Winter 1999

The mut-2 mutator of Caenorhabditis elegans

Kathrine Queta Flint Boese
University of New Hampshire, Durham

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

UMI®
800-521-0600

THE *mut-2* MUTATOR OF *CAENORHABDITIS ELEGANS*

BY

Kathrine Queta Flint Boese

B.A., Purdue University, 1982

B.S., Purdue University, 1984

M.S., University of Virginia, 1991

DISSERTATION

Submitted to the University of New Hampshire

in Partial Fulfillment of

the Requirements for the Degree of

Doctor of Philosophy

in Biochemistry

December 1999

This dissertation has been examinded and approved.

Dissertation Director, John J. Collins, Ph.D.
Associate Professor, Biochemistry and Molecular Biology

Anita Klein, Ph.D.
Associate Professor, Biochemistry and Molecular Biology

Tom Laue, Ph.D.
Professor, Biochemistry and Molecular Biology

Joe Pelliccia, Ph.D.
Professor, Biology, Bates College

Tom Davis, Ph.D.
Associate Professor, Plant Biology

6/25/99
Date

DEDICATION

I would like to dedicate this to my best friend, Jamie Boese.

ACKNOWLEDGMENTS

It is impossible to adequately acknowledge all the people and their support during my time in the Department of Biochemistry and Molecular Biology. For those who go unmentioned on this page, I will say that their friendship, confidence and support were invaluable to me and will always be remembered.

I would like to thank my committee members including John Collins, Tom Davis, Anita Klein, Tom Laue and Joe Pelliccia, for their guidance and patience. And though they might grimace when they hear me say this, I enjoyed teaching Biochemistry 755, first with Andy Laudano and then with Tom Laue. Of course, the success of the teaching labs owes even more to the tireless efforts of Gail Stewart . Anita Klein deserves special thanks for being a mentor, a colleague and a true friend. I also would like to acknowledge Rick Cote, Clyde Denis, and Stacia Sower who always made themselves available to me when I hit the highs and the lows so typical of graduate school and of life in general. Working together with this group of people was a highlight of my time at UNH.

I would like to thank John Collins for introducing me to *C. elegans* and for providing an environment where I had the pleasure of working with many wonderful people. I value most my friendships that were established both in and out of the lab, especially with Deb Audino, Jennifer Hogan, Craig Almeida, Mindy Zhang, Jeremy

Glasner, Rosa Alcazar and Carmen Maftei. Joe Pelliccia deserves special mention for his contribution to this project during his sabbatical and also for his continued support and friendship.

Finally, I want to acknowledge my family for their love and support without which this work would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

THE *mut-2* MUTATOR

OF

CAENORHABDITIS ELEGANS

by

Kathrine Queta Flint Boese

University of New Hampshire, December 1999

The *mut-2* mutator plays multiple regulatory roles in the germ line of *C. elegans*. In addition to regulating germ line transposition of at least four distinct transposon families (Tc1, Tc3, Tc4 and Tc5) *mut-2* is implicated in chromosome segregation. Animals that harbor the *mut-2* mutator produce broods with a higher incidence of males phenotype (Him) as a result of an increase in X chromosome non-disjunction during meiosis.

Using the Him phenotype conferred by *mut-2*(r459), I mapped the gene to the *dpy-14 sem-4* interval on LGI. However, efforts to identify a molecular clone of the gene were hampered because the available phenotypes were unsuitable for standard transformation rescue approaches. This lead me to re-examine the temperature sensitive behavior of *mut-2*(r459); the original *mut-2* isolate, TR674, is sterile at 25°C. Previously it was assumed that the ts behavior was related to its heritage. TR674 is derived from a

hybrid of two common laboratory strains, N2 and Bergerac (BO). N2 is fertile at 25°C but BO is sterile. Analysis of the immediate ancestors and recombinants derived from *mut-2*-bearing strains demonstrated that the ts sterile phenotype cosegregates with *mut-2* and is unrelated to its BO ancestry. Closer inspection revealed that sterility results from a defect in gametogenesis. The gonads of animals raised at 25°C have few, morphologically abnormal sperm. The oocytes appear to be normal but attempts to cross-fertilize with wild type sperm at 25°C failed to produce any viable progeny.

This thesis describes how I took advantage of the ts phenotype to achieve rescue and identify a candidate gene. I demonstrated rescue and restored fertility with a small, genetically defined free duplication, hDp65. Using standard DNA transformation techniques, I defined the relative physical position of the duplication endpoint within the *dpy-14 sem-4* interval thereby reducing the *mut-2* interval to a region spanned by two cosmids, C30F12 and H06O01. Of these two, H06O01 rescued the ts sterile phenotype. The sequence represented in this cosmid includes a gene with similarity to the chromo domain-helicase-DNA binding protein, C HD1. This gene is a member of the SWI/SNF superfamily of transcriptional regulators and encodes a protein with several motifs shared by factors that regulate transcription by remodeling chromatin architecture.

CHD1 represents a compelling candidate for the *mut-2* gene given that *mut-2* plays a role in regulating multiple transposons genome wide and participates in chromosome segregation and gametogenesis. Confirmation awaits rescue by the CHD1 gene and sequencing of the molecular lesion responsible for *mut-2*(r459).

The work that I have described moves us closer to identifying and cloning the *mut-2* gene. Ultimately, the characterization of *mut-2*'s structure and function will elucidate the mechanisms regulating transposition in *C. elegans*, allow better use of transposons as molecular tools and provide insight into the processes that maintain integrity and plasticity of the genome throughout life.

INTRODUCTION

The Biology of Transposable Elements

An important tenet for all organisms is maintenance of genome integrity. This is crucial to preserve an organism's genetic blueprint and to ensure adaptation and survival for future generations. Far from being a static set of instructions, genomes are dynamic and tolerant of alterations and rearrangements. A fundamental type of genetic rearrangement in prokaryotes and eukaryotes alike occurs as a result of transposable element activity. DNA transposable elements are discrete sequences present as dispersed repetitive elements with the unique ability to move independently to new chromosomal locations. While their mobility, prevalence and persistence are interesting phenomena, a far more provocative issue is how host genomes tolerate and regulate the behavior of these unusual guests. This question is addressed here in part by the characterization of the *mut-2* mutator of *Caenorhabditis elegans (C. elegans)*, a participant in mechanisms that control transposable element activity and a likely component of mechanisms that govern host genome integrity and plasticity.

Discovery:

Eukaryotic DNA transposable elements (transposons) were first identified by Barbara McClintock in her pioneering work on the chromatin behavior of maize in the early 1940's. The consequences of transposon activity had been described previously as a heritable instability, or pattern of variegation, in maize (Emerson, 1929; Emerson, 1917;

Emerson, 1914). However, it was McClintock's careful observations that linked the phenotypic patterns of variegation in corn kernels, ears and leaves to physical changes occurring in the chromatin (McClintock, 1951; McClintock, 1947; McClintock, 1946). The mutational occurrences, resulting in the development of pigmented or non-pigmented sectors in kernels or leaves, were far greater in one sector as compared to a twin sector derived from sibling cells. This pattern suggested that a controlling element had been altered or differentially segregated during mitosis of the progenitor cell. Large pigment patches indicated early activity affecting subsequent cellular descendants while small spots resulted from later activity in development affecting fewer cells. From her analyses, McClintock also realized that the "controlling elements" were not fixed but capable of transposition to new genomic sites (McClintock, 1948); an idea that was not well received in the days when chromosomal material was considered a static, immutable catalogue of genetic information. For McClintock, the concept of mobile genetic units was interesting, but what intrigued her the most is revealed in her comment, "The real secret to all of this is control, it is not transposition" (N. Comfort, 1997 Ph.D. Thesis, SUNY Stony Brook, USA; (McClintock, 1987).

Classification and Structure

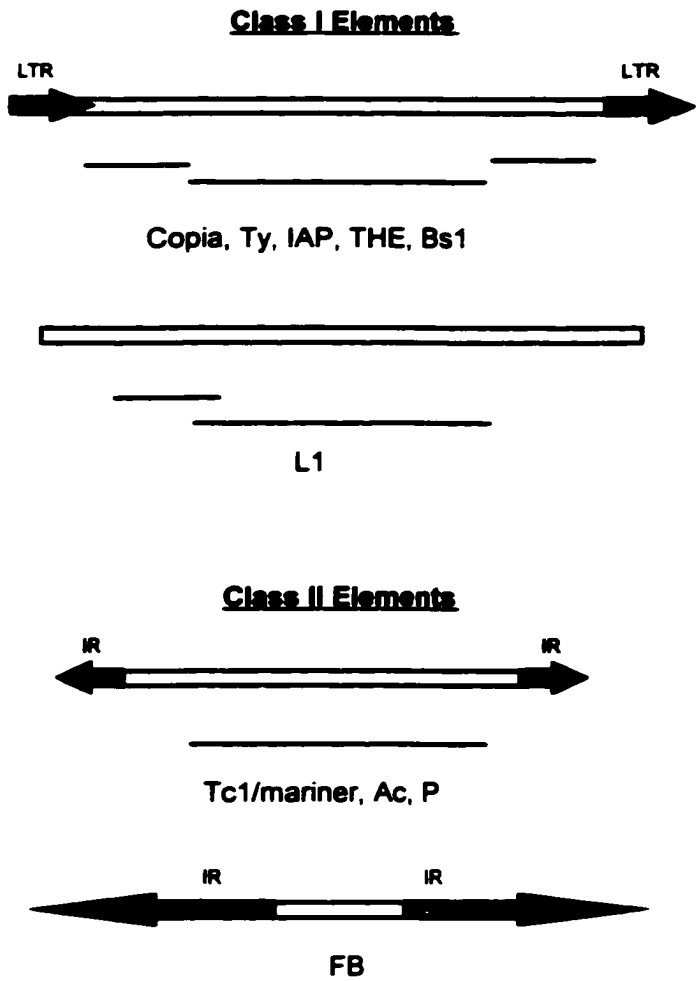
After the identification of transposons in maize, subsequent molecular studies revealed that these elements were ubiquitous to the genomes of most, if not all organisms. In prokaryotes, the simplest form of transposons are represented by insertion sequences (IS). These elements vary considerably in size, sequence and distribution. Prokaryotic transposons are often composites of IS (Tn). IS's and Tn's encode replication and

transposition-dependent proteins and often genes that confer some selective advantage such as antibiotic resistance (i.e. Tn3: amp^R; Tn5: kan^R, $bleo^R$, $strep^R$; Tn10: tet^R; etc.). IS elements represent a fraction of bacterial genomes and are residents on plasmids and bacteriophage as well (reviewed in Berg and Howe, 1989).

