DEFICIENCY MAPPING AND CORRELATION OF THE GENETIC AND PHYSICAL MAPS IN THE *UNC-22(IV)* REGION OF *CAENORHABDITIS ELEGANS*

by

Jacqueline Else Schein

B.Sc., Simon Fraser University, 1991

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the Department

of

Biological Sciences

SIMON FRASER UNIVERSITY

October 1994

APPROVAL

Name: **Jacqueline Schein**

Degree: **Master of Science**

Title of Thesis:

DEFICIENCY MAPPING AND CORRELATION OF THE GENETIC AND PHYSICAL MAPS IN THE UNC-22 (IV) REGION OF CAENORHABDITIS ELEGANS

Examining Committee:

Chair: Dr. R. C. Ydenberg, Associate Professor

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

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

Dr. J. V. Price, Assistant Professor
Department of Biological Sciences, SFU

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

Date Approved Oct 28, 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

Deficiency Mapping and Correlation of the Genetic and Physical Maps in the unc-22(IV) Region of Caenorhabditis elegans

Author: _____
(signature)

Jacqueline Else Schein
(name)

November 1, 1994
(date)

ABSTRACT

In this thesis, the physical arrangement of genes in the *unc-22(IV)* region of the nematode *Caenorhabditis elegans* was studied using the methods of deficiency mapping and germline transformation rescue.

In Part 1, I describe the isolation and characterization of 22 deficiencies affecting the *unc-22* gene. These deficiencies were recovered following treatment with formaldehyde or 254-nm ultraviolet light. The extents of these deficiencies were determined by either standard complementation tests, or, in the case of two of the deficiencies, by a polymerase chain reaction based method. This analysis has resulted in the unambiguous placement of eight genes on the genetic map, and shown that of eight genes which have thus far been molecularly identified in the *let-56* - *unc-22* interval, at least five and as many as seven are not required for viability under laboratory conditions. Analysis of the breakpoints of the 22 new deficiencies, and of 15 previously identified *unc-22* deficiencies, indicate that the *unc-22* - *let-52* interval is a hot spot for *unc-22* deficiency breakpoints.

In Part 2, I describe the systematic correlation of the genetic and physical maps immediately to the right of the *unc-22* gene. This was accomplished by the technique of germline transformation using cosmid clones from the physical map. Three genetic loci, *let-52*, *let-661* and *let-93*, were positioned on the physical map using this procedure. Analysis of these data indicates an average of one essential gene per every two cosmids in the region immediately to the right of the *unc-22* gene, and suggests that the clustering of deficiency breakpoints between *unc-22* and *let-52* is not due to a large physical distance separating the two loci.

To my parents.

ACKNOWLEDGMENTS

I would like to thank my supervisor, Dr. David L. Baillie, for the opportunity to pursue my studies in his lab, and for his guidance, support and encouragement. Thanks also to the members of my supervisory committee, Dr. Bruce Brandhorst and Dr. Jim Price, and to past and present members of the Baillie lab: Denise Clark, Bob Johnsen, Kim McKim, Yousef Amaar, Marcia Wakarchuk, B. J. Allan, Helen Stewart, Thé Ha, Andrea Scouras, Cammy Matheson, Stacy Freeman and Melonie Mondrus. Thanks to Diana Collins for her expert and patient instruction in the technique of micro-injection, Jeff Bryer for his (seemingly) tireless dispensing of computer advice, and Raja Rosenbluth, Erin Gilchrist and Helen Chamberlin for advice, illuminating discussions and for editing this thesis. Many thanks to Brad Barbazuk, Steven Jones, Marco Marra and Dinar Suleman for friendship, helpful discussions and for being a constant source of amusement. Special thanks to Marco Marra for manuscript editing above and beyond the call of duty. I would also like to thank Ruth Davenport, Sharon Gorski, Charlene Mayes, Madhu Singh and many others from the Rose, Price, Smith, Beckenbach, Druehl, Honda and Brandhorst labs for their support and friendship. Finally, I would like to thank Gary Saunders, without whom this would likely not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

The analysis of the genomes of a number of organisms is being undertaken with the aim of determining the number and types of genes present, their functions, interactions, regulation and organization in the genome. These studies will provide a knowledge of the genes and processes defining each organism, and reveal the mechanisms by which development is programmed and executed. Acquiring such an understanding of many different organisms will provide detailed molecular genetic information about biological processes particular to one or a group of organisms, as well as elucidate themes common to all organisms.

While it is ultimately of great interest and benefit to understand the human genome, and thereby human development and the basis of human genetic disorders, the size and complexity of the human genome and the difficulty of genetic analysis are problematic. In contrast, model organisms have been selected and shaped by geneticists specifically for the purpose of facilitating genetic analysis (Fink 1988). It is with a belief in the universality of cellular and developmental mechanisms, such that information gained from the study of one organism can be applied to another, that the genomes of a number of less complex model organisms are being analyzed. The knowledge gained from the genetic and molecular analysis of these more easily studied organisms will facilitate our understanding of higher eukaryotes.

Genetic analysis involves the identification by mutation of genes important for the normal growth, development, reproduction or behavior of an organism. Examination of the consequences of single gene mutations, as well as mutations in multiple loci, provides information about gene functions and can identify gene interactions and regulatory pathways. The isolation of multiple alleles of a gene can provide an array of phenotypes whose characterization allows a clearer understanding of the gene's functions. From a collection of mutations and chromosomal rearrangements, a genetic linkage map can be constructed in which the genes are ordered in a linear array on each linkage group.

Linkage maps provide a tool for genetic analysis, enabling the mapping of additional mutations and the construction of animals with defined genotypes.

While genetic analysis of mutations provides a foundation on which to base theories on the function of individual genes, a complete understanding of the function and regulation of these genes also requires molecular characterization, both at the nucleotide level and at the level of gene products and interactions. This analysis necessitates the identification of the DNA corresponding to the genetic locus in question. In the case of mutations that have been induced by the insertion of known sequences into the genome, for example transposable element insertions in *Drosophila melanogaster* (Cooley et al. 1988) and *Caenorhabditis elegans* (Moerman et al. 1986), cloning is relatively straight forward since the DNA of interest is essentially "tagged". However, for mutations induced by other means it is more difficult to identify the relevant sequences.

The construction of physical maps of the genomes of a number of different organisms has been undertaken to address the problem of cloning genetically identified loci. A physical map of the *Escherichia coli* genome already exists (Kohara et al. 1987), those of *Saccharomyces cerevisiae* (Olson et al. 1986; Riles et al. 1993) and *Caenorhabditis elegans* (Coulson et al. 1986; Coulson et al. 1988; Coulson et al. 1991) are nearing completion, and that of *Drosophila melanogaster* has begun (Ajioka et al. 1991). The primary benefit of a physical map is the immediate correlation of a position on the genetic linkage map with a clone from the physical map. However, the different methods used to construct the two types of maps means that, initially, an immediate correlation is not possible. Distances between markers on the genetic map are based upon recombination frequencies, which may not be uniform along the chromosome. Physical maps may be generated by identification of restriction endonuclease sites (Kohara et al. 1987), sequence tagged sites (STSs) (Olson et al. 1989) or, as is common in the production of most large scale physical maps, by the generation of a collection of contiguous overlapping clones, or contigs (Coulson et al. 1986; Ajioka et al. 1991). The

initial step in correlating the genetic and physical maps is generally placement on the physical map of molecular data, corresponding to mutationally identified loci, accumulated prior to the assembly of the physical map (Coulson et al. 1986; Olson et al. 1986; Kohara et al. 1987).

A clone-based map provides the tools with which to create a high resolution genome map, as these clones are a ready source of DNA for genetic transformation experiments which can be used to establish the correlation between genetic and physical loci. Each cloned gene on the genetic map provides an anchor between the genetic and physical maps. Each anchor then provides a reference point for the cloning of additional genes. Thus, for a given organism, the ease with which one can move from a genetically identified locus to the corresponding physical locus, or vice versa, depends largely on the refinement and completeness of the genetic and physical maps and the number of previously correlated genetic and physical loci in the region of interest.

Recently, as part of the human genome sequencing initiative, efforts to determine the entire genomic sequence of several model organisms have been initiated, the most notable being the nematode *C. elegans* (Sulston et al. 1992; Wilson et al. 1994), the yeast *S. cerevisiae* (Oliver et al. 1992; Ouellette et al. 1993; Dujon et al. 1994), and the bacterium *E. coli* (Yura et al. 1992). Not only will these efforts reveal the complete genetic instructions of each organism, they will also reveal biological similarities between organisms, and provide insight into the structural and functional organization and evolution of their genomes.

Simply acquiring the sequence of all the genes of an organism does not, however, reveal the biological function of these genes. The functions of genes that have been identified by sequence analysis can only be inferred by sequence similarities to other genes of known biological function. In the cases of the *S. cerevisiae* and *C. elegans* genome sequencing projects, a large proportion of the predicted protein products show no informative similarities to sequences in the databanks. In *S. cerevisiae*, some 40 - 44% of

the genes identified on chromosome XI of are of unpredicted function, which is a similar proportion to that calculated for chromosome III genes (Dujon et al. 1994). In *C. elegans*, approximately one-third of the gene products predicted from the first 2.2 Mb of genomic sequence from chromosome III exhibit significant matches to proteins from organisms other than *C. elegans* (Wilson et al. 1994). The biological significance of genes for which no informative database similarities exist will therefore have to be determined by other methods. In *S. cerevisiae*, the method of targeted gene disruption (Rothstein 1993) has been used to investigate potential phenotypic effects caused by disruption of some of the novel genes identified (Oliver et al. 1992; Ouellette et al. 1993). Analogous technology is available to *C. elegans* researchers in the form of a PCR-based method for the selection of Tc1 transposon insertions into sequenced genes (Zwaal et al. 1993). However, this procedure is time consuming and its suitability for large scale recovery of mutations in a defined region of the genome has yet to be proven.

The ability to assign biological functions to most genes in *C. elegans* is therefore predominantly dependent upon conventional mutagen-based methods for the induction of genetic mutations, followed by genetic and molecular characterization. Fortunately, *C. elegans* is well suited for this type of analysis. Its small size, short life cycle (3.5 days at 20^{o}C) and primarily hermaphroditic mode of reproduction (Brenner 1974), contribute to its ease of laboratory cultivation. Males exist, which allows genetic crosses to be performed. Strains may be stored indefinitely in liquid nitrogen. The complete cell lineage is known (Sulston and Horvitz 1977; Kimble and Hirsh 1979; Sulston et al. 1980; Sulston et al. 1983), and the anatomy at the level of the electron microscope has been described (White et al. 1986). Over 1200 genetic loci have been identified to date (Wilson et al. 1994), the majority of which have been identified by mutation (D. Baillie, personal communication). At 1 x 10^8 base pairs (bp) (Sulson and Brenner 1974; J. Sulston, personal communication), *C. elegans* has the smallest haploid genome size of any known metazoan (Coulson et al. 1991). A nearly complete physical map has been

achieved with overlapping cosmid and YAC (yeast artificial chromosome) clones (Coulson et al. 1986; Coulson et al. 1988; Coulson et al. 1991) and the systematic sequencing of the entire genome has begun (Sulston et al. 1992; Wilson et al. 1994). Methods for the introduction of cloned genes into the genome have been described (Stinchcomb et al. 1985; Fire 1986; Mello et al. 1991) that provide a relatively simple and convenient procedure for correlating the genetic and physical maps. In addition, genetic and molecular data are readily available to researchers from the database ACEDB (**a** ***C. elegans*** **database**; R. Durbin and J. Thierry-Mieg, unpublished).

Given that the majority of genes identified by the *C. elegans* Genome Sequencing Consortium have no informative similarities to other known genes (Wilson et al. 1994), it becomes necessary to correlate these genes with genetic mutations in order to ascribe a biological function. In the absence of a convenient method for the recovery of targeted disruptions in such a large number of molecularly identified genes, there is a need for a high resolution genetic map constructed from mutations recovered by non-targeted methods in sequenced regions of the genome. Clones bearing one or more genes can then be assayed for their ability to rescue the mutant phenotype of individuals carrying mutations that map in the region.

Essential genes can be defined as those required for normal development and reproduction. The identification of essential genes in *C. elegans* has been instrumental in the construction of a high resolution genetic map. Several regions of the *C. elegans* genome have been characterized with respect to their essential gene content, including regions on the X chromosome (Meneely and Herman 1979; Meneely and Herman 1981), on chromosome I (Rose and Baillie 1980; Howell et al. 1987; Howell and Rose 1990; McKim et al. 1992), on chromosome II (Sigurdson et al. 1984), on chromosome III (H. Stewart, D. Collins, A. M. Howell and D. L. Baillie, personal communication), on chromosome IV (Rogalski et al. 1982; Rogalski and Baillie 1985; Rogalski and Riddle 1988; Clark et al. 1988; Charest et al. 1990; Clark and Baillie 1992; Marra 1994), and

on chromosome V (Rosenbluth et al. 1988; McKim et al. 1988; Johnsen and Baillie 1991). An advantage of this mutational approach over targeted methods is that a high resolution genetic map can be attained without prior knowledge of the sequence.

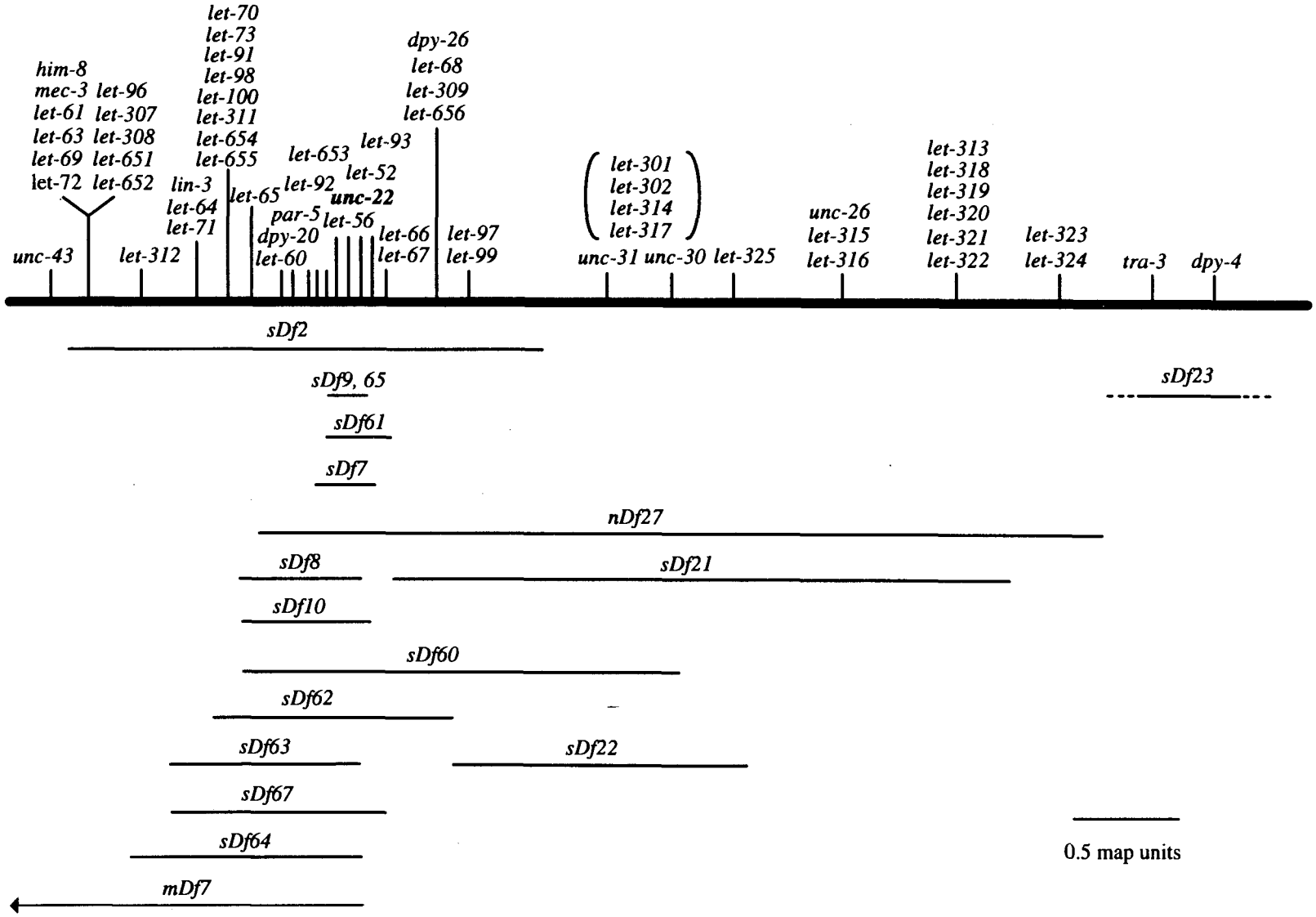
This thesis describes a continuation of efforts to obtain a high resolution genetic map in the *unc-22(IV)* region which is correlated at high resolution to the physical map. Previous mutational analyses in a two map unit interval defined by the breakpoints of the deficiency *sDf2* (Moerman and Baillie 1981), resulted in the identification of 34 essential genes and one maternal effect gene (see Figure 1) (Rogalski et al. 1982; Rogalski and Baillie 1985; Clark et al. 1988; Clark and Baillie 1992). This two map unit interval was estimated to be at most 73% saturated for mutations in essential genes (Clark and Baillie 1992). An additional 15 essential genes were identified and mapped between the right breakpoint of *sDf2* and the right breakpoint of *nDf27* (Charest et al. 1990; see Figure 1). The positioning of these genes relative to each other on the genetic map was facilitated by a set of overlapping deficiencies whose breakpoints subdivide the region into many smaller "zones". However, several clusters remained within which the gene order was not determined.

The availability of numerous overlapping deficiencies simplifies the mapping of newly identified mutations in the region by limiting the number of complementation tests required to determine whether these represent mutations in previously unidentified genes, or new alleles of previously identified genes. In addition, the correlation of genetically identified loci with the physical map, or vice versa, is made easier when the order of the genes with respect to each other is known. As well as providing a genetic tool for the ordering of genes identified by mutation, deficiency breakpoints can be mapped molecularly by identification of junction fragments, thereby providing a link between the genetic and physical maps.

