans*. Genetics 126: 583-592.

Ish-Horowicz, D. and Burke, J. F. (1981). Rapid and efficient cosmid cloning. Nucleic Acids Res. 9: 2989-2998.

Jackon, I. (1991). A reappraisal of non-consensus mRNA splice sites. Nucleic Acids Res. 19: 3795-3798.

Johnsen, R.C. and Baillie, D.L. (1991). Genetic analysis of a major segment [LGV(left)] of the genome of *Caenorhabditis elegans*. Genetics 129: 735-752.

Juriloff, D.M., Harris, M.J., Wong, V. and Miller, J.E. (1992). Studies of a spontaneous lethal mutation at the albino locus in SELH/Bc mice. Genome 35: 342-346.

Killeen, N., Barclay, A.N., Willis, A.C. and Williams, A.F. (1987). The sequence of rat leukosialin (W3/13 antigen) reveals a molecule with O-linked glycosylation of one thired of its extracellular amino acids. EMBO J. 6:4029-4033.

Kimble, J.E., and Hirsh, D.I. (1979). The post-embryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*. Dev. Biol. 70: 396-417.

Kohara, Y., Akiyama, K. and Isono, K. (1987). The physical map of the whole *E. Coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell 50: 495-508.

Kramer, J. M., French, E., Park, E. and Johnson, J. J. (1990). The *Caenorhabditis elegans rol-6* gene, which interacts with the *sqt-1* collagen gene to determine organismal morphology, encodes a collagen. Mol. Cell Biol. 10: 2081-2089.

Krause, M. and Hirsh, D.L. (1987). A *trans*-spliced leader sequence on actin mRNA in *C. elegans*. Cell 49: 753-761.

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

Li, E., Sucov, H.M., Lee, K.F., Evans, R.M. and Jaenisch, R. (1993). Normal development and growth of mice carrying a targeted disruption of the alpha 1 retinoic acid receptor gene. Proc. Natl. Acad. Sci. USA 15: 1590-1594.

Ligtenberg, M.J.L., Vos, H.L., Gennissen, A.M.C. and Hilkens, J. (1990). Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene encoding splice variants with alternative amino termini. J. Biol. Chem. 265: 5573-5576.

MacMorris, M., Broverman, S., Greenspoon, S., Lea, K., Madej, C., Blumenthal, T. and Spieth, J. (1992). Regulation of vitellogenin gene expression in transgenic *Caenorhabditis elegans* - short sequences required for activation of the *vit-2* promoter. Mol. Cell. Biol. 12: 1652-1662.

Marra, M.A., Prasad, S.S. and Baillie, D.L. (1993). Molecular analysis of two genes between *let-653* and *let-56* in the *unc-22*(IV) region of *Caenorhabditis elegans*. Mol. Gen. Genet. 236: 289-298.

McKay, S. (1993). High resolution alignment of the physical and genetic maps of the *dpy-14* region of chromosome I in *Caenorhabditis elegans*. MSc Thesis, University of British Columbia, B.C., Canada.

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

Meneely, P.M. and Herman, R.K. (1979). Lethals, steriles and deficiencies in a region of the X chromosome of *Caenorhabditis elegans*. Genetics 92: 99-115.

Meneely, P.M. and Herman, R.K. (1981). Suppression and function of X-linked lethal and sterile mutations in *Caenorhabditis elegans*. Genetics 97: 65-84.

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

Moerman, D.G. and Baillie, D.L. (1981). Formaldehyde mutagenesis in the nematode *Caenorhabditis elegans*. Mut. Res. 80: 273-279.

Nelson, F.K. and Riddle, D.L. (1984). Functional study of the *Caenorhabditis elegans* secretory/excretory system using laser microsurgery. J. Exp. Zoo. 231: 45-56.

Nelson, F.K., Albert, P.S. and Riddle, D.L. (1983). Fine structure of the *Caenorhabditis elegans* secretory-excretory system. J. Ultrastruc. Res. 82: 156-171.

O'Donnel, J., Mandel, H.C., Krauss, M. and Sofer, W. (1977). Genetic and cytogenetic analysis of the Adh region in *Drosophila melanogaster*. Genetics 86: 853-867.

Okkema, P.G., Harrison, S.W., Plunger, V., Aryana, A. and Fire, A. (1993). Sequence requirements for myosin gene expression and regulation in *Caenorhabditis elegans*. Genetics 135: 385-404.