Eukaryotic transposons are present as dispersed repetitive sequences that comprise a significant fraction of whole genomes. In *Drosophila* they can represent as much as ~10% of the genome (Pimpinelli et al., 1995), ~50% in maize (SanMiguel et al., 1996), and ~35% in humans (Dimitri and Junakovic, 1999; Labrador and Corces, 1997). Based on early studies of the re-association kinetics of *C. elegans* DNA, 17% of the genome is composed of repetitive sequence (Sulston and Brenner, 1974). More recent analysis of the completely sequenced genome reveals that greater than 5 % of the repetitive sequence is represented by specific repeat sequences (i.e. repA), tandem repeats and inverted repeats (Waterston, 1997).

Transposons are broadly classified according to their general structure and mode of transposition (Figure 1; reviewed in (Finnegan, 1989); (Berg and Howe, 1989)). Class I elements are commonly referred to as retrotransposons because they resemble retroviruses in their structure and mode of transposition (Boeke et al., 1985; Garfinkel et al., 1985, Berg, 1989 #53). They encode element-specific proteins including a reverse transcriptase (RT) important for transposition. The RT of Class I elements facilitates transposition via an RNA intermediate. These elements are often, but not always, flanked by long terminal repeats (LTRs). Examples of LTR-bound Class I elements include the BS1 elements in maize, copia-like elements of *Drosophila*, Ty elements of

Figure 1. The General Structure of DNA transposons.



Class I elements are discrete sequences with or without long terminal repeats (LTR) and encode element specific proteins (gag, reverse transcriptase, env and others). Class II elements are bound by inverted repeats (IR). Often, but not always, the internal sequences encode a transposase. Fold-back elements are a special type of Class II elements composed almost entirely of IRs. (Adapted from Finnegan, 1989).

Saccharomyces cerevisiae, iAP elements in mice and THE elements in humans. Non-LTR Class I elements include Cin4 of maize, the jockey elements of *Drosophila*, and LINE (long interspersed elements i.e. L1) and SINE (short interspersed elements i.e. *Alu*) in humans. In *C. elegans*, putative retrotransposon-like elements have only recently been described (Marin et al., 1998; Youngman et al., 1996) and with the availability of the complete worm genome sequence, more are likely to be defined.

Class II elements are sequences of variable size flanked by inverted repeats (IR). Still other elements are composed entirely of IRs with little internal sequence. This latter class can form hairpin structures and are referred to as "fold-back" elements. Class II transposable elements are grouped into "families" according to their ability to interact with each other genetically (Fedoroff, 1989). For many Class II elements, the internal sequences contain a gene that encodes a protein involved in element mobility, termed transposase. The transposase interacts with the transposon to promote excision and re-insertion into new sites. Often the activity of an entire family of transposons (defined as the many copies of that element found in the genome) depends on the expression of its transposase. These elements transpose via a DNA intermediate by means of a "cut-and-paste" mechanism. The best studied members of the Class II group include the *Ac/Ds* (*Activator/Dissociation*) transposon pair of maize, P elements of *Drosophila* and Tc1 elements of *C.* elegans.

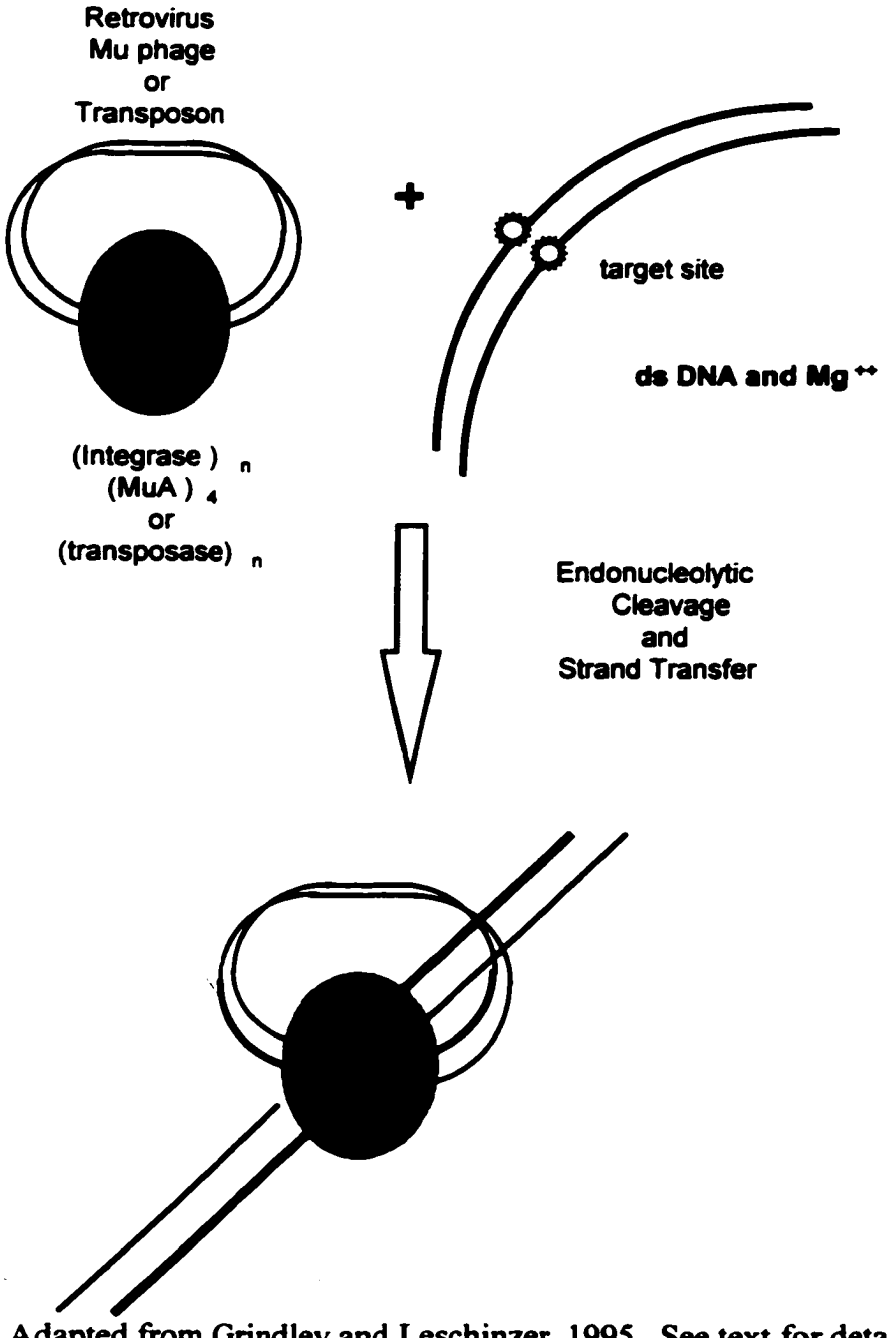
General Mechanism of Transposition:

The unifying hallmark of all transposons is their mobility within the host genome. Regardless of their structure or mode of transposition, transposons are capable of excising

from chromosomal sites they occupy as well as insertion into new sites. Two general strategies for transposition can be described: replicative and non-replicative. Replicative strategies are common to retrotransposons and preserve the original elements while propagating copies which can insert at new sites. This mechanism employs several element encoded proteins that often include reverse transcriptase or integrase, protease, RNAse H and/or gag and env proteins.

Non-replicative strategies are common to many transposons and involve a general "cut-and-paste" mechanism of transposition that requires only the transposase (Figure 2). The mechanism involves three steps: (1) cleavage of the donor DNA, (2) strand transfer and (3) processing at the donor and recipient sites (Mizuuchi, 1992). The element encoded transposase cleaves the DNA at the ends of the transposon and forms a DNA-transposase complex that stabilizes the resulting 3'-OH termini. The complex associates with the target site and strand transfer is initiated in a transesterification reaction as the free 3'-OH groups attack the phosphodiester bond at the genomic site targeted for insertion. Host repair mechanisms are thought to be recruited to process the donor site gap generating a characteristic footprint. The footprint is often a duplication of the original target site. Significant genetic alterations can also be introduced at this point depending on the size of the gap generated (up to regions of a kb or more), the fidelity of the gap-filling process or the template used for repair (ectopic plasmid templates or non-homologous chromosomes) (Gloor et al., 1991).

Figure 2. Cut-and-Paste Model for transposition.



Adapted from Grindley and Leschinzer, 1995. See text for details.

The transposase of eukaryotic transposons (and integrases of prokaryotic elements) are fairly divergent proteins however they all share a well conserved catalytic domain that preserves the relative positions of two aspartates and a glutamate residue or three aspartates arranged in a "DD(X_{34-35})E" or "DD(X_{34-35})D" motif (Doak et al., 1994). The catalytic domain that includes this motif is important for the DNA cleavage required for transposition (Baker and Luo, 1994; Kim et al., 1995; Kulkosky et al., 1992). Crystallographic studies of three structurally distinct transposases from the retro-elements Mu, HIV and ASV reveal that for each, these conserved residues are brought together in the catalytic domain that binds a divalent cation (Mg^{2+} or Mn^{2+}). Binding of the cation is necessary for element excision (Grindley and Leschziner, 1995).

To identify the important residues of the catalytic domain, the *Drosophila mariner* DD(34)D motif was replaced with the DD(34)E motif of its close relative, Tc1 of *C. elegans*. This D to E alteration rendered the transposase non-functional as judged by loss activity of a target transposon in the germ line of the fly. Mutations of other residues within the conserved domain also inhibited transposition even in the presence of wild type transposase (Lohe et al., 1997). One interpretation for inactivation even in the presence of wild type transposase is that the excision reaction involves multiple transposase subunits such that the mutant transposase "poisons" the transposition reaction.

In *C. elegans*, mutation of any of the DDE residues of Tc1A or Tc3A inactivate transposition (van Luenen et al., 1994; Vos and Plasterk, 1994). Together, these results suggest that despite similarities in the transposition mechanism, the process retains

features unique to the system and each family of elements. An interesting question, then, is what other element- or host-dependent factors might participate to facilitate or regulate this process?

Regulation

Both replicative and non-replicative modes of transposition described above are efficient means for element propagation. The efficiency of these strategies poses an interesting dilemma to the host. Transposon activity can easily increase potentially deleterious activities including insertions, excisions and the associated consequences (footprints, frameshift mutations, large deletions, etc). However, activity might also be a source of benign or advantageous alterations by introducing genes of resistance, expanding genetic diversity through gene rearrangements or gene duplications, etc.