Correlation of the genetic and physical maps in the *unc-22* region has previously been achieved by germline transformation using cosmid clones. Using this technique, six

Figure 1

Genetic map , prior to this thesis, of a seven map unit region around the *unc-22* gene on the right end of chromosome IV (adapted from Clark 1990). The heavy line represents the genetic map. Mutationally identified genes are shown above the genetic map. Genes in parentheses were not positioned with respect to *unc-31* and *unc-30*. Deficiencies with breakpoints in the region are shown below the genetic map. All deficiencies, with the exception of *sDf21*, *sDf22* and *sDf23*, delete the *unc-22* gene (indicated by bold lettering). The dashed lines of *sDf23* indicate that the exact breakpoints are unknown. The arrowhead on *mDf7* indicates that it extends beyond the region shown. The deficiency *sDf19*, which is not shown, has its left breakpoint between *let-653* and *let-56*, and its right breakpoint within the *unc-22* gene. *let-654* has been renamed *sem-3* because it has mutant alleles that affect sex muscle development in the hermaphrodite (M. Stern, personal communication). The single allele defining *let-302* has been lost.


him-8
mec-3
let-61
let-63
let-69
let-72
let-96
let-307
let-308
let-651
let-652
let-70
let-73
let-91
let-98
let-100
let-311
let-654
let-655
lin-3
let-64
let-71
let-65
let-653
let-92
let-93
let-52
unc-22
par-5
dpy-20
let-60
let-56
let-66
let-67
dpy-26
let-68
let-309
let-656
let-97
let-99
let-301
let-302
let-314
let-317
unc-43
let-312
unc-31
unc-30
let-325
unc-26
let-315
let-316
let-313
let-318
let-319
let-320
let-321
let-322
let-323
let-324
tra-3
dpy-4
sDf2
sDf9, 65
sDf61
sDf7
sDf23
nDf27
sDf8
sDf21
sDf10
sDf60
sDf62
sDf63
sDf22
sDf67
sDf64
mDf7
0.5 map units

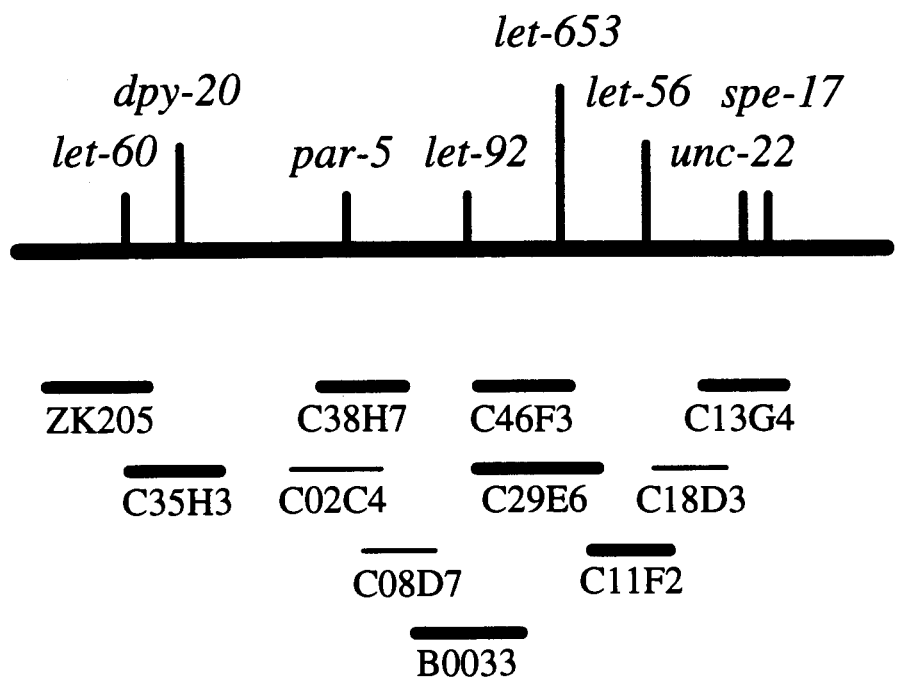
genes in the 0.2 map units to the left of *unc-22* (see Figure 2) have been placed on the physical map: *let-60* (Han and Sternberg 1990), *dpy-20* (Clark 1990), *par-5* (D. Shakes, personal communication), *let-92* (Jones 1994), *let-653* and *let-56* (Clark 1990; Clark and Baillie 1992). These data indicate an average of one essential gene per cosmid in this small interval.

In addition to genetic analysis, extensive molecular analysis has been carried out in the *let-60* - *unc-22* interval. *dpy-20* has been cloned and sequenced (Clark 1990; D. Suleman, D. V. Clark and D. L. Baillie, unpublished), and *let-60* has been found to encode a *C. elegans ras* homologue (Han and Sternberg 1990). Cloning and sequencing of the *let-653* gene has shown it to encode a mucin-like protein (Jones 1994). The *unc-22* gene has also been cloned (Moerman *et al.* 1986) and sequenced (Benian *et al.* 1989, 1993), and encodes a muscle component called twitchin that is thought to interact with myosin (Moerman et al. 1982).

Several coding regions in the *dpy-20* - *unc-22* interval were identified by Prasad and Baillie (1989), who subcloned restriction fragments from four cosmids in the *dpy-20* - *unc-22* contig and, using interspecies cross-hybridization to *Caenorhabditis briggsae* DNA, identified nine putative coding regions. Seven of these nine restriction fragments also hybridized to RNA on *C. elegans* Northern blots, indicating that they contained transcribed genes. Of these seven restriction fragments, three were derived from the cosmid C11F2, which contains the *let-56* gene (Clark 1990; Clark and Baillie 1992), and two were derived from the cosmid C18D3, which lies between, and overlaps with, C11F2 and the *unc-22* - containing cosmid C13G4 (Figure 2). Further investigation of transcribed sequences on the cosmid C11F2 has identified two additional transcribed genes (Marra 1994). In addition, 13 kilobases (kb) of genomic sequence 3' (to the left) of *unc-22* was determined by Benian et al. (1989) during the sequencing of the *unc-22* gene. This 13 kb of sequence is contained within the cosmid C18D3. Preliminary analysis of this sequence identified four potential coding elements (D. L. Baillie and C. Fields,

Figure 2

A portion of the physical and genetic maps in the *let-60* - *unc-22* interval, adapted from ACEDB (R. Durbin and J. Thierry-Mieg, unpublished). The heavy line represents the genetic map. Genes that have been placed on the physical map are shown. The physical map, represented by selected cosmid clones, is shown below the genetic map. Cosmids drawn with a heavy line have been shown by germline transformation rescue to contain the gene(s) directly above them. *let-60* was rescued by ZK205 (Han and Sternberg 1990), *dpy-20* was rescued by C35H3 (Clark 1990), *par-5* has been tentatively assigned to C38H7 (D. Shakes, personal communication), *let-92* was rescued by B0033 (Jones 1994), *let-653* was rescued by C29E6 (Clark 1990; Clark and Baillie 1992) and C46F3 (Jones 1994), *let-56* was rescued by C11F2 (Clark 1990; Clark and Baillie 1992), and both *unc-22* and *spe-17* sequences are on C13G4 (Benian et al. 1989; L'Hernault et al. 1993). The cosmid clones studied by Prasad and Baillie (1989) are also shown. They are C02C4, C08D7, C11F2 and C18D3.


let-653
dpy-20
let-56
spe-17
let-60
par-5
let-92
unc-22
ZK205
C38H7
C46F3
C13G4
C35H3
C02C4
C29E6
C18D3
C08D7
C11F2
B0033
0.1 mu or 100 kb

personal communication). A total of ten coding elements have therefore been identified in an interval to which only a single essential gene, *let-56*, has been mapped.

In Part 1 of this thesis I describe the isolation and characterization of additional *unc-22* deficiencies for use in the further refinement of the genetic map position of genes identified in the *unc-22* region. Twenty-two new deficiencies were isolated, more than doubling the number of deficiencies in the *unc-22* region. These 22 deficiencies define eight new genetic breakpoints, and have moved eight genes from zones containing two or more genes into zones in which they are the sole occupants. In addition, the complementation pattern of several of these deficiencies identifies an unusual property of the DNA immediately to the right of *unc-22*. In a collaborative effort with M. Marra and G. Benian, the 13 kb of sequence 3' of *unc-22* was further characterized for transcribed elements. Three of the four predicted genes were found to be transcribed. By polymerase chain reaction (PCR) (Saiki et al. 1985; Saiki et al. 1988) analysis of two of the deficiencies that I isolated, we determined that none of these four genes, nor any of the genes between *let-56* and *unc-22*, are required for viability under laboratory conditions. Some of the results described in this part of the thesis have been published (Schein et al. 1993).

Part 2 of this thesis describes the systematic correlation of the genetic and physical maps immediately to the right of the *unc-22* gene, using the technique of germline transformation. Three genes, *let-52*, *let-661* and *let-93*, were placed on the physical map. Analysis of these data indicate an average of one essential gene per every two cosmids in the region to the right of the *unc-22* gene.

Materials and Methods

The genetic nomenclature used follows the recommendations of Horvitz et al. (1979).

Nematode strains and culture conditions: Nematodes were maintained on Petri plates containing nematode growth medium (NGM) streaked with *Escherichia coli* OP50 (Brenner 1974). The wild-type strain N2 (var. Bristol) was obtained from the stock collection at the Medical Research Council, Cambridge, England. Mutations used in this work are listed in Table 1. The strain DR789 was obtained from the laboratory of D. Riddle. The strain JT6883 was obtained from the laboratory of J. Thomas.

Mutations in the *unc-22* gene (Brenner 1974) are used throughout this work. All *unc-22* mutations have a recessive "twitcher" phenotype (Brenner 1974; Moerman and Baillie 1979; Waterston et al. 1980) with the exception of *unc-22(m52)* (D. Riddle, personal communication) and the deficiency *sDf19* (Rogalski and Baillie 1985), which have dominant twitcher phenotypes. Mutations in *unc-22* also have a conditionally dominant phenotype such that individuals heterozygous for an *unc-22* mutation twitch in a 1% nicotine solution (Sigma) but are otherwise phenotypically wild-type (Moerman and Baillie 1979). Genotypically wild-type individuals are contracted and paralyzed in nicotine.

Mutagenesis: Mutations affecting *unc-22* were recovered following treatment of wild-type animals with formaldehyde or 254-nm ultraviolet light.

(*i*) **Formaldehyde mutagenesis (JES1):** N2 hermaphrodites were mutagenized following the protocol described by Moerman and Baillie (1981) except that 0.1% formaldehyde was used as recommended by Johnsen and Baillie (1988). The mutagenized worms were spotted onto seeded 90-mm plates for a two hour period (20°C) to allow the worms to recover and expel embryos present in-utero at the time of mutagenesis. After the recovery period, five young, gravid hermaphrodites were placed on each of 95 60-mm Petri plates and allowed to lay eggs for 24 hours. These animals

Table 1

Mutations used in this study

Mutation	Source	Reference
dpy-4(e1166)	M.R.C.[a]	D. Riddle
dpy-20(e1282)	C.G.C.[b]	B. Edgar
dpy-26(n199)	M.R.C.	Hodgkin 1983
hDf13(h12)	A. Rose	L'Hernault et al. 1993
let-52(s42)	D. Moerman	Rogalski et al. 1982
let-52(s2346)	M. Marra	Marra 1994
let-56(s46)	D. Moerman	Rogalski et al. 1982
let-56(s173)	D. Moerman	Rogalski et al. 1982
let-59(s1087)	D. Clark	Clark and Baillie 1992
let-60(s1124)	D. Clark	Clark et al. 1988
let-64(s216)	D. Moerman	Rogalski et al. 1982
let-65(s174)	D. Moerman	Rogalski et al. 1982
let-66(s1739)	D. Clark	Clark and Baillie 1992
let-67(s214)	D. Moerman	Rogalski et al. 1982
let-68(s1081)	D. Clark	Clark and Baillie 1992
let-70(s1132)	D. Clark	Clark et al. 1988
let-71(s692)	T. Rogalski	Rogalski and Baillie 1985
let-73(s685)	T. Rogalski	Rogalski and Baillie 1985
let-73(s1747)	D. Clark	Clark and Baillie 1992
let-91(s753)	L. Donati	Donati 1985
let-91(s1720)	D. Clark	Clark and Baillie 1992
let-92(s504)	D. Pilgrim	Rogalski and Baillie 1985

Table 1 (continued)

Mutation	Source	Reference
let-93(s734)	L. Donati	Donati 1985
let-93(s2254)	M. Marra	Marra 1994
let-93(s2357)	M. Marra	Marra 1994
let-97(s1121)	D. Clark	Clark et al. 1988
let-98(s1117)	D. Clark	Clark et al. 1988
let-99(s1201)	D. Clark	Clark et al. 1988
let-100(s1160)	D. Clark	Clark et al. 1988
let-301(s1134)	D. Clark	Charest et al. 1990
let-309(s1770)	D. Clark	Clark and Baillie 1992
let-311(s1195)	D. Clark	Clark et al. 1988
let-312(s1234)	D. Clark	Clark et al. 1988
let-314(s1206)	D. Clark	Charest et al. 1990
let-315(s1101)	D. Clark	Charest et al. 1990
let-316(s1227)	D. Clark	Charest et al. 1990
let-317(s1182)	D. Clark	Charest et al. 1990
let-323(s1719)	D. Clark	Charest et al. 1990
let-324(s1727)	D. Clark	Charest et al. 1990
let-325(s1738)	D. Clark	Charest et al. 1990
let-653(s1733)	D. Clark	Clark and Baillie 1992
let-654(s1734)	D. Clark	Clark and Baillie 1992
let-655(s1748)	D. Clark	Clark and Baillie 1992
let-656(s1767)	D. Clark	Clark and Baillie 1992
let-660(s1996)	M. Green	Marra 1994
let-661(s2203)	M. Marra	Marra 1994

Table 1 (continued)

Mutation	Source	Reference
let-664(s2374)	M. Marra	Marra 1994
let-662(s2219)	M. Marra	Marra 1994
lev-1(x22)	J. Lewis	Lewis et al. 1980
lin-3(s1750)	D. Clark	Clark and Baillie 1992
nT1 (IV;V) [let-(m435)]	C.G.C.	D. Riddle
par-5(it121)	K. Kemphues	unpublished
sDf23	G. Wild	unpublished
unc-22(s7)	D. Moerman	Moerman and Baillie 1979
unc-30(e191)	C.G.C.	Brenner 1974
unc-31(e169)	M.R.C.	Brenner 1974

[a] M.R.C. is the Medical Research Council, Cambridge, England.

[b] C.G.C. is the *Caenorhabditis* Genetics Centre.

were then transferred to fresh plates and allowed to lay eggs for an additional 24 hours, after which they were removed.

(*ii*) **UV mutagenesis (JES2):** Eleven N2 hermaphrodites at approximately the fourth larval (L4) stage were placed on a seeded 90-mm plate and allowed to produce self progeny. After four days the worms were exposed to 254-nm ultraviolet light (30-W G. E. UV germicidal lamp) following essentially the same protocol described by Stewart et al. (1991) except that the dose applied was 110 J/m^2. The worms were left for a two hour period (20°C) to allow expulsion of embryos in-utero during mutagenesis. After the two hour period, five young, gravid hermaphrodites were placed on each of 40 60-mm Petri plates and allowed to lay eggs for 24 hours. These animals were then transferred to fresh plates and allowed to lay eggs for an additional 24 hours, after which they were removed.

Screening for mutations affecting *unc-22*: The screening protocol is diagrammed in Figure 3. The F_1 generation of both formaldehyde and UV treated P_0 hermaphrodites were screened in 1% nicotine for the presence of twitchers. The self-progeny of each F_1 twitcher selected were screened for the presence of Unc-22 animals. Individual F_1s that produced no fertile Unc-22 progeny were suspected of carrying a deficiency affecting *unc-22* and a neighbouring essential gene or genes. These strains were maintained by selecting phenotypically wild-type individuals that twitched in nicotine every generation until the putative deficiency could be balanced. Lines that gave viable, fertile Unc-22s were maintained as homozygous twitchers and retained for PCR analysis (see below).

Balancing of deficiencies: Deficiency strains were outcrossed to N2 at least once prior to balancing. The deficiencies were balanced over *nT1(IV;V) let-?(m435). nT1(IV;V)* (Ferguson and Horvitz 1985) is a complex rearrangement that causes pseudo-linkage of chromosome IV (right) and chromosome V (left), suppressing recombination on the right arm of chromosome IV and the left arm of chromosome V. The *unc-22* region lies within

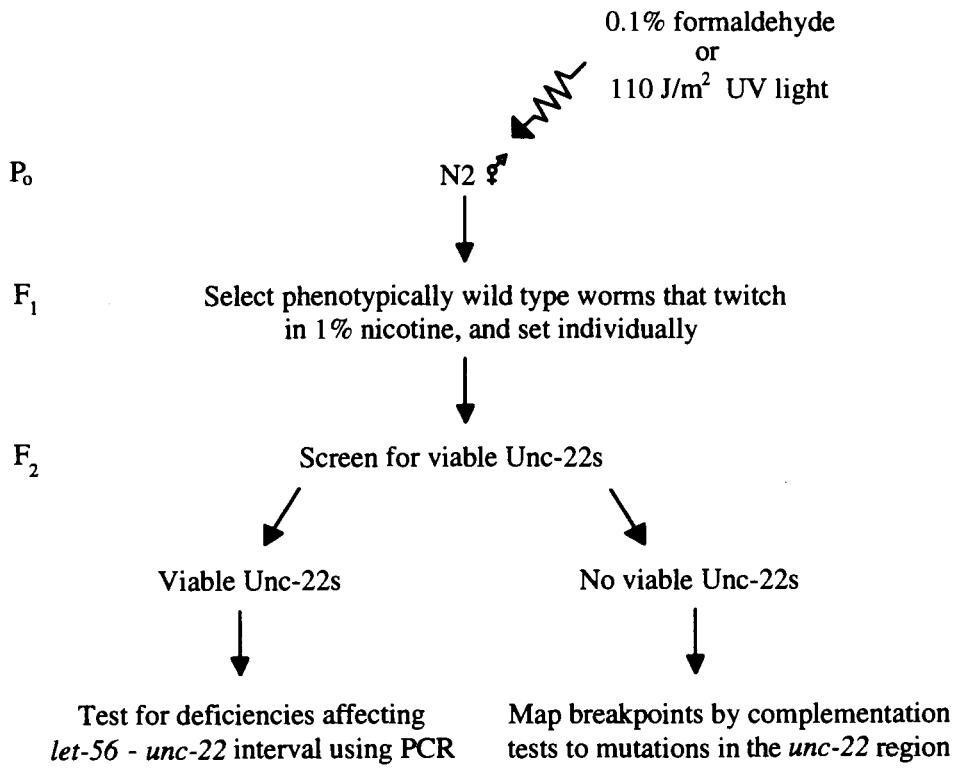


Figure 3

Flow diagram illustrating the protocol used for recovery of formaldehyde and UV radiation induced *unc-22* deficiencies.

the recombinationally suppressed region of chromosome IV. *let-?(m435)* is a recessive lethal mutation linked to the *nT1(IV;V)* rearrangement (Rogalski et al. 1988; T. Rogalski, personal communication).

The deficiencies were balanced by one of two methods. In the first method, L4 stage *sDfx*/+ hermaphrodites were mated to *dpy-13/nT1(IV)* [*let-?(m435)*]; *unc-42/nT1(V)* [*let-?(m435)*] males which arose spontaneously from strain DR789. Individual outcrossed hermaphrodite progeny that twitched in nicotine were selected to separate plates (either *sDfx/dpy-13*; *+/unc-42* or *sDfx/nT1(IV)* [*let-?(m435)*]; *+/nT1(V)* [*let-?(m435)*]). A single line segregating no Dpy-13 Unc-42s was retained for each of the deficiencies balanced in this manner. In the second method, L4 stage *sDfx*/+ hermaphrodites were mated to N2 males. Male progeny that twitched in nicotine (*sDfx*/+) were mated to L4 stage hermaphrodites from the strain DR789 (*dpy-13/nT1(IV)* [*let-?(m435)*]; *unc-42/nT1(V)* [*let-?(m435)*]). Individual outcrossed hermaphrodite progeny that twitched in nicotine were selected to separate plates (either *sDfx/dpy-13*; *+/unc-42* or *sDfx/nT1(IV)* [*let-?(m435)*; *+/nT1(V)* [*let-?(m435)*]). A single line segregating no Dpy-13 Unc-42s was retained for each of the deficiencies balanced in this manner.

Complementation mapping of deficiencies: All complementation tests were performed at 20°C. Putative deficiencies (those producing no viable Unc-22 progeny) were tested for complementation with recessive lethal, maternal effect and visible mutations in the *unc-22* region in order to determine their extents.

All lethal mutations used (including *lin-3(s1750)*) and *par-5(it121)* are linked to an *unc-22* mutation. For these complementation tests, males heterozygous for the *unc-22* linked mutation were crossed to *sDfx*/+ hermaphrodites. Outcrossed progeny were screened for the presence of fertile Unc-22 hermaphrodites. If no fertile Unc-22 hermaphrodites were found in the outcrossed progeny, then the mutation was considered to be uncovered by the deficiency.