Oliver, S.G., Van der Aart, Q.J.M., Agostini-Carbone, M.L., Aigle, M., Alberghina, L., Alexandraki, D., Antoine, G., Anwar, R., Ballesta, J.P.G., Benit, P., Berben, G., Bergantino, E., Biteau, N., Bolle, P.A., Bolotin-Fukuhara, M., Brown, A., Brown, A.J.P., Buhler, J.M., Carcano, C., Carignani, G., Cederberg, H., Chanet, R., Contreras, R., Crouzet, M., Daignan-Fornier, B., Defoor, E., Delgado, M., Demolder, J., Doira, C., Dubois, E., Dujon, B., Dusterhoft, A., Erdmann, D., Esteban, M., Fabre, F., Fairhead, C., Faye, G., Feldmann, H., Fiers, W., Francinques-Gaillard, M.C., Franco, L., Frontali, L.,

Fukuhara, H., Fuller, L.J., Galland, P., Gent, M.E., Gigot, D., Gilliquet, V., Glansdorff, N., Goffeau, A., Grenson, M., Grisanti, P., Grivell, L.A., Dehaan, M., Haasemann, M., Hatat, D., Hoenicka, J., Hegemann, J., Herbert, C.J., Hilger, F., Hohmann, S., Hollenberg, C.P., Huse, K., Iborra, F., Indige, K.J., Isono, K., Jacq, C., Jacquet, M., James, C.M., Jauniaux, J.C., Jia, Y., Jimenez, A., Kelly, A., Kleinhans, U., Kreisl, P., Lanfranchi, G., Lewis, C., van der Linden, C.G., Lucchini, G., Lutzenkirchen, K., Maat, M.J., Mallet, L., Mannhaupt, G., Martegani, E., Mathieu, A., Maurer, C.T.C., McConnell, D., McKee, R.A., Messenguy, F., Mewes, H.W., Molemans, F., Montague, M.A., Muzi Falconi, M., Navas, L., Newion, C.S.,Philippsen, P., Pierard, A., Planta, R.J., Plevani, P., Poetsch, B., Pohl, F., Purnelle, B., Ramezani Rad, M., Rasmussen, S.W., Raynal, A., Remacha, M., Richterich, P., Roberts, A.B., Rodriguez, F., Sanz, E., Schaaff-Gerstenschlager, I., Scherens, B., Schweitzer, B., Shu, Y., Skala, J., Slonimski, P.P., Sor, F., Soustelle, C., Spiegelberg, R., Stateva, L.I., Steensma, H.Y., Steiner, S., Thierry, A., Thireos, G., Tzermia, M., Urrestarazu, L.A., Valle, G., Vetter, L., van Vliet-Reedijk, J.C., Voet, M., Volckaert, G., Vreken, P., Wang, H., Warmington, J.R., von Wettstein, D., Wicksteed, B.L., Wilson, C., Wurst, H., Xu, G., Yoshikawa, A., Zimmermann, F.K. and Sgouros, J.G. (1992). The complete DNA sequence of yeast chromosome III. Nature 357: 38-46.

Olson, M., Hood, L., Cantor, C. and Botstein, D. (1989). A common language for physical mapping of the human genome. Science 245: 1434.

Ouellette, B.F.F., Clark, M.W., Keng, T., Storms, R.K., Zhong, W., Zeng, B., Fortin, N., Delaney, S., Barton, A., Kaback, D.B. and Bussey, H. (1993). Sequencing of chromosome I from *Saccharomyces cerevisiae*: analysis of a 32kb region between *LTE1* and *SPO7* genes. Genome 36: 32-42.

Pearson, W.R. and Lipman, D.J. (1988). Improved tools for biological sequence analysis. Proc. Natl. Acad. Sci. USA 85: 2444-2448.

Perrimon, N. and Gans, M. (1983). Clonal analysis of the tissue specificity of recessive female-sterile mutations of *Drosophila melanogaster* using a dominant female sterile mutation. Dev. Biol. 100: 365-373.

Plasterk, R.H.A. and Groenen, J.I.M. (1992). Targeted alterations of the *Caenorhabditis elegans* genome by transgene instructed DNA double strand break repair following Tc1 excision. EMBO J. 11: 287-290.

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

Rao, M.J.K. and Argos, P. (1988). A conformational preference parameter to predict helices in integral membrane proteins. Biochim. Biophys. Acta 869: 197-214.

Riles, L., Dutchik, J.E., Baktha, A., McCauley, B.K., Thayer, E.C., Leckie, M.P., Braden, V.V., Depke, J.E. and Olson, M. (1993). Physical Maps of the Six Smallest Chromosomes of *Saccharomyces cerevisiae* at a resolution of 2.6 Kilobase pairs. Genetics 134: 81-150.

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

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

Rose, A.M. and Baillie, D.L. (1980). Genetic organization of the region around *unc-15(I)*, a gene affecting paramyosin in *Caenorhabditis elegans*. Genetics 96: 639-648.