In nature, an apparent balance has evolved between host tolerance and transposon parasitism. Constraints are likely to be imposed by both the host and its guests at several levels. These include pre- and post-transcriptional regulation of element encoded proteins, maintenance of protein threshold levels and tissue-specific control of expression or transposition (germ line vs. soma) (Labrador and Corces, 1997). In turn, the transposon benefits from host encoded mechanisms of DNA synthesis, repair and general maintenance.

It has been proposed that transposons and their hosts have co-evolved mechanisms to reduce the deleterious impact and in some instances contribute positively to host fitness and evolution (reviewed in (Kidwell and Lisch, 1997). Mechanisms that might mitigate impact include a bias for insertion into non-coding sequences (Spradling

et al., 1995), pre-mRNA splicing of transposons (Fridell et al., 1990; Rushforth and Anderson, 1996), tissue-specific transposition (Tseng et al., 1991), regulation of copy number by element encoded repressors (Misra and Rio, 1990; Robertson and Engels, 1989) or host-mediated methylation and insertion-dependent regulation of mutant phenotypes (Fedoroff et al., 1995).

In addition to functioning as well known sources of variation, transposons can benefit the host by inserting regulatory elements into a new context (Britten, 1997), by providing new sources of introns (Bradley et al., 1993; Greene et al., 1994), by functioning in telomere replacement (*HET-A* and *TART*) (Pardue et al., 1996), and by providing mechanisms for DSB repair (Moore and Haber, 1996; Teng et al., 1996).

Ultimately, the collective activity of transposons is thought to play a significant role in genome and organism evolution (Fanti et al., 1998). As summarized above, the genetic consequences of these events include spontaneous mutations, alterations in gene expression and reorganization of the genome. Thus transposons and the mechanisms regulating their activity present investigators with interesting paradigms for the study of the "normal" activities that maintain and regulate the dynamic eukaryotic genome.

Transposons as Tools

In addition to generating curiosity because of their novel mobility, transposons were quickly recognized and recruited as molecular tools for a variety of sophisticated genetic manipulations in a broad range of model systems. They have been successfully used to clone genes near which they insert by "transposon tagging" (Bingham et al., 1981) or by "inverse PCR" (Ochman et al., 1988); (Kempken et al., 1998). Transposons are

used as vectors for germ-line transformations of cloned genes (Jasinskiene et al., 1998; Loukeris et al., 1995; Rubin and Spradling, 1982; Spradling and Rubin, 1982). By incorporating a promoter-less reporter gene such as *lacZ* or *gfp* (green fluorescent protein), transposon vectors have been used to "trap enhancers" that regulate levels and patterns of expression of these heterologous genes (O'Kane and Gehring, 1987, Wilson, 1989). Transposon-induced mutations have been generated by targeted insertions or excisions [Ballinger, 1989, Kaiser, 1990, Rushforth, 1993). Another strategy for introducing a desired mutation is to use targeted gene replacement where a transgene serves as the template from which repair is directed following transposon excision (Plasterk and Groenen, 1992, Almeida, 1996).

Clearly, transposons are valuable tools. However, we know very little about what mechanisms, if any, might regulate transposition in response to a variety of signals. Numerous studies from across the phyla reveal that transposon activity responds to multiple factors as diverse as: genetic background of the host, developmental stage of the host, tissue or cell type, copy number of the elements as well as environmental factors imposed by temperature, nutritional status and intentional exposure to mutagens. This spectrum implies that a variety of regulatory mechanisms converge to control transposon activity. Certainly, a better understanding of the factors that regulate these unique elements will help researchers master transposons as precise, sophisticated molecular tools.

C. elegans as a model for studying transposons and their regulation

C. elegans provides an ideal experimental system for the study of transposon activity and regulation in part because of the powerful genetic methods available in this organism and because of the interesting regulation that transposons exhibit in its genome. *C. elegans* is a small free living terrestrial nematode that has a short (3.5 day) life cycle. The worms reproduce primarily as self fertile hermaphrodites. Males occur, although at a low frequency, and provide a means for cross fertilization. The average brood size of a single hermaphrodite is 200-300 (Brenner, 1974). The mature adult is approximately 1 mm in length and is transparent throughout its development. The ability to visualize development has allowed for detailed reconstruction of the complete cell lineage of this organism from the zygote to the resulting 959 somatic cells that comprise the entire worm including the epithelium, muscles, nerves, intestines, and gonad.

The genetics of this worm have been studied extensively, providing investigators with a detailed genetic and physical map of its genome. The genome contains 97 million base pairs coding for 19,099 genes that are distributed among six linkage groups (five autosomes and one sex chromosome). An effort to determine the complete nucleotide sequence was initiated in 1990 and completed in 1998 (Consortium, 1998). With this treasure trove of information, efforts are concentrating on correlating the entire physical sequence with the genetic map. The ultimate goal is to determine the biological role of each genetically defined unit of sequence.

The combination of its genetic history, its ease of maintenance in the laboratory and the ability to apply sophisticated genetic, molecular and biochemical techniques to this organism, makes *C. elegans* an excellent model to use in answering questions concerning the cellular factors that regulate transposon activity. Further, given the role *C. elegans* plays as an important model organism, now with a complete genome sequence, makes it imperative to understand transposon regulation in the organism. By defining the factors which control their activity, transposons can be used to their full potential as molecular tools to take full advantage of the sequence data that is now available.

Transposons in *C. elegans*

In *C. elegans*, the identification of transposons was inevitable but occurred somewhat serendipitously. Tc1 was first characterized in the early 1980's through the concerted efforts of a number of groups working on distinct aspects of the worm's genetic and physical makeup as summarized below. The objective of the community was to begin correlating specific molecular alterations with the corresponding genetically defined mutations. The overall goal was to begin reconstruction of the structural and functional relationships between gene products and the processes necessary for development, differentiation and ultimately evolution.

To address alterations that might be associated with development and to assess the rate of evolutionary divergence between related species, a molecular approach was pursued by Emmons and co-workers (1979). The methods chosen were Southern blot and restriction fragment length polymorphism (RFLP) analyses of genomic DNA probed with randomly generated genomic clones. Included among the strains analyzed were

isolates of Bristol (N2), the standard laboratory strain in use, and Bergerac (BO), another common but geographically distinct isolate. A key to these analyses was that one of the randomly generated clones used to probe the blots contained repetitive DNA from N2. Because Emmons expected to see very little variation between the N2 and BO, he was surprised to see that 15% of the restriction fragments were different or polymorphic between the two strains (Emmons et al., 1979).

Another group in the same lab also uncovered N2/BO polymorphisms while working to clone the actin genes of *C. elegans*. The polymorphisms cross hybridized with the probe Emmons used when he identified the N2/BO RFLP's (Files et al., 1983). The hybridizing probe was subsequently sequenced and characterized (Rosenzweig et al., 1983). The cloned fragment corresponded to a 1.7 kb element, the sequence of which revealed the structural hallmarks of Class II transposons: terminal inverted repeats, internal sequence containing an open reading frame, and dinucleotide repeats that flanked the element. However, the question was whether this element was capable of transposition.

The studies described above provided tantalizing molecular evidence for transposons in *C. elegans* but lacked the functional hallmark of transposons: spontaneous element mobility. The final evidence came from separate genetic studies of *unc-22* and *unc-54*, both genes encoding major muscle proteins in *C. elegans*. These large genes are very useful for genetic studies. Mutants have distinct phenotypes for which there are convenient assays: *unc-22* animals twitch and *unc-54* animals are paralyzed. Spontaneous revertants within a population of *unc* mutants are easily recognized as they

resume normal movement. Similarly, among normal WT populations, spontaneous *unc* mutations are also easily recognized. In either case, spontaneous mutations are rare and subsequent reversion events are even more infrequent. Therefore, using these muscle mutant phenotypes, active mutant hunts were initiated that soon led to the identification of unstable activities later recognized to be characteristic of transposons.

Genetic evidence of mobility came from the work of Moerman and Waterston (1986). They characterized unstable mutations of the *unc-22* gene that were later determined to be the result of Emmons' repetitive element, Tc1 (Moerman et al., 1986). Emmons' work had demonstrated that BO was polymorphic at multiple sites throughout its genome, therefore comparing N2 and BO would be useful for linkage mapping studies to localize mutants that Moerman would generate. During the course of their work, BO exhibited spontaneous germ line mutations in *unc-22* at unusually high rates of 10^{-4} and reversions at a rates of 10^{-3} to 10^{-4}. In comparison, spontaneous mutations in N2 occurred at rates less than 10^{-6}. The instability of the BO derived *unc-22* mutations was reminiscent of the unstable behavior of the transposon-induced mutations McClintock observed in maize (McClintock, 1948; Moerman and Waterston, 1984). Later, the mutagenizing agent was indeed determined to be a transposon in *C. elegans* (Moerman et al., 1986).

At the same time, Phil Anderson was conducting a directed effort to "trap" transposons in the large muscle gene, *unc-54,* by screening for spontaneous mutations with distinct phenotypes. This genetic trick proved unsuccessful until he was prompted to switch the "trap" from the genetic background of N2 to that of BO. Of 18 spontaneous

mutants recovered in BO, 10 were shown to be unstable (reverting to WT spontaneously). Subsequent analysis revealed Tc1 as the causative agent (Eide and Anderson, 1985b; Eide and Anderson, 1985).

The collective results of the molecular and genetic efforts described above defined structurally and then functionally Tc1, the first family of transposons in *C. elegans*. The important key in all these studies was the BO background in which activity was observed. Once revealed as a "mutator" strain, this feature brought BO to the forefront of transposon biology in *C. elegans*. BO sustained high frequencies of spontaneous unstable mutations particularly in the germ line. For the worm community, the Tc1 transposon provided a new and important tool to facilitate the molecular genetic methods necessary to correlate the genetic and physical maps. In addition, BO represented fertile ground for those few who wished to study the regulatory factors responsible for the observed strain-specific and tissue-specific differences.