For complementation tests with mutations causing visible phenotypes (except *dpy-26*), phenotypically wild-type males heterozygous for the visible mutation were mated to *sDfx*/+ hermaphrodites. The outcrossed progeny were screened for the presence of individuals expressing the visible phenotype. If individuals expressing the visible phenotype were found, then the deficiency was considered to uncover the visible mutation.

dpy-26(n199) is a maternal effect, hermaphrodite-specific lethal. The phenotype of *dpy-26(n199)* has been described (Hodgkin 1983). Homozygous *dpy-26(n199)* hermaphrodites from a *dpy-26(n199)*/+ parent are rescued (they are slightly dumpy but are viable). However, the progeny of these homozygotes predominantly arrest as unhatched eggs or larvae, with only a few severely dumpy hermaphrodites from each brood reaching adulthood. Each brood also contains a small number of almost wild-type males. For complementation tests with *dpy-26(n199)*, *sDfx*/+ males were mated to *dpy-20 dpy-26*/*nT1(IV)*; +/*nT1(V)* hermaphrodites. For those deficiencies that uncover *dpy-20*, the outcrossed progeny were screened for Dpy hermaphrodites that twitched in nicotine (*dpy-20 dpy-26*/*sDfx*), and these Dpys were picked to individual plates. If the progeny of the set Dpys consisted primarily of dead eggs, arrested larvae and a few viable males, then the deficiency was considered to uncover *dpy-26*. For those deficiencies that do not uncover *dpy-20*, outcrossed progeny were screened for hermaphrodites that twitched in nicotine (either *dpy-20 dpy-26*/*sDfx* or *sDfx*/*nT1(IV)*; +/*nT1(V)*) and these were picked to individual plates. Worms which were genotypically *sDfx*/*nT1(IV)*; +/*nT1(V)* were identified by the lack of Dpy worms in their progeny, and were discarded. If the progeny of the remaining set worms consisted primarily of dead eggs, arrested larvae and a few viable males, then the deficiency was considered to uncover *dpy-26*.

The results of these complementation tests for JES1 and JES2 deficiencies are presented in the Appendix.

For complementation tests with *hDf13*, which affects both *unc-22* and *spe-17*, *hDf13*/+ males were mated to *sDfx*/+ hermaphrodites. The outcrossed progeny were screened for Unc-22 hermaphrodites (*sDfx/hDf13*), which were selected to individual plates. If these hermaphrodites displayed a Spe-17 phenotype, producing no, or very few, progeny in addition to many unfertilized oocytes (L'Hernault et al. 1993), then the deficiency was considered to uncover *spe-17*.

One of the deficiencies I isolated, *sDf81*, was tested for complementation with *sDf23*. This was accomplished by mating *sDf81*/+ males to hermaphrodites from the strain JT6883 (*unc-22 unc-31 sDf23/nT1(IV); +/nT1(V) unc-(n754dm)*). The presence of fertile Unc-22 Unc-31 hermaphrodite progeny, of the genotype *sDf81/unc-22 unc-31 sDf23*, was taken to indicate that *sDf81* and *sDf23* complemented each other.

Examination of the phenotype of *sDf83* homozygous individuals: Worms were kept at 20°C. Five gravid, young adult *sDf83*/+ hermaphrodites were set on a plate and allowed to lay eggs for 12 hours, after which they were removed. After five days, larval Unc-22s (*sDf83/sDf83*) were picked to a separate plate and allowed to develop for an additional nine days. At this time, approximately 14 days after being laid as eggs, the worms were observed and measured on a dissecting scope using an ocular micrometer. The approximate developmental stage of the worms was determined by comparison of the length of the *sDf83* homozygotes to the *unc-22* growth curve (Rogalski et al. 1982).

Oligonucleotide primers: The oligonucleotide primers used in this work are listed in Table 2, with appropriate references.

PCR techniques: Reagents used for some reactions were those supplied in the GeneAmp PCR Kit (Perkin Elmer Cetus). In some cases reactions were performed using *Taq* polymerase and buffers obtained from Promega, or *Pfu* polymerase and buffers

Table 2

Oligonucleotide primers used for PCR

Primer	Sequence (5' - 3')	Target	Reference[a]
C6-1	GCCAGCAACAAACCGAAG	*adl-1*	1
C6-2	AAGAACAGCAGACAACGC	*adl-1*	1
DB-19	GCTATGATGGAGGAAAGTG	*nhe-1*	2
DC1	GCGGGGTTGCCTTACTGG	pJB8	3
DC2	CGCAAGGTCGGGCTGAAC	pJB8	3
Dup-1	ATTCACAAAACACATCCC	*dup*	1
Dup-2	GTTATGTGACCGATGAGC	*dup*	1
K08-3	TCAAAACGCATCTAAACTGG	LGV	4
K08-7	GCCACACTAAAACAGGATTC	LGV	4
KRp17	CGTCCGGCGCACAGAAGC	Lorist	6
KRp18	GTGCTGAGCCCGGCCAAA	Lorist	6
KT01	CTCCAAGTAACACTACGTCGTGG	*spe-17*	5
KT03	ATGCACTATATTTTCCAATATTTTTCAT	*spe-17*	5
Lunc-1	CCGTTTGAGAGTTGAGCG	*gtl-1*	1
Lunc-2	CGACGAGACAGCCACAAC	*gtl-1*	1
MAM-2	AGGTTTCTCAGTTTTTGG	*nhe-1*	2
Mys-1	AAGTTCGTAAAGATGCCC	*ptl-1*	1
Mys-2	AGATTTGGGCATTTTTCG	*ptl-1*	1
R1	CGGCGGAGAAGAACAAGC	*hdl-1*	2
R8	GCACGGAAGCAAATCAACCC	*hdl-1*	2
179	TGATACTTCCCTTTTTCG	LGV	7
180	CATTACACGGAGAAGACG	LGV	7
265-F	CAAAATGGATGATGAGAAGTGC	LGV	4

Table 2 (continued)

Oligonucleotide primers used for PCR

Primer	Sequence (5' - 3')	Target	Reference[a]
265-R	CAACTGCCCTGACAACTCCGAC	LGV	4

[a] Reference 1 is Schein et al. 1993. Reference 2 is Marra et al. 1993. Reference 3 is D. Collins, personal communication. Reference 4 is K. S. McKim, personal communication. Reference 5 is G. M. Benian and S. L'Hernault, personal communication. Reference 6 is McKay 1993. Reference 7 is W. B. Barbazuk, personal communication.

(Stratagene). The deoxynucleoside triphosphates used in some reactions were obtained from Pharmacia. Reactions were carried out in 0.5 ml microfuge tubes in an Ericomp TwinBlock System, or in glass capillary tubes in an Idaho Technology model 1605 Air Thermo-Cycler. For reactions performed on the Idaho Technology Thermo-Cycler, the reaction buffer used was supplied by Idaho Technology. PCR amplifications were carried out in 25 µl volumes following a whole worm protocol described by Barstead and Waterston (1991), with modifications described by Williams et al. (1992). Cycling conditions used with the Ericomp TwinBlock System were one cycle of 5 minutes @ 94°C, 1 minute @ 58°C, 2 minutes @ 72°C, followed by 25 - 30 cycles of 1 minute @ 94°C, 1 minute @ 58°C, 2 minutes @ 72°C, and one cycle of 1 minute @ 94°C, 1 minute @ 58°C and 10 minutes @ 72°C. Cycling conditions used with the Idaho Technology Air Thermo-Cycler were one cycle of 1 minute @ 94°C, followed by 30 cycles of 10 seconds @ 92°C, 20 seconds @ 59°C, 40 seconds @ 72°C, and one cycle of 2 minutes @ 72°C.

PCR analysis of *unc-22* mutations: For each mutation tested, two to five late larval Unc-22 individuals (for homozygous viable *unc-22* mutations) or 12 to 15 early larval Unc-22 individuals (for *sDf83* homozygotes) were used in PCR reactions. Each reaction contained a primer pair specific for one of the coding elements 5' or 3' of the *unc-22* gene (see Figure 5), plus a control primer pair specific for sequences on LGV. The control primers were included in order to differentiate between a failed PCR and one in which the *unc-22* mutation being tested affected the annealing site of at least one of the experimental primers. *unc-22* mutations that were deficiencies were identified by failure to produce the expected PCR product with one or more of the primer pairs specific for coding elements 5' or 3' of *unc-22*.

Origin of cosmid clones: Cosmid clones used in this work were obtained from A. R. Coulson and J. E. Sulston (Coulson et al. 1986) at the Sanger Centre, Cambridge, England. The cosmids are F20B10, F35G2, K04H1, K08E4, T12G3, T20H7 and ZC197. Cosmids F20B10 and F35G2 contain the vector Lorist6 (Gibson et al. 1987b), cosmids K04H1 and K08E4 contain the LoristB vector (Cross and Little 1986), and cosmids T12G3 and T20H7 contain the Lorist2 vector (Gibson et al. 1987a), all of which contain a kanamycin resistance gene. Cosmid ZC197 contains the vector pJB8 (Ish-Horowicz and Burke 1981) which carries an ampicillin resistance gene.

Preparation of cosmid DNA for germline transformation: DNA from cosmids F20B10 and T12G3 was prepared in the following manner: 40 ml of 2 X YT medium (Sambrook et al. 1989) containing 10 µg/ml kanamycin, were inoculated and left overnight at 37°C in a shaking incubator (250 rpm). Cosmid DNA was isolated using an alkaline lysis miniprep protocol provided by M. Craxton. The pelleted DNA was resuspended in a total volume of 1 ml TE buffer, and prepared for purification by CsCl density gradient centrifugation as per Sambrook et al. (1989), in 11 x 32 mm Quick-Seal centrifuge tubes (Beckman). The gradients were spun at 20°C for 3 hours at 100 000 rpm, followed by 1 hour at 70 000 rpm, using a Beckman TL-100 ultracentrifuge and TLV 100k rotor. Following the protocol described in Sambrook et al. (1989), cosmid DNA bands were collected from the centrifuge tubes, the ethidium bromide extracted with water saturated butanol, and the DNA precipitated . The DNA was re-dissolved in a total volume of 200 µl of TE buffer.

DNA for each of the remaining cosmids was prepared in the following manner: 10 ml of 2 X YT medium containing either 30 µg/ml ampicillin or 10 µg/ml kanamycin was inoculated and left overnight at 37°C in a shaking incubator (250 rpm). The cultures were divided into 1.5 ml Eppendorf tubes, and the cosmid DNA was isolated using the alkaline lysis miniprep protocol and solutions provided with the Pharmacia Miniprep Kit

Plus. The DNA pellets were resuspended in 40 - 50 µl of TE buffer, left overnight at 4°C, and then pooled. No further purification steps were found to be necessary for successful transformations apart from centrifugation of the injection mixture for 30 minutes at 13 000 rpm in a desktop microfuge immediately prior to injection to pellet any cellular debris.

Restriction enzyme digests of cosmid DNA: 0.1 - 0.4 µg of cosmid DNA was digested with 5 units of *Eco*R1 in a 20 µl volume for 1 h at 37°C. The enzyme and appropriate buffer were obtained from Bethesda Research Laboratories.

Agarose gel electrophoresis: PCR products and restriction enzyme digested DNA fragments were resolved by electrophoresis through agarose gels. PCR products were electrophoresed through 0.8 - 1.2 % agarose (w:v) gels in either 0.5 X TBE or 1 X TAE buffer (Sambrook et al. 1989). Restriction enzyme digested DNA fragments were resolved on 0.7 - 0.8 % agarose gels in 1 X TAE. Gels were either ethidium bromide free, or contained 0.05 µg/ml ethidium bromide. Ethidium bromide free gels were used when quantifying cosmid DNA preparations. The gels were electrophoresed in the appropriate buffer (either 0.5 X TBE or 1 X TAE) at 3 - 5 V/cm. Marker DNA was 1 kb ladder (Bethesda Research Laboratories). Following electrophoresis, ethidium bromide free gels were stained for 20 minutes in 0.5 X TE buffer containing 0.5 µg/ml ethidium bromide, and then rinsed briefly in 1 X TAE buffer. The DNA was made visible by exposure to 300-nm UV light, and photographed using either Kodak PLUS-X pan film, or a UVP Image Store 5000 video documentation system.

Quantification of cosmid DNA: Cosmid DNA was quantified by electrophoresing an *Eco*R1 digested sample on an ethidium bromide free agarose gel with a known quantity of marker DNA. After staining and photographing the gel, the concentration of cosmid

DNA was approximated by comparing the relative intensity of the cosmid DNA bands to those of the marker DNA.

Germline transformation techniques: Germline transformation was carried out using the methodology described by Mello et al. (1991). The DNA solution, containing a total of approximately 100 ng/µl DNA in TE buffer, was injected into the large syncitium of one or both gonadal arms of adult hermaphrodites. Included in the injection solution was a plasmid containing the *rol-6(su1006)* gene (Kramer et al. 1990), which acts as a dominant morphological marker for transformation. Individuals that carry the semi-dominant *rol-6(su1006)* gene in an extrachromosomal array display a Rol-6 phenotype (right handed rolling about the long axis). The *rol-6(su1006)*-containing plasmid used was pCes1943, which is a derivative of plasmid pRF4 (Kramer et al. 1990), modified to contain a kanamycin resistance gene (B. Barbazuk and S. Jones, unpublished). pCes1943 therefore carries both an ampicillin resistance gene and a kanamycin resistance gene, providing it with sequences homologous to both Lorist and pJB8 vectors. This sequence homology is important as homologous recombination drives the formation of extrachromosomal arrays (Mello et al. 1991).

Typical injection solutions contained 10 ng/µl cosmid DNA and 90 ng/µl pCes1943 DNA. If no Rol-6 transformants were obtained, presumably because cosmid sequences were "poisonous" at this concentration, then the concentration of the cosmid DNA was decreased to 5 ng/µl. In this case the pCes1943 concentration was raised to 95 ng/µl in order to keep the total DNA concentration at 100 ng/µl.

Confirmation of cosmid incorporation in stably inherited extrachromosomal arrays: For each stably transformed line, two Rol-6 individuals were used in PCR reactions. A cosmid vector-specific primer pair (see Table 2) and a control primer pair were used in each reaction. Control primers used were specific for one of the coding elements 5' of the

unc-22 gene (see Table 2 and Figure 5). The control primers were included in order to differentiate between a failed PCR and one in which vector sequences were not amplified due to the absence of cosmid in the extrachromosomal array.

Complementation tests with strains bearing extrachromosomal arrays: Strains carrying cosmids in stably inherited extrachromosomal arrays were used as duplication bearing strains in complementation tests with lethal mutations to the right of *unc-22* (*let-52*, *let-661*, *let-93* and *let-662*). The protocol is described below, and diagrammed in Figure 4.

Males of the genotype *unc-22 let-x*/+ + or *unc-22 let-x unc-31*/+ + + or *unc-22 let-x lev-1*/+ + + were mated to Rol-6 +/+; *sExn* hermaphrodites. In the F_1 generation, L3 - L4 stage Rol-6 hermaphrodites that twitched in nicotine were selected to separate plates (*unc-22 let-x*/+ +; *sExn* or *unc-22 let-x unc-31*/+ + +; *sExn* or *unc-22 let-x lev-1*/+ + +; *sExn*). The progeny of these hermaphrodites were screened for the presence of Unc-22 Rol-6 or Unc-22 Unc-31 Rol-6 individuals, which were also presumed to be homozygous for the lethal mutation. The Lev-1 phenotype, which is resistance to the antihelminthic drug levamisole, was not screened for. If the extrachromosomal array contained sequences sufficient to rescue the lethal mutation, then these individuals would be viable.

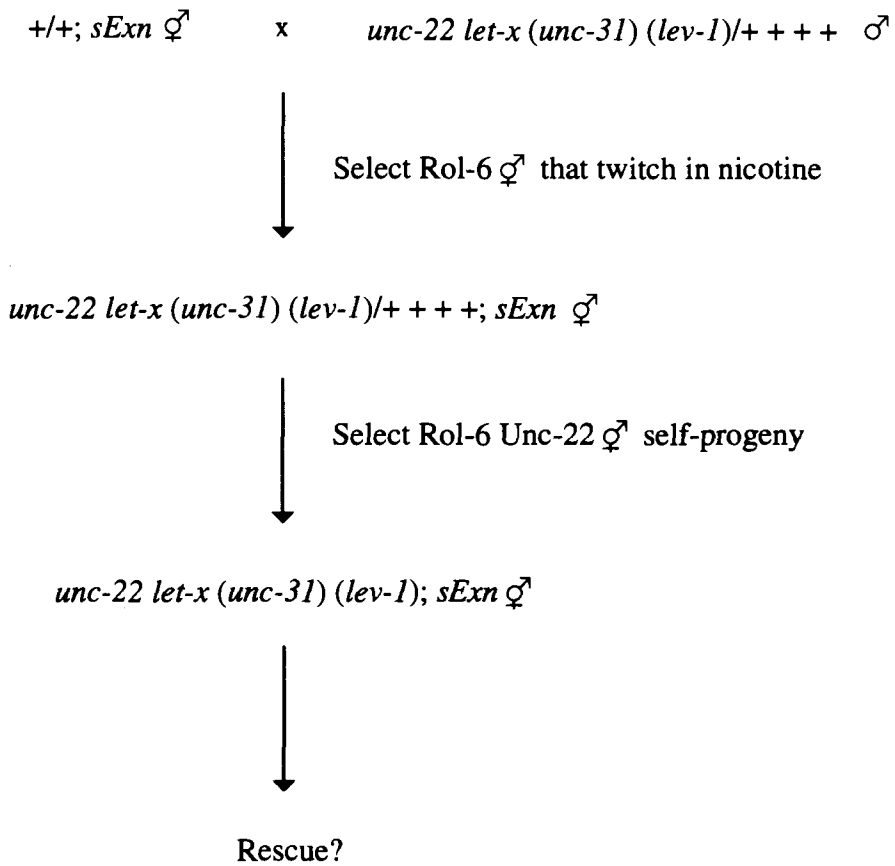


Figure 4

Protocol for complementation tests with transgenic strains. Genes in parentheses were not present in all of the lethal-bearing strains that were tested.

Results

Part 1

Isolation and Characterization of *unc-22* Deficiencies

Isolation of *unc-22* mutations: The results of screens JES1 and JES2 are presented in Table 3. In the JES1 screen 36 F_1 individuals that twitched in 1% nicotine were selected from 190 000 wild-type chromosomes screened after treatment with 0.1% formaldehyde. Of these, 27 were recovered and analyzed. In the JES2 screen 106 F_1 individuals that twitched in nicotine were selected from 81 000 wild-type chromosomes screened after treatment with 110 J/m^2 UV radiation. Of these, 25 were recovered and analyzed. Each of the 52 animals recovered was presumed to be heterozygous for a recessive mutation in the *unc-22* gene. The frequency with which *unc-22* mutations were recovered using 110 J/m^2 UV radiation was approximately twice that for 0.1% formaldehyde (3.3×10^{-4} vs. 1.4×10^{-4} per chromosome, respectively). However, while mutagenesis with UV radiation was more efficient than mutagenesis with formaldehyde, it also resulted in a much higher rate of F_1 sterility. Approximately half (55%) of the F_1 animals selected in 1% nicotine following treatment with UV radiation were sterile or very poorly fertile (Table 3).

Of the 52 F_1 twitchers recovered from the JES1 and JES2 screens, 31 produced fertile Unc-22 progeny, while 21 had no fertile Unc-22 progeny. The relative frequencies of each class were approximately equal between the two screens. Strains were established for each of the 52 mutations as outlined in Materials and Methods.

Analysis of sequence 3' of *unc-22*: Analysis of the 13 kb of genomic sequence 3' of *unc-22* (Benian et al. 1989) for open reading frames and typical nematode introns (consensus donor and acceptor sites, high AT content, and small size (Blumenthal and Thomas 1988)), resulted in the identification of four putative genes (Figure 5). The predicted amino acid sequences of the four genes were used to conduct searches of the non-redundant database, and the database of expressed sequence tags (dbest), at the National Center of Biotechnology Information (NCBI), using the BLAST algorithm (Altschul et al.