Rosenbluth, R.E., Rogalski, T.M., Johnsen, R.C., Addison, L.M. and Baillie, D.L. (1988). Genomic organization in *Caenorhabditis elegans*: deficiency mapping on linkage group V (left). Genet. Res. 52: 105-118.

Rosenbluth, R.E., Cuddeford, C. and Baillie, D.L. (1985). Mutagenesis in *Caenorhabditis elegans*. II A spectrum of mutational events induced with 1500R of gamma radiation. Genetics 109: 493-511.

Rothstein, R.J. (1983). One-step gene disruption in Yeast. Methods Enzymol. 101: 202-211.

Russell, R.J., Healy, M.J. and Oakeshott, J.G. (1992). Molecular analysis of the lethal (1) B214 region at the base of the X chromosome of *Drosophila melanogaster*. Chromosoma 101(7): 456-466.

Rychlik, W. and Rhoads, R.E. (1989). A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in-vitro amplification of DNA. Nucleic Acids Res. 17: 8543-8551.

Sambrook, J., Fritsch, E. F. and Sambrook, J. (1989). Molecular cloning: a laboratory manual. Cold Spring Harbor, New York.

Sigurdson, D.C., Spanier, G.J. and Herman, R.K. (1984). *Caenorhabditis elegans* deficiency mapping. Genetics 108: 331-345.

Singh, S.S. and Sulston, J.E. (1978). Some observations of moulting in *Caenorhabditis elegans*. Nematologica 24: 63-71.

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.

Starr, T., Howell, A.M., McDowall, J., Peters, K. and Rose, A.M. (1989). Isolation and mapping of DNA probes within the linkage group I gene cluster of *Caenorhabditis elegans*. Genome 32: 365-372.

Stewart, H.I., Rosenbluth, R.E. and Baillie, D.L. (1991). Most ultraviolet induced mutations in the nematode *Caenorhabditis elegans* are chromosomal rearrangments. Mut. Res. 249: 37-54.

Strous, G.J. and Dekkler, J. (1992). Mucin-type glycoproteins. Crit. Rev. Biochem. Mol. Biol. 27:57-92.

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

Sulston J.E., Schierenberg, E., White, J.G. and Thompson, J.N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. Dev. Biol. 100: 64-119.

Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierry-Mieg, J., Qui, L., Dear, S., Coulson, A., Craxton, M., Durbin, R., Berks, M., Metzstein, M., Hawkins, T., Ainscough, R. and Waterston R. (1992). The *C. elegans* sequencing project: a beginning. Nature 356: 37-41.

Swida, U., Lucka, L. and Kress, H. (1990). Glue protein genes in *Drosophila virilis*: their organization, developmental control of transcription and specific mRNA degradation. Development 108: 269-280.

Thim, L. (1989). A new family of growth factor-like peptides. FEBS lett. 250: 85-90.

Thornton, D.J., Davies, J.R., Kraayenbrink, M., Richardson, P.S., Sheehan, J.K. and Carlstedt, I. (1990). Mucus glycoproteins from normal human tracheobronchial secretion. Biochem. J. 265: 179.

Verma, M. and Davidson, E.A. (1993). Molecular cloning and sequencing of a canine tracheobronchial mucin cDNA containing a cysteine-rich domain. Proc. Natl. Acad. Sci. USA 90: 7144-7148.

Waterston, R., Martin, C., Craxton, M., Huynh, C., Coulson, A., Hillier, L., Durbin, R., Green, P., Shownkeen, R., Halloran, N., Metzstein, M., Hawkins, T., Wilson, R., Berks, M., Du, Z., Thomas, K., Thierry-Mieg, J. and Sulston J. (1992). A survey of expressed genes in *Caenorhabditis elegans*. Nature Genetics 1: 114-123.

Way, J.C., Wang, L., Run, J.-Q. and Wang, A. (1991). The *mec-3* gene contains cis-acting elements mediating positive and negative regulation in cells produced by asymmetric cell division in *Caenorhabditis elegans*. Cell 54: 5-16.

Weber, J.D. and May, P.E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44: 388-396.

Weinstein, P.P. (1952). Regulation of water balance as a function of the excretory system of the filariform larvae of *Nippostrongylus muris* and *Ancylostoma canium*. Exp. Parasitol. 1:363-376.

White, J.G., Southgate, E., Thompson, J.N. and Brenner, S. (1986). The structure of the nervous system of the nematode *C. elegans*. Phil. Trans. R. Soc. 314: 1-340.

Zotter, S., Hageman, P.C., Lossnitzer, A., van-den-Tweel, J., Hilkens, J., Mooi, W.W. and Hilgers, J. (1988). Monoclonal antibodies to epithelial sialomucins recognize epitopes at different cellular sites in adenolymphomas of the parotid gland. Int. J. Cancer. 3: 38-44.