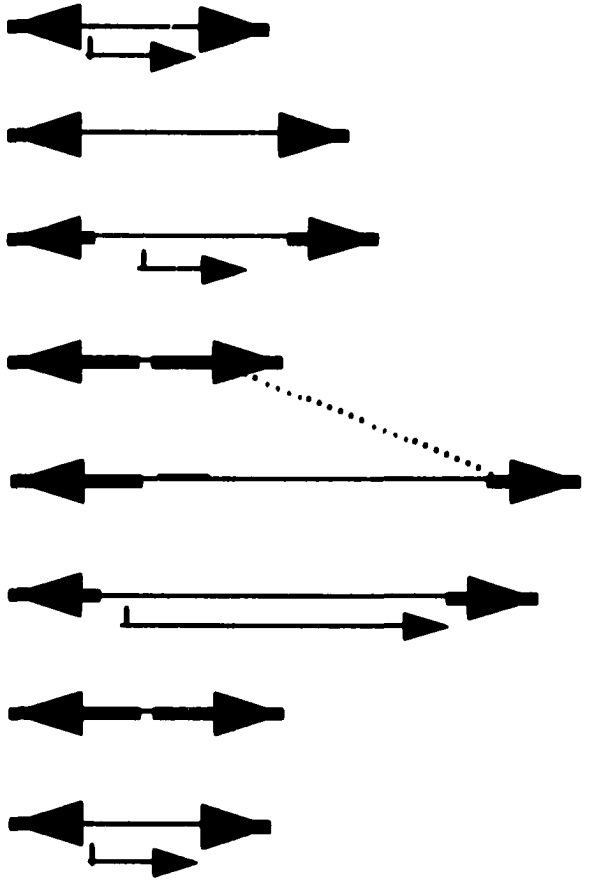
Transposon families in *C. elegans*

Since identification of Tc1, several additional families of transposons have been identified in *C. elegans*. These include Tc2, Tc3, Tc4, Tc5, Tc6 and Tc7. Of these families, those known to be active include Tc1, Tc3, Tc4 and Tc5. Like all Class II elements, they share basic structural features including terminal inverted repeats of variable size, internal sequence also of variable size and often an element encoded transposase (Table 1. Adapted from (Plasterk and van Luenen, 1997).

Table 1. Transposon families in *C. elegans* and their general structure. Adapted from Plasterk and van Luenen, 1997. See text for references.

Element	Size (kb)	IR (bp)	Insertion Site	ORF (AA)	Copy # N2	Copy # BO
Tc 1	1.6	54	TA	343	30-50	300-500
Tc 2	2.1	24	TA	-	6	~20
Tc 3	2.3	462	TA	329	15-18	15-18
Tc 4	1.6	774	CTNAG	-	~20	~20
Tc 4v	3.5	139	CTNAG	-	5	5
Tc 5	3.2	491	CTNAG	532	4-6	4-6
Tc 6	1.6	765	TA	-	~24	~24
Tc 7	0.9	345	TA	-	~30	~30

General Structure of DNA transposons in *C. elegans*



Among the active elements, Tc1 is the most abundant transposon family in *C. elegans*. There are approximately 30 to 50 copies that reside in the genome of N2 while 300 to 500 copies are found in BO's genome (Egilmez et al., 1995; Emmons et al., 1983; Liao et al., 1983). Full length Tc1 elements are 1610 bp in length including inverted repeats of 54 bp. The internal sequence contains an ORF that encodes a 343 amino acid transposase, Tc1A. This protein contains the conserved DD(34)E motif implicated in the "cut-and-paste" transposition mechanism (Doak et al., 1994; Vos and Plasterk, 1994; Vos et al., 1993). Tc1 transposons insert only at TA dinucleotides, duplicating the target sequence upon insertion.

The Tc3, Tc4 and Tc5 were all discovered as causative agents of spontaneous mutations isolated in the background of a "mutator "strain, TR679 (genotype: *mut-2(r459) unc-54 (r323::Tc1rv)*). TR679 was isolated from an EMS mutagenesis screen for mutants that enhance germ line transposition of Tc1. TR679 is a recombinant inbred line that harbors the *mut-2* mutator. The genetic background and derivation of strains harboring the *mut-2* mutator are described below. The features of these strains implicate *mut-2* in a mechanism that regulates many distinct transposons genome-wide in *C. elegans*. The characterization of *mut-2* is the focus of this dissertation and will be described in detail later.

Tc3 elements are found in both N2 and BO at 15 to 18 copies per haploid genome. The Tc3 family is 2335 bp with terminal IRs of 462 bp and internal sequence encoding a 327 amino acid transposase, Tc3A. Tc3 is similar to Tc1: it targets TA dinucleotides, itstransposase, Tc3A, contains the DD(34)E motif, and shares 34 % identity with Tc1A,

and the Tc3 termini share sequence identity with critical nucleotides of the Tc1 IRs (Collins et al., 1989). Despite the similarities to Tc1, Tc3 elements are only mobilized by Tc3A (van Luenen et al., 1993).

Tc4 was identified as unique insertions into *unc-86* and *ced-4* (Finney et al., 1988; Yuan et al., 1991). The Tc4 family is characteristic of the Class II type of fold-back elements. Both N2 and BO harbor approximately 20 copies. It is about 1600 bp in length with 774 bp IRs flanking a short internal sequence that contains no obvious ORF. When it mobilizes, Tc4 inserts at a target site characterized by the sequence, CTNAG.

A subfamily of Tc4, Tc4v, is present at approximately 5 copies per haploid genome in N2 and BO. It is 3500 bp in size due to an insertion of 2343 bp that interrupts one IR, replacing 447 bp of internal sequence. A transcript capable of encoding a 537 amino acid protein has been isolated from Tc4v but lacks similarity to other transposases (Li and Shaw, 1993). While transposase activity has not yet been demonstrated for this protein, it is conceivable that such an enzyme might support the activity necessary to mobilize all Tc4 family members.

Tc5 elements, like Tc3, were identified as novel insertions in *unc-22*. Tc5 elements are found at 4 to 6 copies per haploid genome of both N2 and BO (Collins and Anderson, 1994). They are 3171 bp in length including perfect terminal IRs of 492 bp. The internal sequence encodes an ORF of 532 amino acids (Collins and Anderson, 1994). Tc5 shares some similarities with Tc4 and Tc4v: it targets the sequence CTNAG, the predicted transposases are 33 % identical (Tc5 vs Tc4v) and 8 of 10 terminal nucleotides are identical (T5 vs Tc4). Tc5 shares no significant similarity with Tc1 or Tc3.

The remaining transposon families include elements that exhibit transposition under experimental conditions or in situations where Tc1 is also known to be active suggesting a degree of co-regulation. The Tc2 family of elements vary in size and are about 2074 bp with IRs of 24 bp. The internal sequence includes several potential ORFs with no significant homology to known transposases (Ruvolo et al., 1992). There are about 6 copies of Tc2 present in N2 and about 20 in BO. Analysis of integrated Tc2 elements shows that they targets TA dinucleotides. Germ line transposition of Tc2 has only been noted in the hybrid progeny of crosses between N2 and BO.

Tc6 transposons are fold-back elements similar in structure to Tc4. There are about 24 elements present in all strains of *C. elegans*. Clones of several Tc6 elements appear to be truncated or altered by internal insertions of other Tc6 elements. Intact Tc6 elements are about 1600 bp long with 765 bp IRs that flank a short, unique internal sequence. Tc6 termini share some sequence similarity with the IR termini of both Tc1 and Tc3 but Tc6 has not been shown to move in response to either Tc1A or Tc3A, nor has it been shown to transpose in the soma or germline (Dreyfus and Emmons, 1991).

Tc7 was described as an element that moved under conditions where the Tc1 transposase, Tc1A, was overexpressed, hence its characterization as a "Tc1 hitchhiker" (Rezsohazy et al., 1997). Tc7 elements are present at about 30 copies per haploid genome. They are 921 bp with 345 bp IRs that flank internal sequence lacking any ORF. The termini share similarity to the Tc1A binding site of Tc1 termini providing a possible explanation for its response to overexpressed Tc1A. Analysis of integrated Tc7 elements reveals that Tc7 also targets TA dinucleotides. Transposition in the soma has been

demonstrated under experimental conditions but Tc7 has not been shown to transpose in the germ line (Rezsohazy, pers. comm.).

Identification of these transposon families in *C. elegans* provided the worm community with an important molecular tool that had previously only been available in other systems (insertion elements in prokaryotes, P elements in *Drosophila* and transposable elements in maize). Discovery of transposons in worms provided the scientific community with another model system for studying the precise regulation of these elements.

What do we know about transposition in *C. elegans*?

Most of what is known about transposition in *C. elegans* comes from initial genetic characterization of unstable Tc1-induced *unc-22* and *unc-54* mutations. Subsequent biochemical work with the Tc1 and Tc3 transposases revealed important information about specific mechanistic requirements (van Luenen et al., 1994; van Luenen and Plasterk, 1994; Vos et al., 1993). The model that has emerged provides an explanation for some reaction features including the duplication of the target site and the characteristic footprints at donor sites and is based on the "cut-and-paste" mechanism described previously (Mizuuchi, 1992) (Figure 3). However, many of the precise accessory factors responsible for the regulation of transposition remain a mystery.

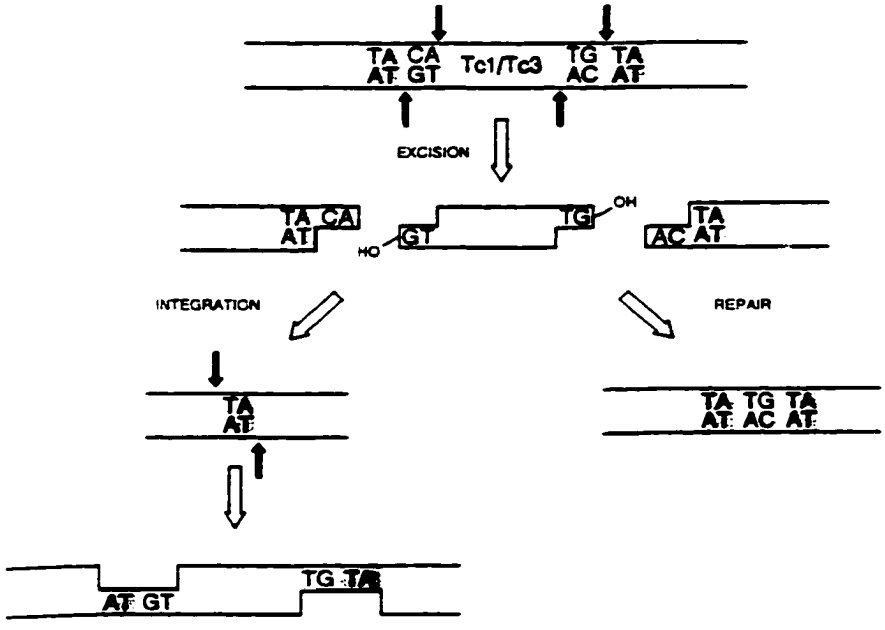
From a broad range of genetic and biochemical work, we know that transposons have distinct sequence specific targets; Tc1, Tc2, Tc3, Tc6 and Tc7 target TA dinucleotides and Tc4 and Tc5 target CTNAG. From studies with Tc1 and Tc3, we know that target site preference is influenced by a weak consensus sequence flanking the target

sequence (Eide and Anderson, 1988; Ketting et al., 1997; Korswagen et al., 1996; Rosenzweig et al., 1983; van Luenen and Plasterk, 1994). Of the many target sites available, there are "hot spots" or preferred sites for insertion. Evidence from studies with targets engineered into relaxed or super-coiled plasmids suggest that chromatin structure might be a common feature that contributes to access and insertion at these preferred sites (Ketting et al., 1997).