Table 3

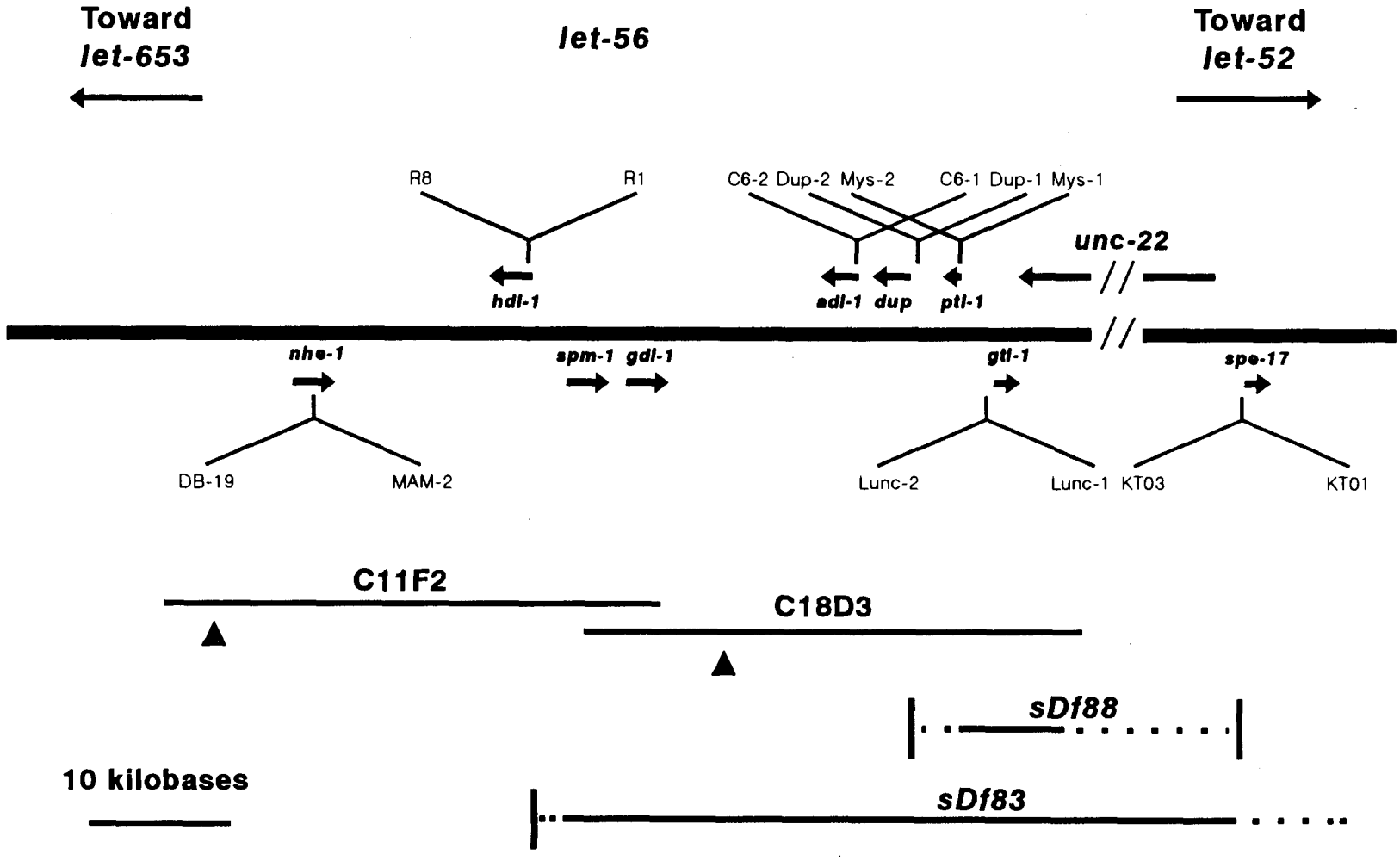
Summary of screens for *unc-22* mutations

	Screen	
	JES1 (0.1% formaldehyde)	**JES2 (110 J/m^2 UV)**
Chromosomes screened in 1% nicotine	190 000	81 000
F_1s selected	36	106
- died	2	5
- sterile, few progeny or dead eggs	2	58
Subsequently discarded, lost or not analyzed	5	18
***unc-22* mutations retained and analyzed (frequency[a])**	27 (1.4×10^{-4})	25 (3.3×10^{-4})
- deficiencies (% of retained mutations)	11 (41%)	11 (44%)
- putative intragenics (% of retained mutations)	16 (59%)	14 (56%)

[a] Per chromosome screened

Figure 5

A physical map of the *let-56 - unc-22* interval. The positions of *let-653* and *let-52* relative to this interval are indicated. The physical map is indicated by the heavy line, and cosmids C11F2 and C18D3 are shown below it. The approximate position of *let-56*, which is contained on cosmid C11F2 (Clark 1990; Clark and Baillie 1992), is shown above the physical map. Arrows indicate genes and their transcriptional orientations. The break in the *unc-22* gene indicates that approximately 28 kb are not represented on this map. PCR primer pairs specific for each gene are indicated (see Table 2 for primer sequences). The dashed lines indicate that the exact molecular breakpoints of *sDf83* and *sDf88* have not been determined. Arrowheads indicate the approximate positions of additonal restriction fragments on cosmids C18D3 and C11F2 that also detect *C. elegans* transcripts (Prasad 1988; Prasad and Baillie 1989).


Toward
let-653
let-56
Toward
let-52
R8
R1
C6-2 Dup-2 Mys-2
C6-1 Dup-1 Mys-1
unc-22
hdl-1
adl-1 dup
ptl-1
nhe-1
spm-1 gdl-1
gtl-1
spe-17
DB-19
MAM-2
Lunc-2
Lunc-1 KT03
KT01
C11F2
C18D3
sDf88
10 kilobases
sDf83

1990). For the deduced amino acid sequences, and sequence analysis additional to that presented here, see Marra (1994).

The first gene to the left (3') of *unc-22* is oriented with opposite transcriptional polarity to *unc-22*. The predicted protein sequence was found to exhibit similarity to a family of glucose transporter proteins. This gene has been named *gtl-1*, for glucose transporter-like (Marra 1994). G. Benian isolated a partial length 870 bp *C. elegans* cDNA corresponding to this gene, which when used to probe Northern blots of total RNA detected a 2.4 kb message. The sequence of this cDNA showed that the polyadenylated message of *gtl-1* overlaps the 944 nucleotide 3' untranslated sequence of the *unc-22* message by 78 nucleotides, on the opposite strand (Schein et al. 1993). This message demonstrated that the first gene predicted from the genomic sequence 3' of the *unc-22* gene was transcribed. The coordinates for the *gtl-1* gene are 43 928 to 42 190 (nucleotide positions correspond to the updated sequence, GenBank accession number L10351, Benian et al. 1993).

The second predicted gene has the same transcriptional polarity as *unc-22*, and extends from nucleotide 46 267 to at least 47 476. Database searches with the predicted protein sequence revealed a similarity to a family of neurotransmitter transporters. The highest similarity found was to a portion of the rat high affinity L-proline transporter (Fremeau et al. 1992). This gene has been named *ptl-1*, for proline transporter-like (Marra 1994). Using oligonucleotide primers designed from the DNA sequence of the predicted gene, M. Marra PCR amplified a 700 bp portion of a cDNA from a *C. elegans* cDNA library, demonstrating the existence of a transcript corresponding to this gene (Schein et al. 1993; Marra 1994). Using this amplified product as a probe, G. Benian detected a 4.2 kb message on Northern blots of total RNA (Schein et al. 1993).

The third predicted gene extends from nucleotide 49 875 to 52 239. Database searches failed to detect any significant similarities to sequences in the non-redundant and dbest databases. We named this gene *dup*, which reflects a duplicated domain in the

predicted protein sequence. Due to the proximity of *dup* to the *ptl-1* gene, and the size of the *ptl-1* message detected, it was possible that this putative gene was in fact additional *ptl-1* sequence. Attempts to PCR amplify a portion of a cDNA from a cDNA library using oligonucleotide primers designed from the predicted DNA sequence of *dup*, or primers encompassing *dup* and *ptl-1* sequences, were unsuccessful (Schein et al. 1993; Marra 1994).

The fourth predicted gene extends from nucleotide 53 438 to the end of the available sequence (at nucleotide 54 963). The predicted protein sequence showed similarity to a number of alcohol dehydrogenase enzymes in the non-redundant database. This gene was named *adl-1*, for **a**lcohol **d**ehydrogenase-**l**ike (Marra 1994). A search of the dbest database revealed identities to two cDNAs (cm12a11 and cm01b8) sequenced by the *C. elegans* Genome Sequencing Consortium (Waterston et al. 1992). In addition, the terminal approximately 600 bp of the Benian et al. (1989) sequence was found to overlap in part with the sequence of a partial cDNA that had been recovered and analyzed by S. Prasad (Prasad and Baillie 1989; S. Prasad, personal communication). This cDNA was used to probe Northern blots of total RNA, and detected a 2.4 kb message (Prasad and Baillie 1989).

The analysis described above demonstrated that three of four genes predicted from the sequence of Benian et al. (1989) 3' of *unc-22* were transcribed. The four genes are contained within sequences present in cosmid C18D3 (Figure 5). Combined with previous molecular analysis (Prasad 1988; Prasad and Baillie 1989), at least five transcribed genes have been demonstrated to lie on cosmid C18D3 . However, no mutations were known to map to this cosmid. An additional five transcribed genes are known to lie on cosmid C11F2 (Prasad 1988; Prasad and Baillie 1989; Marra et al. 1993; Marra 1994), to which only a single essential function (*let-56*) has been mapped. Although no mutations were known to map between *let-56* and *unc-22*, the possibility existed that some of the molecularly identified genes encoded essential functions. We

reasoned that the identification of small *unc-22* deficiencies with breakpoints in the *let-56* - *unc-22* interval could be used to examine whether one or more of these elements were required for development. However, mapping deficiency breakpoints with respect to the identified coding elements using standard complementation tests was not possible. In a collaboration with M. Marra, an alternate strategy based on PCR was used. Primer pairs specific for each of the four sequenced genes on cosmid C18D3 were used in PCRs with individuals homozygous for *unc-22* mutations which mapped within the *let-56* - *unc-22* region. *unc-22* deletion mutations affecting any of the primer annealing sites would be identified by lack of the PCR product expected with the particular primer pair. Once deficiencies affecting these coding elements were identified and the breakpoints determined, the phenotype of individuals homozygous for the deficiencies could be correlated with the coding elements deleted.

Identification of mutations mapping in the *let-56* - *unc-22* interval: The 31 strains from JES1 and JES2 that produced viable Unc-22 progeny were selected for PCR analysis. Either each of these 31 mutations affected only *unc-22* and no neighbouring genes, or any neighbouring genes that were affected were not required for an essential process.

In addition to the 31 Unc-22 strains from JES1 and JES2, we had available to us a further 31 viable Unc-22 strains recovered previously by members of the Baillie laboratory. Twenty-four of these 31 strains were isolated in formaldehyde screens similar to that described for JES1, using either 0.07% or 0.1% formaldehyde (Moerman and Baillie 1981). Six were isolated from screens that used either 1500 R or 2500 R gamma radiation as the mutagen (D. L. Baillie, D. V. Clark, and D. G. Moerman, unpublished results). One mutation was isolated incidentally in an unrelated screen, following mutagenesis with 0.05% formaldehyde (A. M. Rose, unpublished results). We therefore had a total of 62 Unc-22 strains available for PCR analysis.

The mutations carried by the 21 strains from JES1 and JES2 which failed to produce fertile Unc-22 progeny were presumed to be deficiencies affecting *unc-22* and a neighbouring essential gene or genes. I tested these mutations for complementation with mutations in the *unc-22* region in order to ascertain that they were in fact deficiencies, and to determine their extents. Twenty of the 21 *unc-22* mutations failed to complement mutations in at least two known essential genes in addition to *unc-22*. These 20 deficiencies were therefore too large to be of use in the examination of coding elements in the *let-56* - *unc-22* region. The mapping of these deficiencies is described in a following section.

The remaining mutation failed to complement only a single known essential gene, *let-56*, which lies immediately to the left of *unc-22* on the genetic map (Figure 1) and has been positioned on cosmid C11F2 (Clark 1990; Clark and Baillie 1992; see Figure 5). The phenotype of individuals homozygous for this mutation, *sDf83*, was indistinguishable from that described for *let-56 unc-22* homozygous individuals (Marra 1994). Individuals homozygous for *sDf83* are very slow growing animals with an Unc-22 phenotype. Two weeks after being laid as eggs, these animals reach an early adult stage, based on their length as compared to the *unc-22* growth curve (Rogalski et al. 1982). I have never observed *sDf83* individuals to be fertile, although occasionally some appear to carry an oocyte. The phenotype of *sDf83/let-56 unc-22* animals was found to be identical to that of *sDf83* homozygotes. The similarity in phenotype of *sDf83* homozygotes and *let-56 unc-22* homozygotes raised two possibilities - either *sDf83* was actually separate mutations in *let-56* and *unc-22*, or it was a deficiency affecting both *let-56* and *unc-22*, and deletion of the coding elements between them failed to alter the phenotype from that observed for *let-56 unc-22* animals. In order to ascertain whether or not *sDf83* was in fact a deficiency, it was necessary to determine if sequences between *let-56* and *unc-22* were deleted . For this reason, *sDf83* was included with the 62 Unc-22 strains for PCR analysis.

Identification of *unc-22* deficiencies using PCR: The 62 Unc-22 strains, and *sDf83* homozygotes, were initially tested with a primer pair specific for a region 5' of the *gtl-1* gene, which is the first gene 3' of *unc-22* (see Figure 5 for the relative positions of the genes 3' of *unc-22* and the primers that were used). If the *unc-22* mutations were internal to the *unc-22* gene, then PCRs conducted with Unc-22 individuals would be expected to produce a PCR product of the predicted size. However, if any of the *unc-22* mutations were deficiencies that extended 3' of the *unc-22* gene, deleting the *gtl-1* gene and at least one of the primer annealing sites, then the predicted PCR product would not be produced. Of the 62 viable Unc-22 strains tested with the *gtl-1* specific primers, only one, isolated in the JES1 screen, failed to produce the expected product (Figure 6A). This indicated that the mutation carried by this strain was a deficiency affecting at least *unc-22* and the *gtl-1* gene. This mutation was designated *sDf88*. In addition, *sDf83* homozygous individuals did not produce the expected PCR product with the *gtl-1* primers, demonstrating that *sDf83* was a deficiency.

Characterization of *sDf83* and *sDf88*: We performed additional PCRs with primer pairs specific for the *ptl-1*, *dup* and *adl-1* genes. PCRs using *sDf88* homozygous individuals failed to produce the expected amplification product with *ptl-1* primers (Figure 6B), but did produce the expected products with the *dup* (Figure 6C) and *adl-1* primers. This indicated that *sDf88* extended left of *unc-22* at least as far as the annealing site for the Mys-1 primer (Figure 5). Therefore, in addition to deleting *gtl-1*, *sDf88* also disrupted the *ptl-1* gene. The fact that *sDf88* individuals are viable and fertile suggests that the *gtl-1* and *ptl-1* genes are not essential under normal laboratory culture conditions. As expected, PCRs using *sDf83* homozygous individuals failed to produce amplification products with the *adl-1* (Figure 6D), *dup* and *ptl-1* primers.

Figure 6

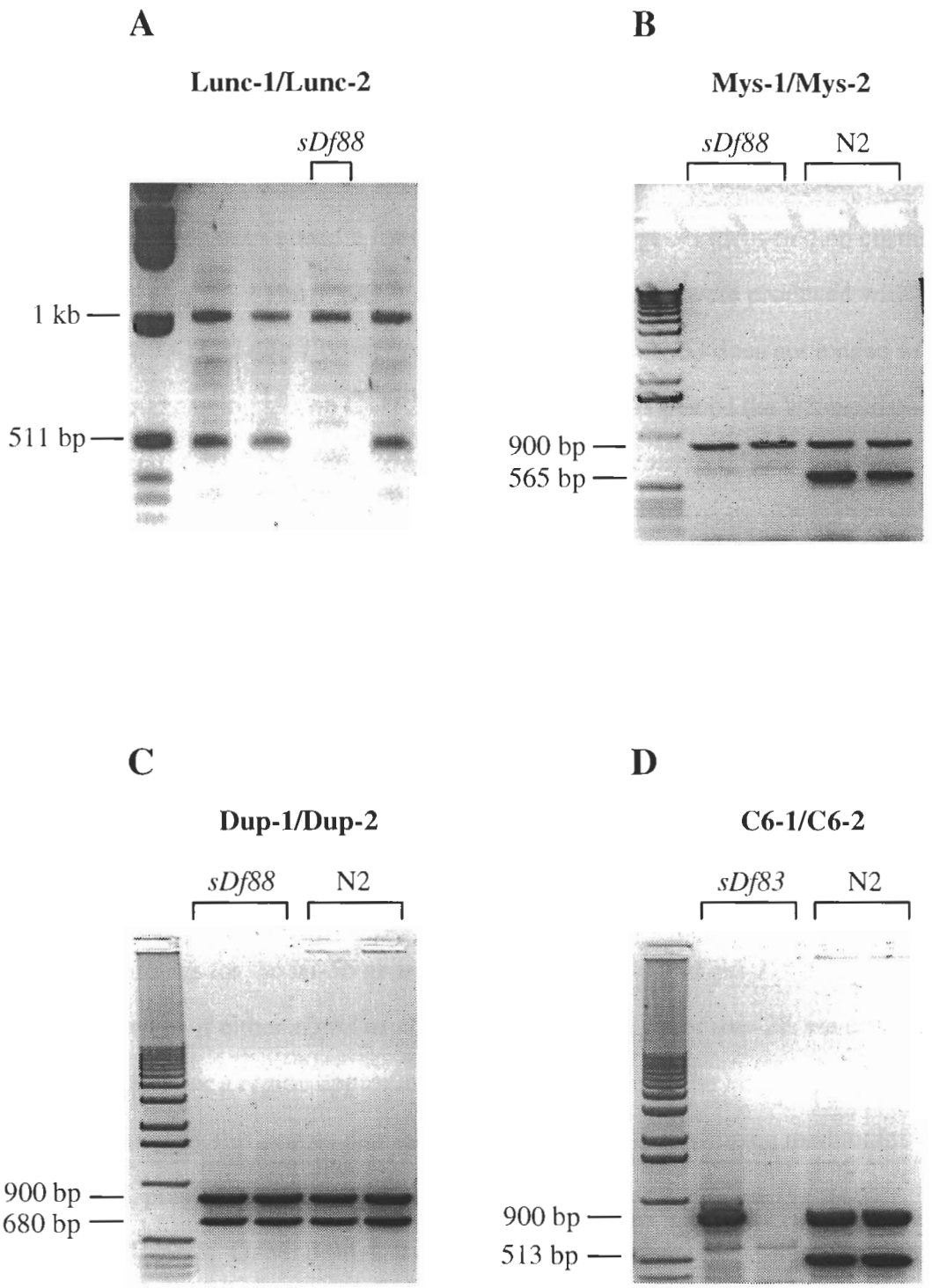
Agarose gels of PCRs conducted with *sDf88* or *sDf83* homozygous individuals using primer pairs specific for coding elements identified on cosmid C18D3. The primer pairs used in the PCRs are indicated above the gels. The positions of the primer pairs relative to coding elements on C18D3 are shown in Figure 5. Reactions conducted with wild-type (N2) worms were used as positive controls for the amplification products expected if no primer annealing sites were affected. The bands migrating at either approximately 1 kb or 900 bp correspond to control amplification products produced by primers specific for sequences on chromosome V (either 265-F and 265-R, or K08-3 and K08-7; Table 2).

(A). Identification of *sDf88* by PCR. DNA from viable Unc-22s were used in PCRs with the Lunc-1/Lunc-2 primer pair, specific for a region approximately 500 bp upstream of the *gtl-1* gene. The 511 bp band corresponds to the product produced by the Lunc-1/Lunc-2 primer pair. This product is not present in the reaction performed with DNA from individuals homozygous for the *sDf88* mutation, indicating that this mutation deletes the *gtl-1* gene in its entirety. Unlabeled lanes are reactions performed with other viable Unc-22 mutants.

(B). *sDf88* deletes *ptl-1*. PCR reactions performed with *sDf88* homozygous individuals and the *ptl-1* specific primers. The 565 bp band in the N2 lanes corresponds to the Mys-1/Mys-2 amplification product . This product is not present in the *sDf88* lanes, indicating that *sDf88* deletes at least the Mys-1 annealing site (see Figure 5).

(C). *sDf88* does not delete *dup*. PCR reactions performed with *sDf88* homozygous individuals and *dup* specific primers. The 680 bp band corresponds to the Dup-1/Dup-2 amplification product.

(D). *sDf83* is a deficiency affecting sequences between *let-56* and *unc-22*. Results of PCR reactions with *adl-1* specific primers are shown. The 513 bp band corresponds to the C6-1/C6-2 amplification product. The second *sDf83* lane is a failed reaction, as evidenced by the absence of the chromosome V specific amplification product.