We know that transposition results in duplication of the target sequence and excision results in footprints at the donor site (Eide and Anderson, 1988; Ruan and Emmons, 1984; van Luenen and Plasterk, 1994). If excision is precise, the footprints can be simple duplications. If excision is imprecise, the footprints can result in large, complex and disruptive alterations with severe consequences for gene expression. Such alterations are thought to be the result of transposon-independent DNA repair processes following excision (Mizuuchi, 1992). The host repair mechanisms make use of the best available template when imprecise excisions remove large portions (up to several kb) of the donor site. In some cases, the template may be an exogenously supplied extrachromosomal template and not the intact homologous sequence. This is a feature exploited in targeted gene replacement strategies (Almeida, 1996; Gloor et al., 1991).

A number of studies have focused on the mechanistic requirements of transposition. The results of *in vitro* cell-free transposition systems with plasmid derived donor and target sequences in the presence of Tc1A or Tc3A suggested that the element-encoded transposases were sufficient for mobility (Vos et al., 1996). These studies helped define the minimal requirements for transposition but do not preclude a role for

Figure 3. Cut-and-paste model for transposition in *C. elegans*



Studies of Tc1 and Tc3 provide the basis for this model which provides an explanation for the footprint often found after excision of an element and the duplication resulting from element insertion into a new target site (figure taken from Plasterk and van Luenen, 1997).

host-encoded factors. Genetic and molecular evidence supports the existence of such host-encoded factors; the mutators, *mut-2* and *mut-7*, are important examples.

We know from early genetic observations of Tc1 and later of Tc3, Tc4 and Tc5, that rates of transposition are influenced by tissue type and genetic background. Somatic transposition events are known to occur frequently (Harris and Rose, 1986; Glasner, 1996). In contrast, germ line excision events are readily detected in mutator strains such as BO but reduced or silent in N2. Analyses of several high copy number strains revealed that germ line transposition was frequent only in BO but not other strains suggesting that genetic background was more important than copy number ((Moerman and Waterston, 1984; Mori et al., 1988). Further support for the influence of tissue specificity and genetic background comes from observations that germ line activation of Tc3, Tc4 and Tc5 occurs only in the presence of the mutator allele, *mut-2(r459)* (Collins et al., 1989; Collins et al., 1987; Collins and Anderson, 1994).

Identification of Mutators - Genetic Factors Regulating Transposition

Screens for the genetic factors responsible for regulating transposition uncovered several mutator alleles from the "high hopper" strain BO. These included *mut-4, mut-5* and *mut-6.* Subsequent studies of these mutators suggested that they were mobile and represented transposons with altered mobility and target specificity. Their apparent regulatory activity was thought to be analogous to that seen between autonomous and nonautonomous pairs of elements where mobility of one member (nonautonomous) depends on the presence or expression of the other (autonomous). Examples of this

system of mobilization are well characterized in maize (*Ac/Ds;* (McClintock, 1951)) and *Drosophila* (P elements and *mariner*; (Engels, 1983; Medhora et al., 1991).

In an effort to uncover tissue specific regulators of transposition, another distinct mutator, *mut-2,* was identified (Collins et al., 1987). The mutator allele, (r459), was derived from an ethyl methanesulfonate (EMS) mutagenesis screen of the strain, TR445 (*unc-54(r323::Tc1*). Characterization of strains harboring *mut-2*(r459), revealed a 50- to 100-fold enhancement of germ line activity of all known transposon families. In addition, a subtle increase in frequency of males, referred to as the "Him" phenotype was observed in populations harboring this mutator. Efforts to map, clone and characterize this gene are the focus of this dissertation and are described below in detail.

Still another mutator, *mut-7,* was identified from a similar EMS mutagenesis screen of N2 (Ketting et al., 1997). The screen resulted in a strain with the ability to mobilize multiple families of transposons in the germ line but the transposition frequencies are 10-fold less than what is seen with *mut-2.* Interestingly, *mut-7* also exhibits a Him phenotype. This mutator maps to LG III and was recently cloned and identified as an RNAse D gene (Ketting, 1999). How an RNAse D participates in regulating transposition is unclear.

Clues for potential roles of both *mut-2* and *mut-7* may be found in recent developments regarding the consequences of widely used molecular techniques such as DNA transformation and interference by ectopically supplied double-stranded RNA (RNAi) provide some. Analysis of transgene expression in the germ line and soma demonstrated that the germ line of N2 will selectively silence or repress expression from

ectopically introduced genes (Kelly et al., 1997). However, there is evidence that the germ line of mutator strains is permissive for transgenic expression (Kelly, pers. comm.). In another investigation of the mechanism underlying RNAi, some mutator strains were found to be resistant to the effects of RNAi (*mut-2* and *mut-7* but not *mut-6*) (Tabara et al., 1999). These two newly identified characteristics of *mut-2* and *mut-7*, further support global roles in the processes that maintain genome integrity specifically in the germ line.

Transposons as tools in *C.elegans*

Of all the elements, Tc1 has been the workhorse for a variety of molecular genetic approaches. The techniques most often used include transposon-tagging for cloning (Greenwald, 1985), targeted gene-replacement strategies (Almeida, 1996; Plasterk and Groenen, 1992), and transposon-induced mutagenesis (Rushforth et al., 1993). An example of the latter that has developed into a large scale effort is the use of Tc1 to generate a large frozen bank of thousands of worm populations harboring Tc1-induced mutations (Zwaal et al., 1993). Mutations of a gene of interest can be recovered in the form of clonal populations by using sequence specific primers and a PCR-based strategy to screen DNA isolated from successive pools of these populations. This provides a tremendous resource available to the worm community to quickly identify mutants and study the associated alterations in normal gene function.

These molecular genetic strategies are widely used but represent the utility of a single abundant transposon family. To better exploit the full potential of all transposons requires a more complete understanding of how their activity is regulated. By identifying the factors that control transposition of all elements, the number and type of transposons

available as tools will be increased, the scope of targets broadened and the techniques already in use refined. In addition, it is apparent that understanding how transposons are regulated in the genome will provide insight into more global aspects of the dynamic processes required to maintain overall genomic integrity and plasticity.

The *mut-2* mutator : A regulator of transposons in *C. elegans*

In an effort to elucidate the mechanism behind germ line specific control of transposon activity, Collins and co-workers conducted a screen of an ethyl methanesulfonate (EMS) treated strain, TR445 (*unc-54(r323::Tc1)* (1987). This strain carries a Tc1 insertion in *unc-54* rendering the worm paralyzed. After mutagenesis, individuals from clonal populations exhibiting enhanced rates of transposition (observed as reversion to wild type movement) were repeatedly crossed to N2 and re-isolated based on enhanced reversion rates of the BO allele, *unc-54*(r323*::Tc1)* . Thus, recombinant inbred mutator lines demonstrating high levels of Tc1 transposition were established.

From this screen, *mut-2(r459)* was isolated as a strain, TR674, that consistently exhibited high frequency transposition and excision of Tc1 in the germ line (Collins et al., 1987). This mutant allele of *mut-2* has been termed the *mut-2* "mutator"; the normal regulatory features of this gene had been altered by mutation. Subsequently, the *mut-2* mutator was shown to "activate" transposition of Tc3, Tc4, and Tc5. These elements are only known to transpose in the presence of the *mut-2* mutator (Collins et al., 1989; Collins and Anderson, 1994; Yuan et al., 1991). Another interesting feature noted upon characterization of the *mut-2* mutator was a higher incidence of males known as the "Him" phenotype. Fertile males arise in natural populations as a result of rare X

chromosome non-disjunction events during meiosis at a frequency of <0.1% (Hodgkin and Brenner, 1977). Males from *mut-2* mutator populations are sterile and occur at frequencies ≥0.4% (Collins et al., 1987). The basis for sterility and the increased frequency is not understood.

As mentioned above, the genomic composition of TR674 (genotype: *mut-2(r459) unc-54(r323::Tc1)*, TR679 (genotype: *mut-2(r459) unc-54(r323::Tc1rv)*, and their derivatives are essentially N2 except for the sequence surrounding the BO derived *mut-2(r459)* and *unc-54(r323::Tc1)* alleles. This can be confirmed by confirming the presence or absence of N2/BO polymorphic markers on the right arm of LGI to which these alleles map (see Results). This feature becomes important for the mapping strategies described below and provides a potential source of physical markers (such as transposons or sequence level polymorphisms) that define the interval harboring the *mut-2* gene.

Additional characterization of *mut-2* phenotypes

Initially, *mut-2* was perceived as a valuable tool that could enhance the utility of Tc1-based cloning and mutagenesis strategies. *mut-2* is useful for two reasons: it results in an increase of germ line transposition events and it expands the menu of available transposons to include Tc3, Tc4 and Tc5. While *mut-2* proved useful to the application of techniques such as transposon tagging and Tc1-induced deletion mutagenesis, the more fundamental question of *mut-2*'s normal biological role remained largely unanswered. My work begins to address this question through efforts to map and identify the gene. During the course of this work, I characterized an additional *mut-2* phenotype:

temperature sensitive (ts) sterility. Worms raised at 25°C that harbor the *mut-2* mutator gene are sterile due to an apparent defect in gametogenesis. The ts phenotype proved invaluable for employing traditional rescue techniques and for providing additional evidence of *mut*-2 's broader role in germ line processes and genome maintenance.

In separate but related studies of phenomena associated with the environment of the germ line, two new phenotypes have been attributed to the *mut-2* mutator. First, there is preliminary evidence suggesting that the *mut-2* mutator de-silences transcription in the normally quiescent germ line (Kelly, pers. comm.). Secondly, in other work, *mut-2* and *mut-7* strains demonstrate altered aspects of RNA metabolism. Specifically, they are unaffected by microinjected double stranded RNA (dsRNA) that otherwise interferes with gene expression in the phenomenon of RNAi (Tabara et al., 1999)).