A
Lunc-1/Lunc-2
sDf88
1 kb
511 bp
B
Mys-1/Mys-2
sDf88
N2
900 bp
565 bp
C
Dup-1/Dup-2
sDf88
N2
900 bp
680 bp
D
C6-1/C6-2
sDf83
N2
900 bp
513 bp

Since *sDf83* fails to complement *let-56*, and *let-56* is contained within sequences on cosmid C11F2 (Clark 1990; Clark and Baillie 1992), we concluded that *sDf83* must extend into the region of the genome represented in cosmid C11F2. Therefore, in order to map the left breakpoint of *sDf83*, we performed PCRs using *sDf83* homozygous individuals and primer pairs specific for the *hdl-1* and *nhe-1* genes identified on cosmid C11F2 (Marra et al. 1993; Marra 1994). The expected products were produced with both the *hdl-1* (Figure 7A) and *nhe-1* primer pairs, indicating that *sDf83* does not extend as far as the R1 primer annealing site on C11F2 (see Figure 5). This placed the left breakpoint of *sDf83* between the R1 annealing site and the right end of cosmid C11F2, a distance of approximately 10 kb. The position of this breakpoint was subsequently refined by M. Marra, who identified restriction fragment length polymorphisms associated with DNA from *sDf83* heterozygotes on Southern blots probed with an *hdl-1* cDNA, placing the left breakpoint of *sDf83* within 200 bp to the right of the R1 annealing site (Marra 1994). The *sDf83* breakpoint therefore lies near the 5' region of the *hdl-1* gene, and consequently *sDf83* must completely delete the *spm-1* and *gdl-1* genes, which reside between *hdl-1* and the right end of cosmid C11F2 (Figure 5). The positioning of the left breakpoint of *sDf83* to this small interval has eliminated the *nhe-1* gene as a potential candidate for *let-56*. Remaining candidates for the *let-56* gene include *hdl-1*, *spm-1* and *gdl-1*.

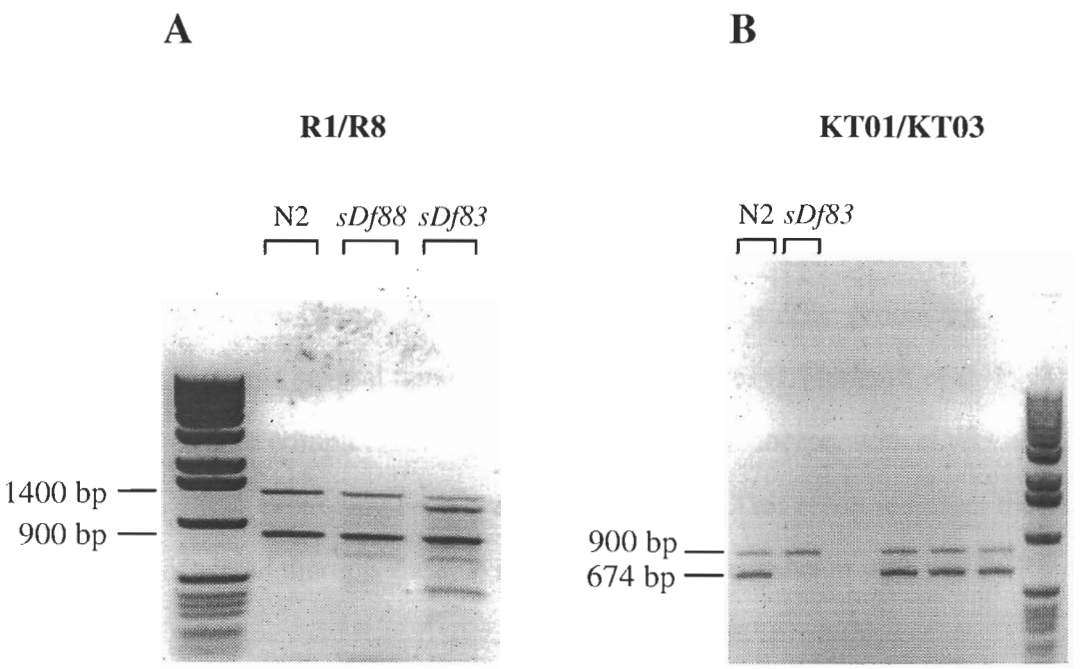
To determine if either *sDf83* or *sDf88* extend to the right of *unc-22*, we used a primer pair specific for a region approximately 2 kb upstream of the *unc-22* gene. The DNA encompassed by this primer pair includes 288 bp 5' of the initiating methionine of the *spe-17* gene, plus all but the last 137 bp of the *spe-17* gene itself (see Figure 5) (L'Hernault et al. 1993; G. M. Benian and S. W. L'Hernault, personal communication). PCRs using *sDf88* homozygous individuals produced the amplification product expected with this primer pair, while PCRs using *sDf83* homozygous individuals did not (Figure 7B). Since *sDf88* does not delete these annealing sites, the right breakpoint of *sDf88* must lie either within the *unc-22* gene, or between the *unc-22* gene and the KT03 annealing

Figure 7

Agarose gels of PCRs conducted to map the left and right breakpoints of *sDf83*. The primer pairs used in the reactions are indicated above the gels. The relative positions of the primer pairs are shown in Figure 5. Reactions conducted with wild-type (N2) worms were used as positive controls for the amplification products expected if no primer annealing sites were affected. The 900 bp band corresponds to control amplification products produced by the chromosome V specific primer pair K08-3 and K08-7.

(A). *sDf83* does not extend as far as the R1 annealing site within the *hdl-1* gene. The 1400 bp product corresponds to the R1/R8 amplification product. This product is present in the *sDf83* lane, indicating that *sDf83* does not extend as far as the R1 primer annealing site within the 5' portion of the *hdl-1* gene.

(B). *sDf83* extends to the right of the *unc-22* gene. The 674 bp band corresponds to the KT01/KT03 amplification product. This product was not observed in reactions using *sDf83* homozygous individuals as template, indicating that *sDf83* extends to the right of *unc-22* at least as far as the KT03 annealing site and therefore likely affects the *spe-17* gene. Unlabelled lanes are reactions performed with other viable Unc-22 mutants isolated in this work.


A
R1/R8
N2
sDf88
sDf83
1400 bp
900 bp
B
KT01/KT03
N2
sDf83
900 bp
674 bp

site. *sDf83* deletes one or both of these primer annealing sites and therefore extends to the right of *unc-22* at least as far as the KT03 annealing site, but as a consequence of complementation test results cannot extend as far as *let-52*. This places the right breakpoint of *sDf83* between the KT03 annealing site and *let-52*.

Given that *sDf83* affects the annealing site of the KT03 primer, I wished to determine whether *sDf83* also affects *spe-17*. To do this, I tested *sDf83* for complementation with *hDf13*, one of two *unc-22* deficiencies that genetically identify the *spe-17* locus (Shakes and Ward 1989; L'Hernault et al. 1993). Null mutations in *spe-17* cause self-sterility in hermaphrodites due to defects in spermatogenesis (Shakes and Ward 1989). *hDf13* deletes the all of the *unc-22* gene as well as *spe-17*, and hermaphrodites homozygous for *hDf13* are phenotypically Unc-22 but are self-sterile, producing on average three progeny and several hundred unfertilized oocytes (L'Hernault et al. 1993). I found that *sDf83* fails to complement *hDf13* for the sterile phenotype, with *sDf83/hDf13* Unc-22 hermaphrodites producing an average of 3 progeny. This result indicates that the right breakpoint of *sDf83* disrupts the *spe-17* gene, which is consistent with our PCR results.

***sEx25* restores partial fertility to *sDf83* homozygous individuals:** The observation that *sDf83* homozygous individuals display a phenotype indistinguishable from that of *let-56 unc-22* homozygous individuals suggests that either the genes on cosmid C18D3 and C11F2 that are deleted by *sDf83* are not required in development prior to the requirement for *let-56*$^+$ activity, or that they provide a dispensable function. PCR analysis of *sDf88* had determined that the *gtl-1* and *ptl-1* genes were dispensable. To test whether any of the additional genes deleted by *sDf83* on cosmid C18D3 were essential for viability, up to but not including fertility (due to disruption of the *spe-17* gene by *sDf83*), M. Marra constructed *sDf83* homozygous animals bearing an extrachromosomal array, *sEx25*, that contains the rightmost 12 kb of cosmid C11F2 (Marra 1994). This extrachromosomal

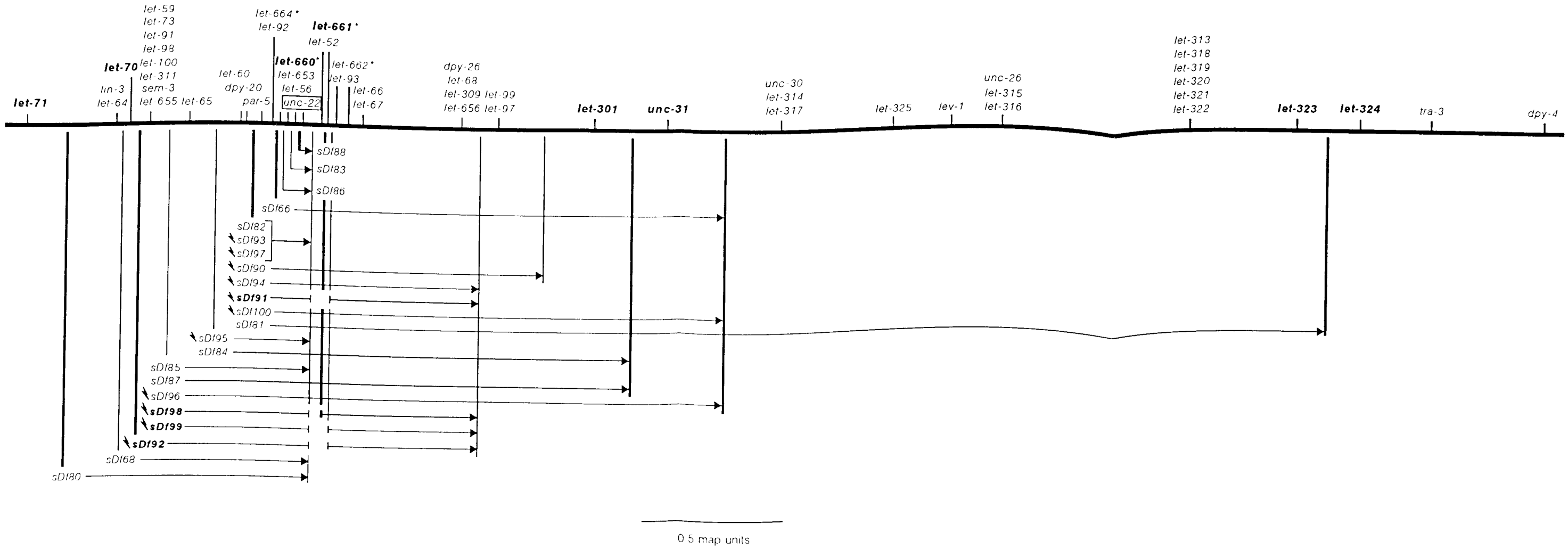
array has been shown to rescue individuals homozygous for *let-56(s173)* (S. Jones and M. Marra, personal communication). *sEx25* does not contain any of the five genes identified on cosmid C18D3. Fourteen individuals of the genotype *sDf83/sDf83*; *sEx25* were examined, and were found to be viable but poorly fertile, producing varying numbers of progeny (0 - 27) and unfertilized oocytes (0 - 40) (Marra 1994). This phenotype is similar to the phenotype described above for *spe-17* null mutants. This result suggests that, under laboratory culture conditions, the five genes identified on cosmid C18D3 are not required for viability.

Complementation mapping of *unc-22* deficiencies identifies eight new breakpoints: The breakpoints of the remaining 20 *unc-22* deficiencies were mapped genetically by complementation tests with mutations in the *unc-22* region. However, not all genes lying between the breakpoints of each deficiency were tested. The results of complementation tests performed with each deficiency are presented in tabular form in the Appendix. A deficiency map depicting the extent of all the deficiencies which I isolated and characterized, including *sDf83* and *sDf88*, is shown in Figure 8.

The mapping of these 20 deficiencies identified eight new genetic breakpoints. With these new breakpoints, I have moved eight genes from zones containing two or more genes to zones in which they are the sole occupants (indicated by bold lettering in Figure 8). Included in these eight genes are two newly identified genes, *let-660* and *let-661* (Marra 1994). Previously existing deficiencies could not resolve the position of *let-660* with respect to *let-92* and *let-664*, or the position of *let-661* with respect to *let-52*. The new deficiencies I isolated have therefore proven useful not only for refining the positions of previously existing mutations in the *unc-22* region, but also for the mapping of newly isolated mutations.

Figure 8

Genetic map of the right end of chromosome IV, showing *unc-22* deficiencies isolated in this work. The heavy horizontal line represents the genetic map. Mutationally identified genes in the region are shown above this line, and deficiencies isolated in this work are shown below it. Gene names in bold lettering indicate genes that have been positioned by this work into zones in which they are the sole occupants. Asterisks indicate genes that have been newly isolated (Marra 1994). Vertical lines below the heavy horizontal line indicate deficiency breakpoints. Heavy vertical lines indicate new deficiency breakpoints. Thin horizontal lines represent the extents of the deficiencies, with the right breakpoint indicated with an arrow. *sDf83*, *sDf86* and *sDf88* are indicated beside their right breakpoint. All other deficiencies are indicated beneath their left breakpoint. Lightning bolts denote deficiencies induced by UV radiation. Deficiencies with complementation islands (see text) are indicated with bold lettering. The complementation island of *sDf98* covers *let-52* but not *let-661*, thereby placing *let-52* and *let-661* in separate zones. The placement of *let-661* in the deleted region to the right of the *sDf98* complementation island, as opposed to the deleted region to the left, was accomplished by correlation of *let-52* and *let-661* to the physical map (see Part 2).


let-71
let-70
lin-3
let-64
let-59
let-73
let-91
let-98
let-100
let-311
sem-3
let-655
let-65
let-60
dpy-20
par-5
let-664 *
let-92
let-660 *
let-653
let-56
unc-22
let-661 *
let-52
let-662 *
let-93
let-66
let-67
dpy-26
let-68
let-309 let-99
let-656 let-97
let-301
unc-31
unc-30
let-314
let-317
let-325
lev-1
unc-26
let-315
let-316
let-313
let-318
let-319
let-320
let-321
let-322
let-323
let-324
tra-3
dpy-4
sDf88
sDf83
sDf86
sDf66
sDf82
sDf93
sDf97
sDf90
sDf94
sDf91
sDf100
sDf81
sDf95
sDf84
sDf85
sDf87
sDf96
sDf98
sDf99
sDf92
sDf68
sDf80
0.5 map units

***sDf91, sDf92, sDf98* and *sDf99* have an "island":** Of the eleven UV radiation induced deficiencies (indicated with a lightning bolt in Figure 8), four appear to be more complex rearrangements. These deficiencies, *sDf91*, *sDf92*, *sDf98* and *sDf99*, exhibit a small "island" of complementation within the deleted region. Although all four were isolated from different P_0 plates (i.e. they were independently derived), in all four cases the island lies between *unc-22* and *let-93* (Figure 8). This interval contains two genes, *let-52* and *let-661*. With three of the four deficiencies, *sDf91*, *sDf92* and *sDf99*, the island covers both *let-52* and *let-661*, while with the fourth deficiency, *sDf98*, only *let-52* is covered. I did not observe this phenomenon with any of the formaldehyde induced deficiencies that I isolated, nor was it found with any of the 15 previously existing *unc-22* deficiencies, all but two of which (*mDf7* and *nDf27*) were induced with formaldehyde. This island phenomenon may therefore reflect the type of DNA damage caused by UV radiation and/or the mechanism(s) involved in repair of this damage in *C. elegans* germ cell nuclei. The fact that the islands in all four deficiencies are localized to the same small region suggests that this phenomenon is also due in part to properties inherent in the DNA immediately to the right of *unc-22*. This topic is addressed further in the Discussion.

The *unc-22* - *let-52* interval is a hot spot for *unc-22* deficiency breakpoints: Consistent with the notion that the DNA immediately to the right of *unc-22* has unusual properties is the observation that a large percentage of *unc-22* deficiencies have genetic breakpoints in the *unc-22* - *let-52* interval. Fourteen of the 22 deficiencies I isolated, including *sDf91*, *sDf92*, *sDf98* and *sDf99*, have right breakpoints that map genetically between *unc-22* and *let-52* (Figure 8). If previously isolated *unc-22* deficiencies are included, a total of 19 out of 37 deficiencies (51%) have right breakpoints which map genetically to this interval. However, the result of a complementation test performed with an *unc-22* deficiency and an *unc-22* linked *let-52* mutation would not distinguish between a deficiency that breaks within the *unc-22* gene and one that breaks in the interval

between *unc-22* and *let-52* since in both cases fertile Unc-22 progeny would be recovered. The apparent clustering of genetic breakpoints would therefore also be observed if the deficiencies actually broke within the *unc-22* gene which, at 38 308 bp (Benian et al. 1989), is physically very large.

The possibility that the right breakpoints of these 19 deficiencies lie within the *unc-22* gene, as opposed to extending into the *unc-22 - let-52* interval, was investigated by determining whether any of the deficiencies affect *spe-17*, which lies between *unc-22* and *let-52*. This was accomplished either by performing PCRs with *spe-17* gene-specific primers or by complementation tests with *hDf13*, which deletes *spe-17*. PCR results described above with *sDf83* and *sDf88*, using the *spe-17* gene-specific primers, showed that *sDf83* deleted at least the KT03 annealing site and consequently must extend into the *unc-22 - let-52* interval. M. Marra conducted additional PCRs with these primers, with the remaining five JES1 deficiencies having genetic breakpoints in the *unc-22 - let-52* interval (*sDf68*, *sDf80*, *sDf82*, *sDf85* and *sDf86*), and four of the five previously isolated *unc-22* deficiencies with genetic breakpoints in the interval (*sDf8*, *sDf63*, *sDf64* and *mDf7*) (Marra 1994). The other previously identified *unc-22* deficiency with a genetic breakpoint in this interval, *sDf19*, has been shown to break in the 3' end of the *unc-22* gene (Benian et al. 1989). M. Marra's PCR results showed that, with the exception of *mDf7* and *sDf64*, all the deficiencies he tested had breakpoints in the interval between *unc-22* and *let-52* (Marra 1994).

I performed complementation tests against *hDf13* with the seven UV radiation-induced deficiencies that break genetically in the *unc-22 - let-52* interval (*sDf91*, *sDf92*, *sDf93*, *sDf95*, *sDf97*, *sDf98* and *sDf99*). *hDf13* deletes all of the *unc-22* gene as well as the *spe-17* gene (L'Hernault et al. 1993). The *spe-17* gene is contained within a 1 kb region that lies approximately 2 kb 5' of the *unc-22* gene, and the Spe-17 phenotype of *hDf13* homozygous animals is rescued by an extrachromosomal array containing this 1 kb of sequence (L'Hernault et al. 1993). I reasoned, therefore, that an *unc-22* deficiency that

extends 5' of the *unc-22* gene and affects this region would fail to complement *hDf13* for the Spe-17 phenotype. With the exception of *sDf95*, all of the UV radiation induced deficiencies I tested failed to complement *hDf13* for the Spe-17 phenotype. Therefore, of the 19 deficiencies with genetic breakpoints between *unc-22* and *let-52*, 14 affect sequences in this interval as well.

Given that the interval between *unc-22* and *let-52* appears to be a hot spot for deficiency breakpoints, the question arises as to why this is so. Is the DNA in this region more susceptible to breakage, or is there simply a relatively large physical distance between the two genes within which many deficiency breakpoints fall? Recombinationally, *let-52* maps approximately the same distance to the right of *unc-22* as *let-56* does to the left of *unc-22* (0.07 vs. 0.03 map units, respectively) (Rogalski et al. 1982). Since *let-56* lies approximately one cosmid length from *unc-22*, these map data suggest that *let-52* should lie a similar physical distance from *unc-22*. However, calculated map distances may be subject to localized differences in recombination frequencies within the genome. Correlation of the genetic and physical maps in the *lin-12(III)* region (Greenwald et al. 1987) has supported the proposal that gene clusters on the autosomes result from decreased recombination frequency within the clusters, as opposed to a nonrandon distribution of genes along the chromosome. Distances calculated in map units, therefore, do not directly translate into physical distances. Indeed, the region to the right of *unc-22* appears relatively gene sparse with respect to the region to the left of *unc-22* (Figure 8). It remains to be determined whether the apparent sparseness of genes to the right of *unc-22* reflects a relatively low gene density, or a relatively increased rate of recombination, with respect to the region to the left of *unc-22*. The question of the physical distance separating *unc-22* and *let-52*, then, can only be answered by determining the location of the *let-52* gene on the physical map. This is addressed in Part 2 of my thesis, which describes the correlation to the physical map of three genes immediately to the right of *unc-22*, including *let-52*.

Part 2

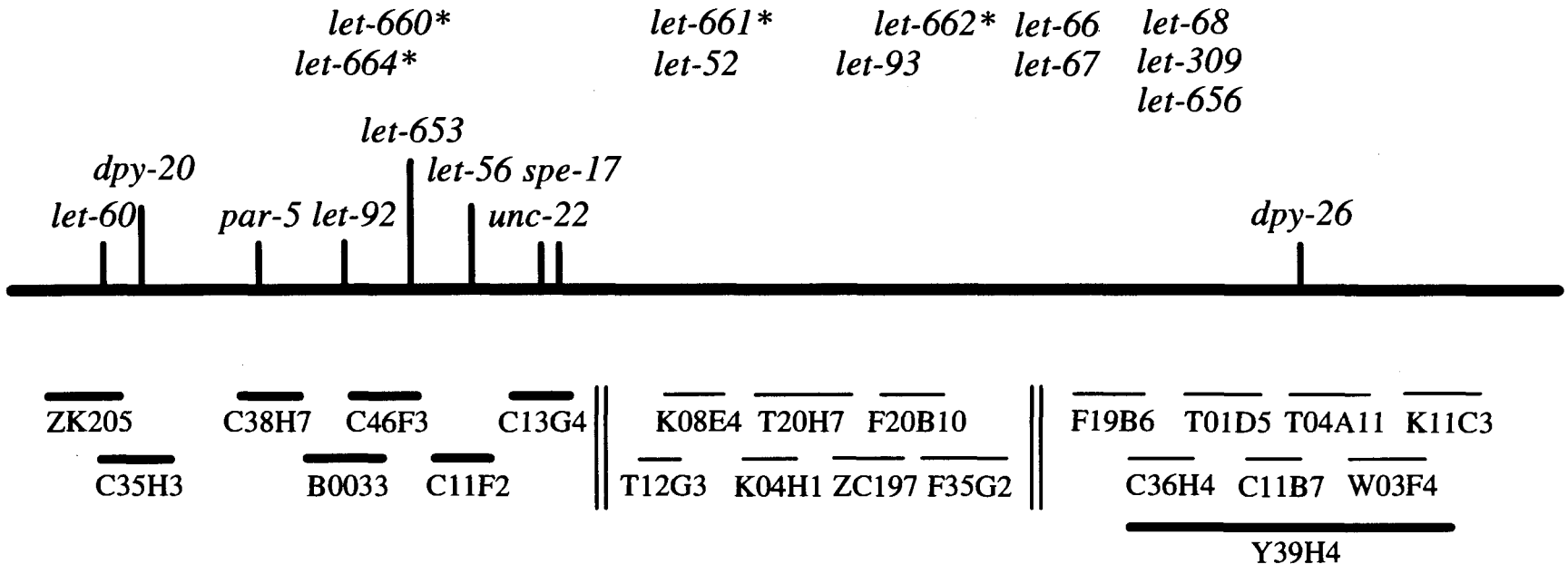
Physical Mapping of Genes to the Right of *unc-22*

Physical map to the right of *unc-22*: In contrast to the region immediately to the left of *unc-22*, relatively little is known about the relationship of the genetic and physical maps to the right of *unc-22*. Perusal of ACEDB (R. Durbin and J. Thierry-Mieg, unpublished) shows that eight mutationally identified genes to the right of *unc-22* have been correlated to the physical map, but these genes are spread out over approximately seven map units. The nearest mutationally identified gene to the right of *unc-22* that has been placed on the physical map is *dpy-26*, which was rescued with a YAC (B. Capowski, personal communication). In comparison, six mutationally identified genes have been correlated with the physical map in the 0.2 map units immediately to the left of *unc-22*, in the *let-60* - *unc-22* interval (Figure 9). These data indicate an average of one essential gene per cosmid in the *let-60* - *unc-22* interval. Genetically, fewer genes have been identified per map unit to the right of *unc-22* than to the left (J. Hodgkin, R. Durbin and M. O'Callaghan, personal communication; see Figure 8), which suggests that *unc-22* is near the rightmost edge of the gene cluster on chromosome IV.

The physical map to the right of *unc-22* contains numerous YAC bridges, which are regions in which two adjacent cosmid contigs have been joined only by overlapping YAC clones. Contig gaps generally occur in regions where the genomic DNA is not cloneable in available cosmid vectors (Coulson et al. 1991), although approximately one third of these gaps were found to be artifacts resulting from short contig overlaps that had not been detected (Coulson et al. 1991). The first of these YAC bridges lies immediately to the right of the *unc-22* gene, separating *unc-22* from a small cosmid contig delineated by the cosmids T12G3 and F35G2. This small contig is joined to the *dpy-26* contig by another YAC bridge (Figure 9). Based on the essential gene density per cosmid to the left of *unc-22*, and the position of the *dpy-26* contig relative to *unc-22*, I chose the small contig to the right of the *unc-22* gene to begin the search for cosmids capable of rescuing the lethal phenotypes of mutations immediately to the right of *unc-22*.

Figure 9

A portion of the physical and genetic maps in the *let-60* - *dpy-26* interval prior to this study (adapted from ACEDB), showing the correlation between the genetic and physical maps. The heavy line represents the genetic map. Genes that have been placed on the physical map by germline transformation rescue are indicated above vertical lines intersecting the heavy line. Genes without intersecting vertical lines have not been correlated with the physical map. Asterisks indicate newly identified genes (Marra 1994). The physical map is represented with selected cosmid and yeast artificial chromosome (YAC) clones. Clones drawn with a heavy line have been shown to contain the gene(s) directly above them. Note that *let-660* has been genetically mapped between *let-92* and *let-653*, and positioned within genomic DNA encompassing cosmid B0033 and 2.4 kb to the right of B0033 (Jones 1994). Double vertical bars between contigs represent YAC bridges. *dpy-26* was rescued by YAC Y39H4 (B. Capowski, personal communication). *let-60* was rescued by ZK205 (Han and Sternberg 1990), *dpy-20* was rescued by C35H3 (Clark 1990), *par-5* has been tentatively assigned to C38H7 (D. Shakes, personal communication), *let-92* was rescued by B0033 (Jones 1994), *let-653* was rescued by C29E6 (Clark 1990; Clark and Baillie 1992) and C46F3 (Jones 1994), *let-56* was rescued by C11F2 (Clark 1990; Clark and Baillie 1992), and both *unc-22* and *spe-17* sequences are on C13G4 (Benian et al. 1989; L'Hernault et al. 1993).


let-660*
let-664*
let-661*
let-52
let-662*
let-93
let-66
let-67
let-68
let-309
let-656
let-653
dpy-20
let-60
let-56 spe-17
par-5 let-92
unc-22
dpy-26
ZK205
C35H3
C38H7
B0033
C46F3
C11F2
C13G4
K08E4
T12G3
T20H7
K04H1
F20B10
ZC197
F35G2
F19B6
C36H4
T01D5
C11B7
T04A11
W03F4
K11C3
Y39H4
0.1 mu or 100 kb

Construction of transgenic strains bearing extrachromosomal arrays: I obtained seven cosmid clones representative of the contig immediately to the right of the *unc-22* gene from the Sanger Centre in Cambridge, England (cosmids T12G3, K08E4, K04H1, T20H7, ZC197, F20B10 and F35G2). The positions of these cosmids are shown in Figure 9. Transgenic strains bearing these cosmids in stably inherited extrachromosomal arrays were constructed by injection into wild-type hermaphrodites (Table 4). Two transgenic strains were established for each cosmid, with the exception of ZC197 for which only one transgenic strain was retained. The extrachromosomal arrays in each strain are essentially free duplications, and transgenic strains carrying these arrays can be used in genetic crosses. The cosmids were tested for the ability to rescue the lethal phenotype of individuals homozygous for mutations in the *let- 52 - let-662* interval by performing complementation tests between transgenic strains and strains carrying the lethal mutations. The protocol is described in Materials and Methods. The rescue data are described below, and summarized in Table 5.

The rescue of *let-52*: *unc-22(s7) let-52(s42)* homozygous individuals were rescued by an extrachromosomal array bearing cosmid T12G3 (*sEx57*; see Table 5), indicating that *let-52*$^{+}$ gene sequences are present on this cosmid. T12G3 is the nearest cosmid to the right of *unc-22* (Figure 9). An extrachromosomal array carrying cosmid K08E4, which overlaps with the right end of cosmid T12G3 and extends to the right, did not rescue *unc-22(s7) let-52(s42)* homozygous individuals. This suggests that *let-52* gene sequences likely lie on the left end of cosmid T12G3, but may lie across the area of overlap between T12G3 and K08E4.

Expression from the *sEx57* array appeared to be at the threshold at which rescue could be achieved. *unc-22(s7) let-52(s42); sEx57* individuals were sickly and slow developing. Many were sterile, and those that were fertile had very small broods (1 - 15 progeny after 5 - 12 days). It is also possible that expression of other sequences on

Table 4

Construction of transgenic strains carrying extrachromosomal arrays[a]

Cosmid	[Cosmid]	[pCes1943]	Array[b]	Strain
F20B10	10 ng/µl	90 ng/µl	*sEx51*	BC4624
			sEx52	BC4640
F35G2	5 ng/µl	95 ng/µl	*sEx88*	BC4712
			sEx89	BC4713
K04H1	10 ng/µl	90 ng/µl	*sEx53*	BC4641
			sEx54	BC4642
K08E4	10 ng/µl	90 ng/µl	*sEx55*	BC4643
			sEx56	BC4644
T12G3	10 ng/µl	90 ng/µl	*sEx57*	BC4645
			sEx58	BC4646
T20H7	10 ng/µl	90 ng/µl	*sEx65*	BC4647
			sEx66	BC4648
ZC197	10 ng/µl	90 ng/µl	*sEx75*	BC4659

[a] Cosmids were injected into N2 (wild-type) hermaphrodites from strain BC842.

[b] Presence of cosmid in the extrachromosomal array was confirmed by PCR (see Materials and Methods).

Table 5

Summary of cosmid rescue data

Mutation tested[a]				Cosmid tested[b]			
Strain	**Gene**	**Allele**	**Method[c]**	**Cosmid**	**Array**	**Strain**	**Rescue**
BC2902	*let-52*	*s42*	Cross	F20B10	*sEx52*	BC4640	NO
			Cross	K04H1	*sEx54*	BC4642	NO
			Cross	K08E4	*sEx55*	BC4643	NO
			Cross	T12G3	*sEx57*	BC4645	YES
BC3888	*let-52*	*s2346*	Injection	T12G3[d]			YES[e]
BC1291	*let-93*	*s734*	Cross	F20B10	*sEx51*	BC4624	YES[f]
			Cross	F20B10	*sEx52*	BC4640	YES[f]
			Cross	ZC197	*sEx75*	BC4659	NO
			Cross	F35G2	*sEx89*	BC4713	NO
BC4431	*let-93*	*s2357*	Cross	F20B10	*sEx52*	BC4640	NO
			Cross	T12G3	*sEx57*	BC4645	NO
			Cross	T20H7	*sEx65*	BC4647	NO
			Cross	ZC197	*sEx75*	BC4659	NO
			Cross	F35G2	*sEx89*	BC4713	NO
BC3745	*let-661*	*s2203*	Cross	F20B10	*sEx51*	BC4624	NO
			Cross	F20B10	*sEx52*	BC4640	NO
			Cross	K04H1	*sEx54*	BC4642	YES[e]
			Cross	K08E4	*sEx55*	BC4643	NO
			Cross	K08E4	*sEx56*	BC4644	NO
			Cross	T20H7	*sEx65*	BC4647	NO

Table 5 (continued)

Mutation tested[a]				Cosmid tested[b]			
Strain	**Gene**	**Allele**	**Method[c]**	**Cosmid**	**Array**	**Strain**	**Rescue**
BC3761	*let-662*	*s2219*	Cross	F20B10	*sEx51*	BC4624	NO
			Cross	F20B10	*sEx52*	BC4640	NO
			Cross	K04H1	*sEx54*	BC4642	NO
			Cross	K08E4	*sEx55*	BC4643	NO
			Cross	T20H7	*sEx65*	BC4647	NO
			Cross	ZC197	*sEx75*	BC4659	NO
			Cross	F35G2	*sEx88*	BC4712	NO

[a] Strains indicated are hermaphroditic. In the cases where a genetic cross was used to introduce the extrachromosomal array, hermaphrodites of lethal bearing strains were outcrossed to wild-type males, and F_1 males heterozygous for the lethal mutation were mated to hermaphrodites bearing the appropriate extrachromosomal array (see Materials and methods).

[b] In those cases where the extrachromosomal array was introduced via a genetic cross, the names of the parental transgenic strains and the arrays they carry are given.

[c] Indicates whether the extrachromosomal array was introduced via a genetic cross (Cross), or by germline transformation (Injection).

[d] Concentration of DNA injected: T12G3, 10 ng/µl; pCes1943, 90 ng/µl.

[e] Larval lethality was rescued, i.e. the worms developed past the early larval stage to adulthood, but were sterile.

[f] The worms have a few progeny (3 - 12), many of which hatch internally. Progeny may develop, but die before having progeny of their own.

cosmid T12G3, present in multicopies in the extrachromosomal array, has a detrimental effect on the health and/or fertility of the transgenic individuals. The strain in which *sEx57* was constructed is quite healthy and fertile, but appears to develop more slowly than the transgenic strains containing the other cosmids. It is possible, therefore, that the infertility of rescued individuals is due to the expression of sequences present on cosmid T12G3 at levels that are poisonous to the worm.

I also observed that most broods from these rescued worms contained a small number of sterile adult twitchers that did not display the Rol-6 phenotype. The expression from the *sEx57* array was therefore sufficient in these worms to allow development past the early larval arrest phase to that of a sterile adult, but insufficient to produce a rolling phenotype. These sterile, non-rolling Unc-22 animals are most likely mosaic for the extrachromosomal array. Stinchcomb et al. (1985) observed that approximately half of the transgenic worms they examined were mosaic for the extrachromosomal array. This showed that extrachromosomal arrays could be lost at any time during development. The presence of non-rolling Unc-22 adult animals in the progeny of *unc-22 let-52(s42); sEx57* individuals suggests that the *let-52*$^+$ gene product may only be required at low levels of expression, or only early in development. These sterile, non-rolling Unc-22 animals would therefore be individuals in which the extrachromosomal array, and therefore the *let-52*$^+$ gene product, was present in a sufficient number of cells early in development to provide rescue from early larval arrest, but was lost in subsequent cell divisions such that the Rol-6 phenotype was not produced. The sterility of these individuals may also indicate an adult requirement for the *let-52* gene product, which would not be provided in these individuals due to loss of the extrachromosomal array.

Rescue of a second allele of *let-52*, *s2346* (Marra 1994), that also causes developmental arrest in the early larval stage, was achieved by injection into *unc-22 let-52(s2346) lev-1/nT1(IV;V)* hermaphrodites, using the same DNA mixture with which *sEx57* was obtained. A stably transformed line was obtained that segregated adult Rol-6

Unc-22 animals. However, complete rescue was not achieved as the Unc-22 individuals were not fertile.

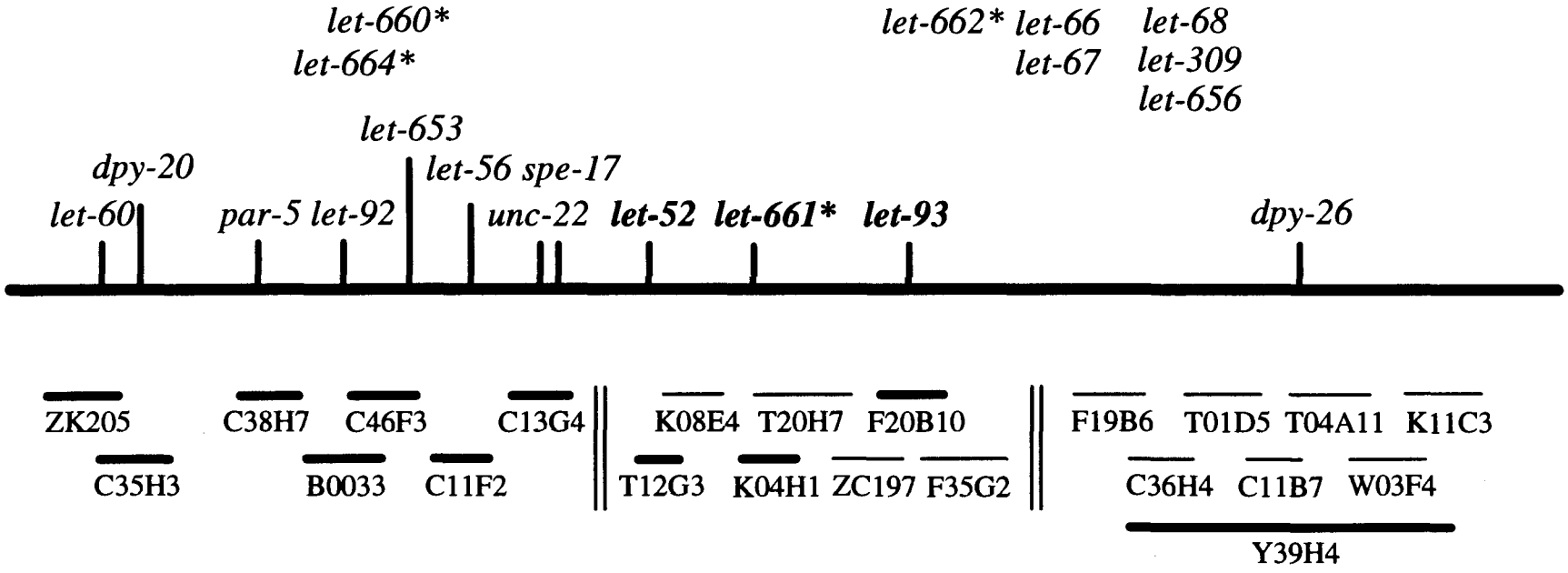
The rescue of *let-661*: The *sEx54* array, carrying cosmid K04H1, was able to rescue *unc-22(s7) let-661(s2203) lev-1(x22)* homozygous individuals from the early larval arrest normally associated with *let-661(s2203)* homozygotes (Table 5). This rescue was not complete, however, as these individuals achieved adulthood but were sterile. Neither *sEx55*, containing cosmid K08E4, nor *sEx65*, containing cosmid T20H7, were able to rescue *let-661(s2203)* homozygotes. The *let-661* gene must therefore be present within the sequences of cosmid K04H1. However, the absence of a complete rescue suggests that either the *let-661* gene is not contained in its entirety on K04H1, or that sequences necessary for proper transcriptional control are either not present or not recognized in the context of the array.

Localization of *let-661* to cosmid K04H1 places it to the right of *let-52*. Prior to my work, *let-661* and *let-52* had been mapped to the same zone, as they could not be separated by existing deficiencies (Marra 1994). One of the UV radiation induced deficiencies I isolated, *sDf98*, complemented *let-52* but not *let-661*. However, because *sDf98* is one of the four deficiencies with an "island" (see Part 1), complementing *let-52* but deleting genes on both sides (Figure 8), I was unable to determine whether *let-661* lay to the left or the right of *let-52*. Correlation of both *let-52* and *let-661* to the physical map has determined that *let-661* lies to the right of *let-52* (Figure 10).