In the case of germ line transcription, Kelly and co-workers (1997) investigated the basis for regulated expression of transgenic arrays microinjected into the gonads of *C. elegans*. They found that the germ line of WT worms selectively silences expression from transgenic arrays. Silencing can be relieved in worms that are defective for a group of transcriptional repressors, the maternal effect sterile genes (*mes-2, mes-3, mes-4* and *mes-6*) (Kelly, 1998). Sterility in a *mes* mutant background is due to a defect in gametogenesis such that few to no gametes ever develop in the progeny of the affected worms. Of the *mes* genes, *mes-2* and *mes-6,* are similar to the Polycomb genes which are thought to regulate transcription by remodeling chromatin structure (Holdeman, 1998; Korf, 1998). Additional screens for de-silencing mutations revealed that the germ line of the *mut-2* and *mut-7* mutators are also de-silencing (Kelly, pers. comm.).

The other phenomenon is related to the widely used technique of RNAi. This has become an important molecular tool that is widely used but mechanistically not well understood, not unlike transposon mobility in *mut-2* mutator backgrounds. RNAi is a method to ectopically reproduce the null mutant phenotype of a gene of interest by microinjection with dsRNA generated from cloned cDNA fragments of the gene (Izant and Weintraub, 1984; Nellen and Lichtenstein, 1993). Studies have determined that dsRNA is more effective than sense or anti-sense single stranded RNA alone (Fire et al., 1998). Screens for mutations that affect this aspect of RNA metabolism revealed that *mut-2* and *mut-7* strains are resistant to interference by dsRNA (Tabara et al., 1999).

An interesting question is how *mut-2* and *mut-7* participate in this process. This has been addressed in part by evidence that *mut-7* is an RNAse D homologue which targets dsRNA (Ketting, 1999). One model that links activated germ line transposition and RNAi resistance proposes that inappropriate read-through transcription of entire transposons fold back on themselves and produce dsRNA intermediates. These intermediates might prime specific RNA metabolic pathways that will then target ectopically introduced dsRNA and/or their most likely source: the endogenous gene. This provides an explanation for RNAi resistance of mutator strains but does not readily address the Him and Mut phenotypes.

Some biology of the germ line

The question that I wish to address is how *mut-2* participates in multiple aspects of germ line biology: transposition, chromatin behavior, gametogenesis and RNA metabolism. All the phenotypes that characterize the *mut-2* mutator provide important clues. Global activation of multiple transposon families and the Him phenotype are superficially distinct events but share a number of features with transcriptional silencing and RNA metabolism. They all are processes that occur in the germ line and they involve activity of complexes that monitor the status of chromatin. These complexes include repair complexes, transcriptional activators/repressors and the chromatin itself.

As a specialized tissue, the germ line maintains a highly constrained chromatin environment. To successfully generate viable gametes, the germline must silence inappropriate transcription and monitor the status of chromatin throughout meiosis. The *mut-2* mutator phenotypes likely represent the consequences of disrupting the coordinated germ line processes that maintain genome integrity. Therefore to understand *mut-2*'s role in this environment, we must consider the dynamics of chromatin in the germ line and some related aspects of transposition.

Signals of chromatin status: Double strand breaks

One example of indicators of chromatin integrity include double strand breaks (DSBs) of DNA. DSB's in chromatin are important phenomena because they signal damage or are intermediates during cell cycle progression. During transposition, DSBs also play an important role. The transposition model proposes that a transposase complex

binds its element and cleaves it from the donor site, producing DSBs. The transposase also identifies a target sequence and stabilizes a transposon:target complex during double strand cleavage and strand transfer (Grindley and Leschziner, 1995; Mizuuchi, 1992). The gaps created at the donor site and at the transposon:target site are presumably repaired by host mechanisms that are signaled and recruited to the complex. It is not clear whether the signals for repair are transduced solely by DSBs, by other components of the transposase complex or both.

DSBs are also an integral part of meiosis, a process unique to the germ line (Kleckner, 1996; Zetka and Rose, 1995). Meiotic cell division is initiated by the alignment and synapse of homologous chromosomes. During this process, DSBs promote cross-overs between homologues that may be resolved by recombination. Meiosis is highly coordinated and monitored by cell cycle check points that halt division and signal repair pathways in response to unresolved DSB intermediates. In *C. elegans*, Him mutations are those that decrease meiotic recombination and increase non-disjunction of the X chromosome (Hodgkin et al., 1979).

Signals of chromatin status: Double strand RNA

Other important indicators of chromatin status are dsRNAs. These nucleic acid intermediates have been shown to effectively interfere with transcription across a broad range of phyla including plants (Hart et al., 1992; Voinnet et al., 1998), worms (Fire et al., 1998; Montgomery and Fire, 1998), flies (Kennerdell and Carthew, 1998) and protozoa (Ngo et al., 1998). The mechanism of dsRNA interference is unclear but several studies have revealed a number of characteristic features. Microinjection of dsRNA to an

upstream gene in an operon did not interrupt expression of the downstream gene suggesting that interference by dsRNA is gene specific and does not involve initiation or elongation. (Montgomery et al., 1998). Examination of the endogenous target revealed that RNAi does not alter the primary sequence of the target gene (Montgomery and Fire, 1998; Montgomery et al., 1998). Injections at very low dsRNA concentrations were also effective suggesting that the dsRNA does not associate stoichiometrically with endogenous mRNA or the transcription unit (Fire et al., 1998). dsRNA appears to suppress or silence expression post-transcriptionally; dsRNA's derived from exon sequences interfere, dsRNA's derived from intron sequences don't (Fire et al., 1998). Studies in plants and *C. elegans* show that the signal and effects of dsRNA are transmissible between cells and subsequently amplified (Fire et al., 1998; Tabara et al., 1998; Voinnet et al., 1998). In the worm, interference is observed even after feeding on bacteria expressing dsRNA (Timmons and Fire, 1998).

Degradation of dsRNA intermediates are thought to be distinct from the SMG system of mRNA surveillance that targets aberrant transcription products (Montgomery and Fire, 1998; Montgomery et al., 1998). One model for the existence of a dsRNA interference mechanism proposes that it serves as a host defense mechanism induced by retroviral replication intermediates to protect the host from infection. Another compelling model that does not exclude a defensive role, is that RNAi might represent a natural method for gene specific silencing.

Signals of chromatin status: Modifiers of Architecture

The structure of chromatin and mechanisms that modify it also serve as important regulators of transcriptional states. Genomic DNA is an unwieldy molecule that is extensively packaged. The basic subunit is a nucleosomal core of chromatin wrapped around an octamer of histones that are further condensed through several levels of compaction. This architecture effectively serves a repressive function by restricting access to genes, promoters and *cis*-acting regulatory factors.

Several such factors have been identified and have been shown to contribute to maintaining repressive chromatin architecture. These factors assemble to form the yeast silencing complexes, SWI/SNF complexes, or Polycomb Group (PcG) of developmentally specific repressors. Many of these complexes are thought to function by assembling on or interacting with DNA binding proteins that recruit them to specific sites. Once assembled, these complexes effectively repress expression in general (silencers) or in a developmentally specific manner (PcG) (reviewed by Pirotta, 1997). Access to transcriptional units or cis-acting sequences can also be regulated by the degree of the DNA's association with the nucleosomal core. Under-acetylated or de-acetylated histones associate tightly with the DNA wrapped around them but acetylation at specific lysines of the histones will destabilize the association and permit access by remodeling factors (reviewed by Tsukiyama and Wu, 1997).

All these remodeling complexes and modifiers contribute to the general maintenance of a healthy germ line by restricting access and transcription especially

during periods of DNA synthesis and replication. Mutations affecting the processes that patrol the germ line environment and signal quiescence (or activation as the need arises) result in defects in chromatin behavior. Ultimately, the effect of a single mutation or class of mutations can have broad ranging affects on the overall viability of the germ line.

The goal: What is *mut-2's* role in maintaining chromatin stability

The *mut-2* mutator provides a unique opportunity to investigate a host encoded factor that participates in regulating germline activity of transposable elements and contributes to the global dynamics of the genome. My goal is to understand the function of this interesting gene. To address *mut-2*'s role, I set out to precisely map and identify the best candidate gene taking full advantage of the molecular and genetic techniques available for *C. elegans* and its completed genomic sequence.

MATERIALS AND METHODS

Nematode Culture Methods

All nematode strains were grown and manipulated according to standard procedures (Brenner, 1974). Nematodes were maintained at 20°C except when determining the viability of *mut-2* mutator strains at 25°C. For genetic mapping and interaction experiments, strains were cultured and maintained by feeding with *Escherichia coli* strain OP50 seeded onto solid nematode growth media (NGM: 3 g/L NaCl, 17 g/L agar, 2.5 g/L peptone supplemented with 1 mM $CaCl_2$, 1 mM $MgSO_4$, 1 mM $KHPO_4$ pH 6.0, and 5 μg/ml (ethanol) cholesterol). For isolation of nucleic acids, strains were cultured in S-medium (0.1 M NaCl, 0.05 M $KHPO_4$ (pH 6), 5 mg/L cholesterol (5 mg/ml ethanol stock), 10 mM KCitrate (pH 6), 3 mM $CaCl_2$, 3 mM $MgSO_4$ and 10 mls/L trace metals (trace metals stock: 5 mM Na_2EDTA, 2.5 mM $FeSO_4 \cdot 7H_2O$, 1 mM $MnCl_2 \cdot 4H_2O$, 1 mM $ZnSO_4 \cdot 7H_2O$, 0.1 mM $CuSO_4 \cdot 5H_2O$). Worms grown in liquid culture were fed a 40 X stock of *E. coli* RR1, a bacterial strain that does not maintain plasmids therefore reducing opportunities for bacterial-derived plasmid contamination.

C. elegans strains and Nomenclature

Most nematode strains were obtained from the *Caenorhabditis* Genetics Center (CGC). TR445, TR674, TR1034 and TR1175 were obtained from Phil Anderson. The transgenic line, BC5732 was obtained from Diane Janke in David Baillie's lab and the transgenic lines, JH1271 and JH1272 were generated by Kim Reese in Geraldine

Seydoux's lab. I generated all other strains during the course of this work. These strains are listed in Table 2 along with their genotypes.

The wild type strains commonly used in research laboratories are N2 variety Bristol (Bristol or N2) or N2 variety Bergerac (BO). For studies investigating transposon activity, BO strains have been the preferred strains because they harbor many more copies of active transposons than Bristol strains, hence the designation "high hopper". TR445 is a recombinant inbred strain that bears the BO-derived allele *unc-54(r323::Tc1)* that is a transposon (Tc1)-induced muscle mutation resulting in paralysis. TR445 was mutagenized with ethyl methanesulphonate (EMS) and subsequent populations were screened for enhanced germ-line transposition of Tc1. Transposition was detected as a significant increase in the frequency of reversion to wild type movement (Collins et al., 1987). The EMS screen generated the *mut-2* mutator strain, TR674 (*mut-2(r459) unc-54(r323)::Tc1)*, which exhibits the greatest frequency of Tc1 transposition and excision (1×10^{-3}).

All mapping strains used in this study carry genetic markers on Linkage Group I derived from 100% Bristol N2 genetic backgrounds (Table 2A). Except for the interval that contains *mut-2*, all recombinant strains derived from mapping crosses represent replacement of the BO genotype with a greater proportion of Bristol chromosomal complement. This can be verified by assaying for the presence of known N2/BO polymorphisms by PCR or Southern analysis. For example, sP1 and hP9 are polymorphisms on LG1 that are present in BO but not N2.

Table 2. List of strains and their genotypes used in mapping and transgenic strain construction.

A. Mapping Strains:	
N2	WT variety Bristol
EM1002	WT variety Bergerac (aka BO)
BC244	*dpy-14 (e188) let-86(s141)+unc-13(e51)/++unc-15(e73)+*
CB188	*dpy-14 (e188)*
MT3184	*sem-4 (n1378) unc-13 (e51)*
TR445	*unc-54 (r323::Tc1)*
TR674	*mut-2 (r459) unc-54 (r323::Tc1)*
TR679	*mut-2 (r459) unc-54 (r323::Tc1rv)*
TW195	*dpy-5 (e61)*
TW404	*dpy-5 (e61) mut-2(r459)*
TW409	*mut-2 (r459) sem-4 (n1378) unc-13 (e51)*
TW410	*mut-2 (r459) sem-4 (n1378)*
B. Reporter Strains	
TR1034	*unc-22 (r644::Tc5)*
TR1175	*unc-22 (r765::Tc4)*
C. Duplication Strains	
KR16	*unc-11(e47) dpy-5(e61) I; sDp2 (I, f)*
KR1755	*dpy-14 (e188) unc-13 (e51) I; hDp65 (I,f)*
KR1758	*dpy-5 (e61) dpy-14 (e188) rec-1 (s180)? I; hDp62 (I,f)*
TW417	*mut-2 (r459) sem-4 (n1378); sDp2 (I,f)*
TW422	*mut-2 (r459) sem-4 (n1378) unc-13 (e51); hDp65 (I,f)*
TW423	*dpy-14 (e188) mut-2 (r459) sem-4 (n1378) unc-13 (e51); hDp62 (I,f)*
D. Transgenic strains	
BC5732	N2; *sEx744*[C48B6(I) + pCes1943(*rol-6(su1006)*]]
JH1271/2	N2; *axEx* [H06O01(I) + pCes1943(*rol-6(su1006)*]]
TW414/415	*mut-2 (r459) sem-4 (n1378); sEx* rol-6::C48B6
TW420	*mut-2 (r459) sem-4 (n1378); axEx* rol6::H06O01
TW426	N2; cjEx[N2 + T21G5, C48B6, T10B11, C30F12, H06O01, E02H2 (I)+ pCes1943[*rol-6(su1006)*]]
TW427	N2; *cjEx* [N2 + T10B11 (I) + pCes1943 [*rol-6(su1006)*]]
TW428	N2; *cjEx* [N2 + C30F12 (I) + pCes1943 [*rol-6(su1006)*]]
TW429	N2; *cjEx* [N2 + H06O01 (I) + pCes1943 [*rol-6(su1006)*]]

Strains carrying free duplications that span the entire region or cover segments of the *mut-2* interval were obtained from CGC and maintained according to recommendations. Free duplications are generally maintained as single extrachromosomal copies which segregate in a non-Mendelian fashion. The duplications are maintained in a reference strain with a mutation that is rescued by the extrachromosomal element. Progeny that segregate characteristic phenotypes allow one to readily identify animals that maintain the duplication (Herman et al, 1976; 1979).

Using standard techniques described below, transgenic strains were constructed by germ-line co-injection of a plasmid carrying the dominant marker *rol-6(su1006)* with cosmids selected from the *mut-2* interval. The plasmids used were either pRF4 or pCes1943. pCes1943 is essentially pRF4 plus a kanamycin^R cassette to facilitate recombination with kan^R-carrying cosmids used in microinjection protocols. The co-injected DNAs form arrays which are maintained extrachromosomally. The resulting arrays were moved into the appropriate *mut-2* mutator background by standard genetic crosses and the relevant phenotypes were assayed as described below.

Phenotype Analysis (Him, Twitchers and the temperature sensitive phenotype)

Several phenotypes were used to screen populations of worms for the presence of *mut-2(r459)*. Mutator strains are characterized by a high incidence of males (Him) and a greater frequency of transposon insertion and excision events (Mut). During the course of this study, I also identified a temperature sensitive sterility phenotype (viable at 20°C but sterile at 25°C) that co-segregated with the Him phenotype in *mut-2* mutator strains. The

sterile phenotype was used to assess rescue of the *mut-2* mutator strains when genomic fragments carrying candidates for the normal wild type *mut-2*(+) gene were supplied.

The Him phenotype Putative *mut-2(r459)* bearing populations derived from crosses for mapping or genetic introduction of rescuing fragments were screened and scored for the occurrence of males. Strains were scored by seeding plates with 1 to 10 young L4 hermaphrodites and allowing them to grow to a population size of at least 200 to several thousand (one to several generations). Populations were screened and males counted and removed (to prevent cross fertilization) if the plates were retained for additional scoring. Populations with frequencies greater than 0.4% were considered to be Him. Candidate strains were then crossed to two different reporter strains: *unc-22(r745)::Tc4* and *unc-22(r644)::Tc5*.

The Twitcher Assay for enhanced transposition The *unc-22* gene encodes Twitchin, a member of a superfamily of giant muscle proteins (Benian et al., 1989; Benian et al., 1996). Both Tc4 and Tc5 insertions affect the *unc-22* gene by disrupting the normal coding region and altering the structure of the UNC-22 Twitchin protein. Strains bearing the transposon-induced mutations exhibit a characteristic twitching and are referred to as "twitchers". Subsequent populations derived from the progeny of a test cross between the reporter strains and the candidate *mut-2* mutator strains were screened for reversion of the twitchers to non-twitchers indicating the precise excision of the transposon. Excision of Tc4 or Tc5 is only known to occur in the *mut-2(r459)* genetic background providing a *mut-2* specific assay.

The temperature sensitive (ts) assay The temperature sensitive phenotype manifests as an effect of sterility for *mut-2* mutator animals raised at 25°C. Eggs plated at 25°C develop to adulthood but never produce eggs. L4 hermaphrodites plated at 25°C produce a reduced brood size which hatch and mature but are also sterile.

Interval Mapping of *mut-2*

Precise genetic mapping of the *mut-2(r459)* allele was achieved by scoring recombinants derived from standard mapping crosses between *mut-2* strains and strains with well characterized genetic markers. For example, the strain, TW404, carries the mutant allele *dpy-5* closely linked to *mut-2* on the right arm of Linkage Group I (LGI). It was constructed by mating TR674 (*mut-2(r459) unc-54::Tc1*) L4 hermaphrodites with TW195 (*dpy-5)* heterozygous males (parental generation/P0). F1 wild type cross progeny were picked individually and allowed to self-fertilize. Animals that segregated 25% Dpy progeny were retained and F2 wild type individuals were isolated and allowed to self fertilize. F3 progeny were screened for Dpy Unc recombinants which were then used to establish Dpy Unc lines to be tested for the presence of *mut-2(r459)* based on two of the phenotypes described above. These Dpy Unc lines were scored for the Him phenotype and enhanced excision rates of Tc1, Tc4, and Tc5 which resulted in reversion to wild type movement. Upon verifying the presence of *mut-2(r459)*, a line of Dpy non-Unc (*unc-54::Tc1* revertants) was established and designated TW404 for use in subsequent mapping crosses.

The relative position of *mut-2(r459)* on LGI was determined with respect to convenient N2 genetic markers, *dpy-14, sem-4* and *unc-13* in the following series of crosses . TW404 L4 hermaphrodites were mated to MT3184 (Egl Unc) males (P0). F1 wild type progeny were picked individually and allowed to self-fertilize. Dpy F2 progeny were isolated, self-fertilized and screened for Dpy Unc and Dpy Egl Unc recombinants. Lines of each class were established and screened for the presence of the *mut-2* mutator as described above. In addition to refining the map position of *mut-2*, the recombinants were used to build mutator strains that lack either the *dpy-5* genetic marker: TW409 (*mut-2 sem-4 unc-13*), or that lack both *dpy-5* and *unc-13*: TW410 (*mut-2 sem-4*). These strains were used in subsequent assays to assess the ability of cloned fragments to rescue the temperature sensitive sterile phenotype of *mut-2* bearing strains.

DNA Isolation and Southern Analysis

Whole genomic DNA, cosmids, plasmids or PCR products were prepared using a variety of methods depending on the source of DNA and its ultimate use.

Nematode Genomic DNA preparation

Large scale preparations of nematode genomic DNA were isolated and purified for use in restriction digests, southern analyses and PCR-based screens by the following method. Worm cultures were initiated on 1 or 2 large NGM plates (150 mm X 15 mm) seeded with *E. coli* OP50. Nearly confluent cultures of worms were washed from the plates with 2-3 mls of M9 buffer and used to inoculate 10 - 20 mls of S-medium (0.1 M NaCl, 0.05 M $KHPO_4$ (pH 6), 5 mg/L cholesterol (5 mg/ml ethanol stock), 10 mM KCitrate (pH 6), 3 mM $CaCl_2$, 3 mM $MgSO_4$ and 10 mls/L trace metals (trace metals

stock: 5 mM Na_2EDTA, 2.5 mM $FeSO_4 \cdot 7H_2O$, 1 mM $MnCl_2 \cdot 4H_2O$, 1 mM $ZnSO_4 \cdot 7H_2O$, 0.1 mM $CuSO_4 \cdot 5H_2O$)) in 1 or 2 large petri dishes (150 mm X 15 mm). Worms grown in liquid culture were fed a 40 X stock of *E. coli* RR1, a bacterial strain that does not maintain plasmids therefore reducing opportunities for bacteria-derived plasmid contamination. Worms were grown at 20°C with constant mixing (genie mixer set at ~150 r.p.m.'s) until the bacterial food was nearly exhausted. Worms were harvested by pelleting the liquid cultures at 3,000 r.p.m. in an IEC clinical centrifuge for 10 seconds. The supernatant was discarded and the pellet of bacteria and worms was washed 3 times with 10 mls of cold milli-Q water. Intact worms were then purified away from bacterial and worm debris by floating on or spinning through a 35 % cold sucrose "cushion". The cushion was prepared by completely resuspending the washed worm/bacteria pellet in 7 mls of cold milli-Q water and adding 7 mls of 70 % sucrose. The suspension was thoroughly mixed by inversion and centrifuged at 3,000 r.p.m. for 3 minutes. The topmost layer of gradient purified worms was carefully removed and transferred to a fresh 15 ml tube. The worms were washed and spun three times in DNA disruption buffer (0.2 M NaCl, 0.1 M Tris-HCl (pH 8.5), 0.05 M EDTA, and 0.5% SDS) as described above. The final worm pellet was resuspended in 500 µl of DNA disruption buffer and stored at -20°C until ready for DNA purification.