The rescue of *let-93*: *unc-22 let-93(s734) unc-31* homozygous individuals arrest development in the mid larval stage (Clark et al. 1988). When these individuals carry an extrachromosomal array bearing cosmid F20B10 (*sEx51* or *sEx52*) they are able to reach adulthood, although they are small, sickly and develop slowly. They are weakly fertile (1 - 12 progeny) and all progeny appear to hatch internally. The progeny may develop, but

Figure 10

A portion of the physical and genetic maps in the *let-60* - *dpy-26* interval (adapted from ACEDB), illustrating the extent of the correlation between the genetic and physical maps after this study. The heavy line represents the physical map. Genes that have been placed on the physical map are indicated above vertical lines which intersect the heavy line. Genes without intersecting vertical lines have not been correlated with the physical map. Genes correlated to the physical map in this study are indicated with bold lettering. Asterisks indicate newly identified genes (Marra 1994). The physical map is represented with selected cosmid and YAC clones. Double vertical bars between contigs represent YAC bridges. Clones drawn with a heavy line have been shown to contain the gene(s) directly above them. Note that *let-660* has been genetically mapped between *let-92* and *let-653*, and positioned within genomic DNA encompassing cosmid B0033 and 2.4 kb to the right of B0033 (Jones 1994). *dpy-26* was rescued by YAC Y39H4 (B. Capowski, personal communication). *let-60* was rescued by ZK205 (Han and Sternberg 1990), *dpy-20* was rescued by C35H3 (Clark 1990), *par-5* has been tentatively assigned to C38H7 (D. Shakes, personal communication), *let-92* was rescued by B0033 (Jones 1994), *let-653* was rescued by C29E6 (Clark 1990; Clark and Baillie 1992) and C46F3 (Jones 1994), *let-56* was rescued by C11F2 (Clark 1990; Clark and Baillie 1992), and both *unc-22* and *spe-17* sequences are on C13G4 (Benian et al. 1989; L'Hernault et al. 1993).


let-660*
let-664*
let-662* let-66 let-68
let-67 let-309
let-656
let-653
dpy-20
let-56 spe-17
let-60
par-5 let-92
unc-22
let-52
let-661*
let-93
dpy-26
ZK205
C38H7
C46F3
C13G4
K08E4
T20H7
F20B10
F19B6
T01D5
T04A11
K11C3
C35H3
B0033
C11F2
T12G3
K04H1
ZC197
F35G2
C36H4
C11B7
W03F4
Y39H4
0.1 mu or 100 kb

die before having progeny of their own. Extrachromosomal arrays bearing cosmids to the left or right of F20B10 (cosmids ZC197 and F35G2, respectively; see Figure 10) are unable to rescue *let-93(s734)* homozygous individuals (Table 5). These observations indicate that *let-93* sequences are present on cosmid F20B10. However, the fact that *unc-22 let-93(s734) unc-31* individuals carrying *sEx51* or *sEx52* are sickly and either die before reaching adulthood, or reach adulthood but produce few eggs which they are unable to lay before dying, suggests that either the entire gene is not present within cosmid F20B10, or that there is an inadequate level of expression of the *let-93* gene within the array.

Individuals homozygous for another allele, *let-93(s2357)*, arrest development at the L2 stage (Marra 1994), which is earlier than that for *let-93(s734)* homozygotes, suggesting that the *s2357* mutation causes a greater loss of *let-93*$^+$ function than the *s734* mutation. *unc-22 let-93(s2357) lev-1* individuals were not rescued by the *sEx52* array (Table 5). The expression of the *let-93* gene within the *sEx52* array is therefore sufficient for marginal rescue of individuals homozygous for the *s734* allele, but insufficient for rescue of individuals homozygous for the more severe loss of function *s2357* allele.

An average of one essential gene every two cosmids in the *unc-22* - *let-93* interval: Previous transformation rescue data has determined an average of one essential gene per cosmid in the *let-60* - *unc-22* interval (Figure 10). The transformation rescue data that I have obtained indicates an average of one essential gene for every two cosmids in the *unc-22* - *let-93* interval. Although additional rescue data would provide a more accurate estimation, as a first approximation it would appear that the lower density per map unit of genetically identified essential genes to the right of *unc-22*, relative to that to the left of *unc-22*, is at least in part a reflection of a lower physical density of essential genes. This does not rule out the possibility that recombination frequencies are also higher to the right of *unc-22* than to the left.

Discussion

In Part 1 of this thesis, I described the isolation of 22 *unc-22* deficiencies recovered after treatment of wild-type worms with formaldehyde or 254-nm UV light, and the characterization of these deficiencies with respect to both genetically and molecularly identified genes in the *unc-22* region. This analysis has refined the genetic map in the *unc-22* region by placement of eight genetic loci that had previously resided in clusters of two or more unordered genes into zones in which they are the sole occupants. PCR analysis of two of these deficiencies, *sDf83* and *sDf88*, has determined that there are likely no genes essential for development in the *let-56* - *unc-22* interval. This has shown that, of eight genes that have thus far been identified molecularly in this 50 kb interval, at least five and as many as seven do not supply an indispensable function under laboratory conditions.

In Part 2, I described the correlation of three genes, *let-52*, *let-661* and *let-93*, to the physical map. These genes are the three nearest mutationally identified genes to the right of *unc-22*, and their placement on the physical map is therefore the initial step in the systematic correlation of the genetic and physical maps to the right of the *unc-22* gene. The analysis of this data suggests that, at one essential gene per every two cosmids, the density of essential genes immediately to the right of *unc-22* is half that immediately to the left of *unc-22*.

This thesis provides the first description of *unc-22* deficiencies isolated following UV mutagenesis. Of the 15 previously existing *unc-22* deficiencies, all but two were recovered following formaldehyde mutagenesis. The remaining two, *mDf7* and *nDf27*, were recovered following gamma irradiation. It is therefore of interest to compare the breakpoint distribution of deletions recovered following UV mutagenesis with those recovered following formaldehyde mutagenesis, although any conclusions drawn from this comparison are somewhat speculative due to the sample size.

From Figure 8, it is apparent that the genetic breakpoints of the 11 UV radiation induced deficiencies I isolated show a more biased distribution than that of the 11

formaldehyde induced deficiencies I isolated. Many of the UV radiation induced deficiencies share common genetic breakpoints, while the formaldehyde induced deficiencies show a greater distribution of breakpoints (with the exception of the *unc-22 - let-52* interval, which is discussed below). The 15 previously existing *unc-22* deficiencies also show a greater distribution of breakpoints relative to those of the UV radiation induced deficiencies. These observations may reflect a greater bias in the distribution of UV radiation induced DNA damage relative to that caused by formaldehyde. The two primary types of photolesions induced in DNA by UV radiation are cyclobutane pyrimidine dimers and pyrimidine(6-4)pyrimidone photoproducts. Hot spots for these photolesions have been demonstrated to lie preferentially within pyrimidine runs (Brash and Haseltine 1982; Sage et al. 1992; Sage 1993). *C. elegans* is unable to repair UV radiation induced pyrimidine dimers by photoreactivation (Keller et al. 1987; Hartman et al. 1989). However, both excision repair of the two types of photoproducts, and the ability to replicate through these non-instructional lesions (trans-lesion synthesis), have been demonstrated in *C. elegans* embryos (Hartman et al. 1989; Hartman et al. 1991). In terms of the mutagenic potential of the photolesions, UV radiation has been found to predominantly cause point mutations in other systems studied (reviewed in Sage 1993). In *C. elegans*, on the other hand, the majority of UV radiation induced mutations were found to be chromosomal rearrangements such as deletions, duplications and translocations (Stewart et al. 1991). However, the mechanism by which *C. elegans* mis-repairs UV radiation induced photolesions into chromosomal rearrangements is unknown.

The mechanism by which formaldehyde causes mutations is not well understood (Auerbach et al. 1977). Formaldehyde is known to undergo condensation reactions with amino groups of purines and proteins (Feldman 1973). Auerbach et al. (1977) suggest that, in regard to the mutagenic effect of formaldehyde, the reaction with the amino groups of purines in locally denatured regions of DNA is especially significant, since this

can result in dimers formed by methylene cross-links between neighbouring purines in a strand of DNA. Others (Poverenny et al. 1975; Magaña-Schwencke and Ekert 1978) suggest that it is the cross-linking of purines in DNA with DNA-associated proteins that is the primary mutagenic lesion. In *E. coli* and *S. cerevisiae*, an excision repair mechanism responsible for the removal of UV radiation induced pyrimidine dimers appears to also be responsible for repair of formaldehyde induced DNA damage (Nishioka 1973; Magaña-Schwenke et al. 1978). Therefore, although it is not known how the DNA damage caused by either of these mutagens is mis-repaired to cause deletions, the difference in the sites of action of formaldehyde and UV radiation may in large part explain the difference in the distribution of deficiency breakpoints observed. The greater bias of breakpoints found with UV radiation induced deficiencies suggests that UV radiation induced lesions are more dependent upon sequence context than are formaldehyde induced lesions.

One surprising result from the mapping of the 22 deficiencies I isolated is the island of complementation falling between *unc-22* and *let-93*, found with four of the UV radiation induced deficiencies. Three of these four deficiencies complement both genes in this interval, *let-52* and *let-661*, while the fourth complements only *let-52*. This island phenomenon was not observed with any of the formaldehyde induced deficiencies I isolated, nor with any of the pre-existing formaldehyde or gamma radiation induced *unc-22* deficiencies. Therefore it appears that the formation of these islands is peculiar to UV mutagenesis.

An explanation for the island phenomenon comes from observations of high frequencies of coincident mutations following UV irradiation in *E. coli* (Kubitschek 1980) and *B. subtilis* (Kubitschek and Venema 1976). Each of the four deficiencies could therefore be alternatively viewed as two separate deficiencies, one deficiency deleting *unc-22* and extending to the left and the other deleting *let-93* (or *let-661* in the case of *sDf98*) and extending to the right, with a region of undeleted DNA between them. This

may not be the first case of coincident mutagenesis in *C. elegans* following UV irradiation. In a screen for deletions affecting the *fem-3(IV)* gene and at least one neighbouring essential gene, the lethality associated with the putative deficiencies recovered was later found to have recombined away from the *fem-3* mutation, an observation that was explained by the induction of coincident mutations, one affecting *fem-3* and another affecting at least one linked essential gene (P. S. Hartman, D. De Wilde and V. N. Dwarakanath, personal communication). However, the fact that the region of undeleted DNA is nearly identical in all four deficiencies, and that the right breakpoint of the deleted region to the right of the island is identical in all four, indicates that this is not likely a simple case of coincident mutagenesis, and may be due in part to properties inherent in the DNA to the right of *unc-22*.

A second possible explanation for the observed phenomenon is the occurrence of a spontaneous duplication in the wild-type strain prior to UV mutagenesis. A duplication covering the *let-52 - let-661* region would give the appearance of an island of complementation in an otherwise deleted region. However, since only four out of eight UV radiation induced deficiencies which extend through this region display the island phenomenon, such a duplication was either not homogeneous within the population of worms that were originally mutagenized, or was lost in some of the recovered mutants but not in others. In addition, in order to account for the fact that *sDf98* fails to complement *let-661*, the duplication carried by the *sDf98* strain would have to have sustained a mutation affecting the *let-661* gene. For the purpose of the remainder of the discussion, I will consider the four island-containing deficiencies to be the result of coincident mutagenesis.

In contrast to the general observation that formaldehyde and UV radiation induced *unc-22* deficiencies have different breakpoint distributions is the observed clustering of both formaldehyde and UV radiation induced deficiency breakpoints in the *unc-22 - let-52* interval. Twelve of the 22 deficiencies I isolated have a breakpoint in this interval.

There is apparently no mutagen bias for breakpoints in this region as six of the 12 deficiencies were formaldehyde induced and six were UV radiation induced. The latter number includes the four deficiencies with complementation islands in this region. If the 15 previously existing *unc-22* deficiencies are taken into account, a total of 14 out of 37 (38%) of *unc-22* deficiencies are known to have a breakpoint in this interval. The *unc-22* - *let-52* interval therefore appears to be a hot spot for breakpoints of *unc-22* deficiencies induced by either mutagen.

This hot spot is not likely to be merely a coincidence of there being a large physical distance between the two genes since *let-52* was shown to lie on cosmid T12G3, which is the nearest cosmid on the physical map to the right of the cosmid containing *unc-22* (ACEDB, R. Durbin and J. Thierry-Mieg, unpublished; see Figure 10). These two cosmids are joined only by a YAC bridge which, given a mean insert size of 200 - 250 kb (Coulson et al. 1991), could theoretically be upwards of 200 kb in size. However, overlaps of several smaller YACs in the gap suggest it is likely to be less than 100 kb. It may be that the hot spot coincides with a region of tandem or inverted repeats, which are often associated with YAC bridges (Wilson et al. 1994; A. Coulson and S. Jones, personal communication). Tandem repeats have been noted in ACEDB for YACs bridging the *unc-22* - *let-52* cosmid contig gap. Whether or not this hot spot is a result of some physical property of the DNA between *unc-22* and *let-52* will likely only be answered by determination of the sequence between the two genes, and determination of the molecular breakpoints of the deficiencies that break in the interval.

In addition to the physical positioning of *let-52*, the *let-661* and *let-93* genes were also anchored to the physical map. These data indicate that, at one essential gene per every two cosmids, the density of essential genes to the right of *unc-22* is half that to the left of *unc-22*. This would seem to correlate well with the genetic map, which shows a relatively lower density of essential genes to the right of *unc-22* than to the left. It has been observed that each of the *C. elegans* autosomes contains a gene cluster in which the

majority of identified genes are located (Brenner 1974). The genetic data in the *unc-22* region therefore indicate that *unc-22* is at the right edge of the gene cluster on chromosome IV.

It has been suggested that the observed clustering of genes on the autosomes is due to a decrease in recombination frequency within the cluster relative to that outside the cluster (Greenwald et al. 1987). The correlation data obtained in this study suggest that it may also be due to an actual decrease in the physical density of genes outside the clusters, which is in agreement with previous observations that the regions outside the clusters are genuinely gene poor (Coulson et al. 1991). It must be remembered, however, that these observations of gene density are based upon the genetic identification of genes which, as evidenced by sequence analysis in this and other studies in the *unc-22* region (Prasad and Baillie 1989; Marra et al. 1993; Marra 1994), and in the 2.2 Mb of *C. elegans* chromosome III sequence (Wilson et al. 1994), fails to identify the majority of genes. Further correlation of genes to the right of *unc-22*, for which reliable recombination data are available, will be necessary to provide an estimate of the relationship between physical and genetic distances outside the gene cluster.

The results of the characterization of *sDf83* and *sDf88* may provide an explanation as to why many genes identified from genomic sequence have not been identified genetically. The analysis of these two deficiencies has determined that there are likely no genes essential for viability in the interval between *let-56* and *unc-22*. Eight genes have been thus far molecularly identified in this approximately 50 kb interval (Prasad and Baillie 1989; Marra et al. 1993; Schein et al. 1993; Marra 1994), although the complete genomic sequence of the interval has not been obtained. Three of these genes, *hdl-1*, *spm-1* and *gdl-1* are candidates for the *let-56* gene (Marra et al. 1993; Marra 1994), which means that at least five and as many as seven of these genes are not required for viability. This analysis explains why previous screens for mutations in essential genes in the *unc-22* region failed to identify any genes in this interval.

Rather than suggesting that a large proportion of genes in *C. elegans* are not essential, these results indicate that not all genes are essential under laboratory culture conditions. The genetic identification of genes is dependent upon our ability to discern a change in phenotype following mutagenic treatment. In terms of the results with *sDf83* and *sDf88*, the possibility cannot be ruled out that the deleted coding elements provide functions required under non-laboratory conditions, or that deletion of these coding elements causes more subtle phenotypes such as uncoordinated movement that may be masked by the Unc-22 phenotype of these individuals, or behavioral abnormalities that would only be apparent under particular assay conditions. It is also possible that one or more of these elements are members of multicopy gene families and as such may encode redundant or partially redundant functions.

Although only a fraction of the estimated 17 800 genes in the *C. elegans* genome (Wilson et al. 1994) have been identified genetically, the rate of their identification has remained ahead of our ability to genetically map them in order to produce a linear linkage map. In addition, the positioning of these genes on the physical map and subsequent determination of their molecular nature is also lacking, although the development and application of a simple protocol for obtaining and identifying transgenic worms (Mello et al. 1991) has begun to address this problem. This study has been a continuation of efforts to produce a high resolution genome map in the *unc-22* region of chromosome IV - that is, a high resolution genetic map that is correlated at high resolution to the physical map. With respect to the region immediately to the left and right of the *unc-22* gene, in the *let-65* - *let-662* interval, this has nearly been achieved. Of the 14 genes that have been mutationally identified in this interval, all but two (*let-92* and *let-664*) have been ordered with respect to each other, and all but four (*let-65*, *let-660*, *let-662* and *let-664*) have been positioned on the physical map by transgenic rescue. In terms of the identification of additional essential genes in this region, Marra (1994) suggests that the practical upper limit for the density of genes in this interval, which when mutated slow or arrest

development of the worm, has been achieved. Therefore, at least in this small interval, increased resolution of the genome map will likely only be achieved through the cloning of additional genes, and the molecular mapping of deficiency breakpoints.

With regard to the genetic map to the left and right of the *let-65* - *let-662* region, there are several clusters of genes for which the gene order has yet to be determined, including two large gene clusters just to the left of *let-65* (see Figure 1). Due to the fact that it has been difficult to obtain deficiencies with breakpoints that fall within these clusters, the best approach for both ordering the genes in these two clusters and anchoring them to the physical map is likely to be that of transgenic rescue. Three of the genes in this region have already been positioned on the physical map: *mec-3* (Way and Chalfie 1988), *him-8* (S. Broverman, personal communication) and *lin-3* (Hill and Sternberg 1992). The anchoring of these three genes to the physical map provides a starting point for the selection of clones for additional transgenic rescue experiments. Correlating the genes within the clusters to cosmid clones on the physical map will simultaneously order the genes with respect to each other and align them with the physical map. This approach to aligning the physical and genetic maps within a cluster of unordered genes has been used in the *dpy-14* region of chromosome I (McKay 1993), and in the *rol-3* region of chromosome V (W. B. Barbazuk, personal communication).

The creation of a highly refined genome map in a previously poorly characterized region on chromosome III is now being undertaken in collaboration with the *C. elegans* Genome Sequencing Consortium (H. Stewart, D. Collins, A. M. Howell and D. Baillie, personal communication). Undertakings of this sort are essential for providing functional information about the genomic sequence that is being generated on chromosome III. Likewise, a high resolution genome map in the *unc-22* region will prove to be indispensable during the analysis of genomic sequence as it is generated in this region by the *C. elegans* Genome Sequencing Consortium.

Appendix

Complementation test results of JES1 and JES2 deficiencies. A "+" indicates that the deficiency complements the mutation tested. A "-" indicates that the deficiency fails to complement the mutation tested. Deficiencies *sDf66* through to *sDf87* were recovered in the JES1 screen. Deficiencies *sDf90* through to *sDf100* were recovered in the JES2 screen.

Zone	Lethal	*sDf66*	*sDf68*	*sDf80*	*sDf81*	*sDf82*	*sDf83*	*sDf84*	*sDf85*	*sDf86*	*sDf87*
	dpy-4	+			+			+			+
	tra-3										
15	*let-324*				+						
14	*let-323*				-						
13	*let-313*										
	let-318										
	let-319										
	let-320										
	let-321										
	let-322										
	let-315				-						
	let-316				-						
	unc-26										
12	*let-325*	+			-			+			+
11C	*unc-30*	+						+			+
	let-314	+						+			+
	let-317	+						+			+
11B	*unc-31*	-			-			+			+
11A	*let-301*	-						-			-
10	*let-97*				-			-			-
	let-99	-						-			-
9	*let-68*										
	let-309										
	let-656										
	dpy-26										
8	*let-66*										
	let-67										
7B	*let-662*										
7A	*let-93*	-		+	-	+	+	-	+	+	-
6B	*let-661*										
6A	*let-52*	-	+	+	-	+	+	-	+	+	-
5B	*unc-22*	-	-	-	-	-	-	-	-	-	-
5A	*let-56*				-		-				
4B	*let-653*	-	-	-	-	-	+	-	-	-	-
4A2	*let-660*	-								+	
4A1	*let-92*	+			-	-				+	
	let-664										
3B	*par-5*				-	-					
3A	*dpy-20*	+	-	-	+	+	+	-	-	+	-
	let-60							-			
2	*let-65*			-	+			+	-		-
1D2	*let-59*		-						+		+
	let-73		-					+	+		+
	let-91		-						+		+
	let-98		-						+		+
	let-100		-						+		+
	let-311		-						+		+
	let-654		-	-					+		+
	let-655		-	-					+		+
1D1	*let-70*		-						+		+
1C2	*let-64*		+	-				+	+		+
	lin-3		+	-							
1C1	*let-71*		+	+				+	+		
1B	*let-312*			+							

Zone	Lethal	*sDf90*	*sDf91*	*sDf92*	*sDf93*	*sDf94*	*sDf95*	*sDf96*	*sDf97*	*sDf98*	*sDf99*	*sDf100*
	dpy-4	+		+				+			+	
	tra-3											
15	*let-324*											
14	*let-323*											
13	*let-313*											
	let-318											
	let-319											
	let-320											
	let-321											
	let-322											
	let-315											
	let-316											
	unc-26											
12	*let-325*	+	+					+		+		+
11C	*unc-30*		+					+				+
	let-314							+				+
	let-317	+	+		+			+				+
11B	*unc-31*	+		+				-			+	-
11A	*let-301*	+	+	+		+		-			+	-
10	*let-97*	-	+	+		+		-		+	+	-
	let-99	-	+	+		+		-		+	+	-
9	*let-68*		-	-		-	+	-		-	-	-
	let-309	-	-	-		-		-		-	-	-
	let-656	-	-	-		-		-		-	-	-
	dpy-26	-	-	-		-				-	-	
8	*let-66*	-	-	-	+	-	+			-	-	-
	let-67	-	-	-	+	-	+			-	-	-
7B	*let-662*											
7A	*let-93*	-	-	-	+	-	+	-	+	-	-	-
6B	*let-661*	-	+	+	+	-	+		+	-	+	-
6A	*let-52*	-	+	+	+	-	+	-	+	+	+	-
5B	*unc-22*	-	-	-	-	-	-	-	-	-	-	-
5A	*let-56*	-	-	-	-	-	-		-	-	-	-
4B	*let-653*	-	-		-	-				-		
4A2	*let-660*											
4A1	*let-92*	-	-		-	-		-	-			
	let-664	-	-		-	-			-			
3B	*par-5*	-	+/-		-	-			-			-
3A	*dpy-20*	+	+	-	+	+	-	-	+	-	-	+
	let-60		+	-	+	+	-		+	-	-	+
2	*let-65*			-			+	-		-	-	+
1D2	*let-59*			-				+		+	+	
	let-73			-				+		+	+	
	let-91			-				+		+	+	
	let-98			-				+	+	+	+	
	let-100			-				+		+	+	
	let-311			-				+	+	+	+	
	let-654			-				+	+	+	+	
	let-655			-				+		+	+	
1D1	*let-70*			+				+		+	+	
1C2	*let-64*											
	lin-3											
1C1	*let-71*											
1B	*let-312*			+				+		+		

References

Ajioka, J. W., D. A. Smoller, R. W. Jones, J. P. Carulli, A. E. C. Vellek, D. Garza, A. J. Link, I. W. Duncan and D. L. Hartl, 1991 *Drosophila* genome project: one-hit coverage in yeast artificial chromosomes. Chromosoma **100:** 495-509.

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

Auerbach, C., M. Moutschen-Dahmen and J. Moutschen, 1977 Genetic and cytogenetical effects of formaldehyde and related compounds. Mutat. Res. **39:** 317-362.

Barstead, R. J., and R. H. Waterston, 1991 Vinculin is essential for muscle function in the nematode. J. Cell Biol. **114:** 715-724.

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

Benian, G. M., S. W. L'Hernault and M. E. Morris, 1993 Additional sequence complexity in the muscle gene, *unc-22*, and it's encoded protein, twitchin, of *Caenorhabditis elegans*. Genetics **134:** 1097-1104.

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

Brash, D. E., and W. A. Haseltine, 1982 UV-induced mutation hotspots occur at DNA damage hotspots. Nature **298:** 189-192.

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

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

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

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

Cooley, L., R. Kelley and A. Spradling, 1988 Insertional mutagenesis of the *Drosophila* genome with single P elements. Science **239:** 1121-1128.

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

Coulson, A., R. Waterston, J. Kiff, J. Sulston and Y. Kohara, 1988 Genome linking with yeast artificial chromosomes. Nature **335:** 184-186.

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

Cross, S. H., and P. F. R. Little, 1986 A cosmid vector for systematic chromosome walking. Gene **49:** 9-22.

Donati, L. A. M., 1985 A genetic analysis of the right arm of linkage group IV of *Caenorhabditis elegans*, with emphasis on the *sDf2* region. M.Sc. Thesis, Simon Fraser University, Burnaby, B. C., Canada.

Dujon, B., D. Alexandraki, B. André, W. Ansorge, V. Baladron, J. P. G. Ballesta, A. Banrevi, P. A. Bolle, M. Bolotin-Fukuhara, P. Bossier, G. Bou, J. Boyer, M. J. Buitrago, G. Chéret, L. Colleaux, B. Daignan-Fornier, F. del Rey, C. Dion, H. Domdey, A. Düsterhöft, S. Düsterhus, K-D. Entian, H. Erfle, P. F. Esteban, H. Feldmann, L. Fernandes, G. M. Fobo, C. Fritz, H. Fukuhara, C. Gabel, L. Gaillon, J. M. Garcia-Cantalejo, J. J. Garcia-Ramirez, M. E. Gent, M. Ghazvini, A. Goffeau, A. Gonzaléz, D. Grothues, P. Guerreiro, J. Hegemann, N. Hewitt, F. Hilger, C. P. Hollenberg, O. Horaitis, K. J. Indge, A. Jacquier, C. M. James, J. C. Jauniaux, A. Jimenez, H. Keuchel, L. Kirchrath, K. Kleine, P. Kötter, P. Legrain, S. Liebl, E. J. Louis, A. Maia e Silva, C. Marck, A-L. Monnier, D. Möstl, S. Müller, B. Obermaier, S. G. Oliver, C. Pallier, S. Pascolo, F. Pfeiffer, P. Philippsen, R. J. Planta, F. M. Pohl, R. Pöhlmann, D. Portetelle, B. Purnelle, V. Puzos, M. Ramezani Rad, S. W. Rasumussen, M. Remacha, J. L. Revuelta, G-F. Richard, M. Rieger, C. Rodrigues-Pousada, M. Rose, T. Rupp, M. A. Santos, C. Schwager, C. Sensen, J. Skala, H. Soares, F. Sor, J. Stegemann, H. Tettelin, A. Thierry, M. Tzermia, L. A. Urrestarazu, L. van Dyck, J. C. van Vliet-Reedijk, M. Valens, M. Vandenbol, C. Velela, S. Vissers, D. von Wettstein, H. Voss, S. Wiemann, G. Xu, J. Zimmermann, M. Haasemann, I. Becker and H. W. Mewes, 1994 Complete DNA sequence of yeast chromosome XI. Nature **369:** 371- 378.

Feldman, M. Ya., 1973 Reactions of nucleic acids and nucleoproteins with formaldehyde. Prog. Nucleic Acid Res. Mol. Biol. **13:** 1-49.

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

Fink, G. R., 1988 Notes of a bigamous biologist. Genetics **118:** 549-550.

Fire, A., 1986 Integrative transformation of *Caenorhabditis elegans*. EMBO J. **5:** 2673-2680.

Fremeau, R. T. Jr., M. G. Caron and R. D. Blakely, 1992 Molecular cloning and expression of a high affinity L-proline transporter expressed in putative glutamatergic pathways of rat brain. Neuron **8:** 915-926.

Gibson, T. J., A. R. Coulson, J. E. Sulston and P. F. R. Little, 1987a Lorist2, a cosmid with transcriptional terminators insulating vector genes from interference by promoters within the insert: effect on DNA yield and cloned insert frequency. Gene **53:** 275-281.

Gibson, T. J., A. Rosenthal and R. H. Waterston, 1987b Lorist6, a cosmid vector with *Bam*HI, *Not*I, *Sca*I and *Hin*dIII cloning sites and altered neomycin phosphotransferase gene expression. Gene **53:** 283-286.

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

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

Hartman, P. S., J. Hevelone, V. Dwarakanath and D. L. Mitchell, 1989 Excision repair of UV radiation-induced DNA damage in *Caenorhabditis elegans*. Genetics **122:** 379-385.

Hartman, P., J. Reddy and B-A. Svendsen, 1991 Does trans-lesion synthesis explain the UV-radiation resistance of DNA synthesis in *C. elegans* embryos? Mutat. Res. **255:** 163-173.

Hill, R. J., and P. W. Sternberg, 1992 The gene *lin-3* encodes an inductive signal for vulval development in *C. elegans*. Nature **358:** 470-476.

Hodgkin, J., 1983 X chromosome dosage and gene expression in *Caenorhabditis elegans*: two unusual dumpy genes. Mol. Gen. Genet. **192:** 452-458.

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

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

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

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

Johnsen R. C., and D. L. Baillie, 1988 Formaldehyde mutagenesis of the *eT1* balanced region in *Caenorhabditis elegans*: dose response curve and the analysis of mutational events. Mutat. Res. **201:** 137-147.

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

Jones, S. J. M., 1994 The molecular characterization of the essential gene *let-653* in *Caenorhabditis elegans*. M.Sc. Thesis, Simon Fraser University, Burnaby, B. C., Canada.

Keller, C. I., J. Calkins, P. S. Hartman and C. S. Rupert, 1987 UV photobiology of the nematode *Caenorhabditis elegans*: action spectra, absence of photoreactivation and effects of caffeine. Photochem. Photobiol. **46:** 483-488.

Kimble, J. and D. Hirsh, 1979 The postembryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*. Dev. Biol. **70:** 396-417.

Kohara, Y. , K. Akiyama and K. Isono, 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., R. P. French, E-C. Park and J. J. Johnson, 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.

Kubitschek, H. E., 1980 Double mutations induced in *Escherichia coli* by ultraviolet light. J. Bacteriol. **142:** 724-725.

Kubitschek, H. E., and G. Venema, 1976 Multiple mutations induced in *Bacillus subtilis* by ultraviolet light. Mutat. Res. **35:** 325-328.

Lewis, J. A., C-H. Wu, H. Berg and J. H. Levine, 1980 The genetics of levamisole resistance in the nematode *Caenorhabditis elegans*. Genetics **95:** 905-928.

L'Hernault, S. W., G. M. Benian and R. B. Emmons, 1993 Genetic and molecular characterization of the *Caenorhabditis elegans* spermatogenesis-defective gene *spe-17*. Genetics **134:** 769-780.

Magaña-Schwenke, N., B. Ekert and E. Moustacchi, 1978 Biochemical analysis of damage induced in yeast by formaldehyde, I. Induction of single-strand breaks in DNA and their repair. Mutat. Res. **50:** 181-193.

Magaña-Schwencke, N., and B. Ekert, 1978 Biochemical analysis of damage induced in yeast by formaldehyde, II. Induction of cross-links between DNA and protein. Mutat. Res. **51:** 11-19.

Marra, M. A., 1994 Genome analysis in *C. elegans*: genetic and molecular identification of genes tightly linked to *unc-22(IV)*. Ph.D. Thesis, Simon Fraser University, Burnaby, B. C., Canada.

Marra, M. A., S. S. Prasad and D. L. Baillie, 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. J., 1993 High resolution alignment of the physical and genetic maps of the *dpy-14* region of chromosome I in *Caenorhabditis elegans*. M.Sc. Thesis, University of British Columbia, Vancouver, B. C. , Canada.

McKim, K. S., M. F. P. Heschl, R. E. Rosenbluth and D. L. Baillie, 1988 Genetic organization of the *unc-60* region in *Caenorhabditis elegans*. Genetics **118:** 49-59.

McKim, K. S., T. Starr and A. M. Rose, 1992 Genetic and molecular analysis of the *dpy-14* region in *Caenorhabditis elegans*. Mol. Gen. Genet. **233:** 241-251.

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

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

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

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

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

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

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

Nishioka, H., 1973 Lethal and mutagenic action of formaldehyde in Hcr^+ and Hcr^- strains of *Escherichia coli*. Mutat. Res. **17:** 261-265.

Oliver, S. G., Q. J. M van der Aart, M. L. Agostoni-Carbone, M. Aigle, L. Alberghina, D. Alexandraki, G. Antoine, R. Anwar, J. P. G. Ballesta, P. Benit, G. Berben, E. Bergantino, N. Biteau, P. A. Bolle, M. Bolotin-Fukuhara, A. Brown, A. J. P. Brown, J. M. Buhler, C. Carcano, G. Carignani, H. Cederberg, R. Chanet, R. Contreras, M. Crouzet, B. Daignan-Fornier, E. Defoor, M. Delgado, J. Demolder, C. Doira, E. Dubois, B. Dujon, A. Dusterhoft, D. Erdmann, M. Esteban, F. Fabre, C. Fairhead, G. Faye, H. Feldmann, W. Fiers, M. C. Francingues-Gaillard, L. Franco, L. Frontali, H. Fukuhara, L. J. Fuller, P. Galland, M. E. Gent, D. Gigot, V. Gilliquet, N. Glansdorff, A. Goffeau, M. Grenson, P. Grisanti, L. A. Grivell, M. de Haan, M. Haasemann, D. Hatat, J. Hoenicka, J. Hegemann, C. J. Herbert, F. Hilger, S. Hohmann, C. P. Hollenberg, K. Huse, F. Iborra, K. J. Indge, K. Isono, C. Jacq, M. Jacquet, C. M. James, J. C. Jauniaux, Y. Jia, A. Jimenez, A. Kelley, U. Kleinhans, P. Kreisl, G. Lanfranchi, C. Lewis, C. G. van der Linden, G. Lucchini, K. Lutzenkirchen, M. J. Maat, L. Mallet, G. Mannhaupt, E. Martegani, A. Mathieu, C. T. C. Maurer, D. McConnell, R. A. McKee, F. Messenguy, H. W. Mewes, F. Molemans, M. A. Montague, M. Muzi Falconi, L. Navas, C. S. Newlon, D. Noone, C. Pallier, L. Panzeri, B. M. Pearson, J. Perea, P. Philippsen, A. Pierard, R. J. Planta, P. Plevani, B. Poetsch, F. Pohl, B. Purnelle, M. Ramezani Rad, S. W. Rasmussen, A. Raynal, M. Remacha, P. Richterich, A. B. Roberts, F. Rodriguez, E. Sanz, I. Schaaff-Gerstenschlager, B. Scherens, B. Schweitzer, Y. Shu, J. Skala, P. P. Slonimski, F. Sor, C. Soustelle, R. Spiegelberg, L. I. Stateva, H. Y. Steensma, S. Steiner, A. Thierry, G. Thireos, M. Tzermia, L. A. Urrestarazu, G. Valle, I. Vetter, J. C. van Vliet-Reedijk, M. Voet, G. Volckaert, P. Vreken, H. Wang, J. R. Warmington, D. von Wettstein, B. L. Wicksteed, C. Wilson, H. Wurst, G. Xu, A. Yoshikawa, F. K. Zimmermann and J. G. Sgouros, 1992 The complete DNA sequence of yeast chromosome III. Nature **357:** 38-48.

Olson, M. V., J. E. Dutchik, M. Y. Graham, G. M. Brodeur, C. Helms, M. Frank, M. MacCollin, R. Scheinman and T. Frank, 1986 Random-clone strategy for genomic restriction mapping in yeast. Proc. Natl. Acad. Sci. (USA) **83:** 7826-7830.

Olson, M. V., L. Hood, C. Cantor and D. Botstein, 1989 A common language for physical mapping of the human genome. Science **245:** 1434-1440.

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

Poverenny, A. M., Yu. A. Siomin, A. S. Saenko and B. I. Sinzinis, 1975 Possible mechanisms of lethal and mutagenic action of formaldehyde. Mutat. Res. **27:** 123-126.

Prasad, S. S., 1988 Molecular analysis within the *dpy-20* to *unc-22* region of chromosome IV in *Caenorhabditis elegans*. Ph.D. Thesis, Simon Fraser University, Burnaby, B. C., Canada.

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

Riles, L., J. E. Dutchik, A. Baktha, B. K. McCauley, E. C. Thayer, M. P. Leckie, V. Braden, J. E. Depke and M. V. Olson, 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 D. L. Baillie, 1985 Genetic organization of the *unc-22 IV* gene and the adjacent region in *Caenorhabditis elegans*. Mol. Gen. Genet. **201:** 409-414.

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

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

Rogalski, T. M., A. M. E. Bullerjahn and D. L. Riddle, 1988 Lethal and amanitin-resistance mutations in the *Caenorhabditis elegans ama-1* and *ama-2* genes. Genetics **120:** 409-422.

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

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

Rothstein, R. J., 1983 One-step gene disruption in yeast. Methods Enzymol. **101:** 202-211.

Sage, E., 1993 Distribution and repair of photolesions in DNA: genetic consequences and the role of sequence context. Photochem. Photobiol. **57:** 163-174.

Sage, E., E. Cramb and B. W. Glickman, 1992 The distribution of UV damage in the *lacI* gene of *Escherichia coli*: correlation with mutation spectrum. Mutat. Res. **269:** 285-299.

Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich and N. Arnheim, 1985 Enzymatic amplification of ß-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science **230:** 1350-1354.

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis and H. A. Erlich, 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:** 487-491.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular cloning: a laboratory manual*. 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Schein, J. E., M. A. Marra, G. M. Benian, C. Fields and D. L. Baillie, 1993 The use of deficiencies to determine essential gene content in the *let-56 - unc-22* region of *Caenorhabditis elegans*. Genome **36:** 1148-1156.

Shakes, D. C., and S. Ward, 1989 Mutations that disrupt the morphogenesis and localization of a sperm-specific organelle in *Caenorhabditis elegans*. Dev. Biol. **134:** 307-316.

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

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

Stinchcomb, D. T., J. E. Shaw, S. H. Carr and D. Hirsh, 1985 Extrachromosomal DNA transformation of *Caenorhabditis elegans*. Mol. Cell. Biol. **5:** 3484-3496.

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

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

Sulston, J. E., D. G. Albertson and J. N. Thomson, 1980 The *Caenorhabditis elegans* male: postembryonic development of nongonadal structures. Dev. Biol. **78:** 542-576.

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

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

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

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

Way, J. C, and M. Chalfie, 1988 *mec-3*, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in *C. elegans*. Cell **54:** 5-16.

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

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

Wilson, R., R. Ainscough, K. Anderson, C. Baynes, M. Berks, J. Bonfield, J. Burton, M. Connell, T. Copsey, J. Cooper, A. Coulson, M Craxton, S. Dear, Z. Du, R. Durbin, A. Favello, A. Fraser, L. Fulton, A. Gardner, P. Green, T. Hawkins, L. Hillier, M. Jier, L. Johnston, M. Jones, J. Kershaw, J. Kirsten, N. Laisster, P. Latreille, J. Lightning, C. Lloyd, B. Mortimore, M. O'Callaghan, J. Parsons, C. Percy, L. Rifken, A. Roopra, D. Saunders, R. Shownkeen, M. Sims, N. Smaldon, A. Smith, M. Smith, E. Sonnhammer, R. Staden, J. Sulston, J. Thierry-Mieg, K. Thomas, M. Vaudin, K. Vaughan, R. Waterston, A. Watson, L. Weinstock, J. Wilkinson-Sproat and P. Wohldman, 1994 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. Nature **368:** 32-38.

Yura, T., H. Mori, H. Nagai, T. Nagata, A. Ishihama, N. Fujita, K. Isono, K. Mizobuchi and A. Nakata, 1992 Systematic sequencing of the *Escherichia coli* genome: analysis of the 0-2.4 min region. Nucleic Acids Res. **20:** 3305-3308.

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