To digest degradatory enzymes and other contaminating proteins away from the worm DNA, 10 µl of 10 mg/ml proteinase K was added to 500 µl of DNA in disruption buffer. The mix was incubated at 65°C for 30 minutes, spiked with another 5 µl of proteinase K and incubated an additional 30 minutes. The enzyme digested mixture was

extracted three times with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) and two times with an equal volume of chloroform:isoamyl alcohol (24:1). After each extraction, the aqueous layer and organic layer were partitioned by centrifugation in a microcentrifuge at 14,000 r.p.m.'s for 5 minutes at room temperature. The nucleic acids were precipitated from the final aqueous layer by addition of 1/10th volume of 3 M NaOAc and 2.5 volumes of ice-cold ethanol. The precipitation mixture was incubated on ice for at least 20 minutes and the nucleic acids were pelleted by centrifugation as above for 10 minutes at 4°C. The supernatant was removed, the pellet was rinsed in 500 µl of ice-cold 70 % ethanol and re-spun for 10 minutes at 4°C. After removal of the supernatant, the pellet was air-dried for 5 to 10 minutes and resuspended in 200 µl 1 X TE (10 mM Tris-HCl, 1 mM EDTA pH 8).

RNA was removed from the DNA preparation by enzymatic digestion at 37° C with two successive additions of 2 µl RNase A (5 mg/ml stock) at 30 minutes each. To remove enzyme and free ribonucleotides, the DNA mix was extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1), once with chloroform:isoamyl alcohol (24:1) and ethanol precipitated as described above. The final DNA pellet was resuspended in 100 µl of 1 X TE. Electrophoresis of 1-2 µl through a 1 % agarose gel in 1 X TAE buffer (Tris-HCl, Acetate, EDTA) was used to assess DNA integrity, concentration and the presence/absence of RNA. Diluted samples (1:100 in water) were used to determine the DNA concentration spectrophotometrically using UV absorbance at 260 nm.

Genomic DNA preparations from single worms

For use in PCR-based reactions, nematode DNA was isolated from individual worms by picking young adult hermaphrodite worms from NGM plates and resuspending in 30 µl of worm lysis buffer. One µl of proteinase K (10 µg/ml stock) was added and the mixture was flash-frozen in a dry ice-ethanol bath for 15 minutes to help break open the cuticle of the worm. After thawing at room temperature and following at least one hour incubation at 65°C, the proteinase K was heat inactivated by boiling for 10 minutes. The DNA preparation was then stored at -20°C until ready for use. For most PCR reactions, unless otherwise noted, 5 to 10 µl were used in a final reaction mix of 50 µl.

Growth and maintenance of cosmid and plasmid DNA:

Cosmids were obtained by requesting the desired stocks from Alan Coulson at the Sanger Center, Wellcome Trust Genome Center, Hinxton, Cambridge, England (e-mail: alan@sanger.ac.uk). The cosmids used in this study included T21G5, C48B6, T10B11, C30F12 and H06O01. Cosmids were streaked for isolation of single colonies on LB agar plates containing the appropriate antibiotic (T = kanamycin; C = ampicillin; H = chloramphenicol). The plasmid containing the dominant marker, *rol-6(su1006),* was obtained from the CGC stock center and cultured similarly. Liquid cultures inoculated with a single colony were harvested after 12 to 16 hours of shaking at 37°C. Cosmid or plasmid DNA was prepared according to the recommendations of a commercially available alkaline lysis method available as a kit from Qiagen (Qiagen MidiPrep #12144).

Southern Analysis

Whole genomic DNA from N2 and BO strains were digested to completion with a panel of restriction endonucleases including *Bam* HI, *Eco* RI, *Hinc* II, *Hind* III, *Kpn* I, *Pst* I, *Sac* I, *Xba* I, and *Xho* I. The digestions were separated by gel electrophoresis through a 15 cm gel of 1 % agarose in 1X TAE buffer overnight at 23 V. The fractionated DNA digests were prepared for transfer to a charged nylon support (Magna Nylon, MSI, Westboro, MA) by depurinating the gel in 0.25 M HCl for 10 minutes, denaturing in 1.5 M NaCl, 0.5 M NaOH for two 15 minute periods and neutralizing in 1.5 M NaCl, 1 M Tris-HCl pH 7.5 for 30 minutes. The gel was rinsed briefly in de-ionized water and soaked for 5 to 10 minutes in 10 X SSC (1.5 M NaCl, 0.15 M NaCitrate). Digested DNA was transferred to a nylon membrane by capillary action using 20 X SSC (3 M NaCl, 0.3 M NaCitrate). Upon completion of transfer, the membrane was marked appropriately with pencil (date and lane numbers), rinsed briefly in 6 X SSC (0.5 M NaCl, 0.15 M NaCitrate) to remove any agarose remnants and allowed to air dry. The transferred DNA was immobilized on the membrane by illuminating the DNA side with UV in a UV Crosslinker (Fisher Biotech 1000) set for optimal crosslinking (1200 $\mu J/cm^2$).

Hybridization was carried out using slight modifications to the method described by Church and Gilbert (Church and Gilbert, 1984). Nylon membranes were prepared for hybridization by incubating in pre-hybridization buffer (0.5 M $NaHPO_4$ (pH 7.4), 100 µg/ml sheared, boiled salmon sperm) for 3 to 4 hours at 65°C in a shaking water bath while radio-labeled probes were prepared. Cosmid probes were labeled to high specific

activity with ^{32}P-dCTP (NEN #BLU513H) using Stratagene's kit for Random Prime labeling (Stratagene #300392). Unincorporated radiolabel and free nucleotides were removed by column purification using Boehringer Mannheim's Quick Spin G-50 Sephadex Columns and recommended procedures (catalog #1273-965). Prior to adding to the hybridization solution (0.5 M $NaHPO_4$), probes were boiled for 5 to 10 minutes to denature the double stranded probes. The membranes were sealed in bags with hybridization solution plus approximately 1 X 10^6 counts per minute/ml hybridization solution and incubated overnight at 65°C. Excess, non-specifically bound probe was removed by washing the membrane twice in 2 X SSC, 0.1 % SDS for 15 minutes each at 65°C, twice in 0.2 X SSC, 0.1 % SDS for 15 minutes each and once in 0.1 X SSC, 0.1 % SDS at room temperature. The moist membrane was wrapped in saran wrap and exposed to Kodak medical X-ray film (Eastman Kodak Co., Rochester, NY) at -80°C. Films were developed after 5 to 7 days exposure.

Polymerase Chain Reaction (PCR)

Variations of the polymerase chain reaction were used for amplification and comparative analysis for strain specific sequence level polymorphisms (N2 vs Bergerac vs *mut-2* mutator strains) or amplification of fragments for sequencing.

Standard PCR reaction Genomic DNA prepared from liquid cultures or single worms was used as template in PCR reactions. A typical PCR reaction mix contained 0.5 µg/ml of genomic template or 5 to 10 µl of a 50 µl single worm DNA lysis preparation mixed with final concentrations of 2 µM each of the 5' and 3' primers, 2 µM dNTPs (0.5 µM

each), 10 mM Tris-HCl (pH8.8 or pH9.2), 1.5 mM $MgCl_2$, 25 mM KCl and 0.1-1.0 unit of Taq in a final volume of 50 µl. Salt concentrations were optimized using Stratagene's OptiPrime kit (Stratagene #200422). Amplification was performed after denaturing the DNA mix at 94°C for 2 minutes followed by 30 cycles of 94°C for 20 seconds, 54° to 60°C for 30 seconds and 72°C for 30 seconds to 1.5 minutes and followed by an extension of 10 minutes at 72°C.

PCR products were analyzed by gel electrophoresis on 1 % agarose gels and stained with ethidium bromide (0.5mg/ml). The desired bands were excised from the agarose gel, purified and concentrated using the Qiaquick Gel extraction kit (Qiagen #28704). Concentrations of aliquots were determined by agarose gel analysis or spectrophotometric analysis. Purified samples were sequenced by the UNH DNA Sequencing Facility.

Bioinformatics

Once the *mut-2* interval was genetically defined, I searched ACeDB (Waterston, 1997; http://www.wormsrv1.sanger.ac.uk/cgi-bin/web_ace/) for the relevant cosmids that spanned the interval. Using the sequence information and Genefinder's ORF predictions (Green and Hillier, 1998), I selected candidates to analyze further. Based on the available sequence information, I designed primers using DNAStar's (Madison, WI) PrimerSelect program. I used Amplify 2.5β (Madison, WI) to test primers in "virtual" PCR reactions in an attempt to optimize reactions and circumvent non-specific priming reactions. Selected primers were synthesized by Operon Technologies (http://www.operon.com) and used to amplify fragments from N2, BO and *mut-2* mutator strains. Cycle Sequencing

reactions were performed by the UNH DNA Sequencing facility on PCR fragments gel purified as described above . Sequence analyses including resolution of ambiguous bases, construction of contiguous reads and sequence alignments were performed using the DNAStar's LaserGene (version 1996-1998) suite of programs (Seqed, EditSeq, Seqman and MegAlign).

Transformation rescue

Both cosmid and plasmid DNA(pRF4 or pCes1943) were isolated as described above. Cosmid DNA (10μg) was linearized with restriction enzymes that produce blunt ends (*Sma*I or *Pst*I). Each reaction mix was purified by agarose gel electrophoresis through a 0.8 % gel in 1 X TAE (Tris Acetic Acid EDTA). Bands were excised with a clean razor blade and purified as above. Concentrations were determined by absorbance at 260nm of a 1:100 dilution or by running one or two μl aliquots on a 0.8% agarose gel and comparing the intensity of ethidium bromide fluorescence to that of a molecular weight marker of known concentration.

N2 genomic DNA was isolated and prepared as described above. Approximately 10 to 20 μg of DNA was digested with *Sma*I or *Sca*I to generate blunt ended cuts. The DNA mixture was diluted to a concentration of approximately 1.0 to 0.6 μg/μl.

Transgenic lines were constructed by standard germline transformation techniques (Kelly and Fire, 1998; Mello et al., 1991). Microinjections with mixtures of "simple" arrays contained 100 μg/ml of a plasmid containing the *rol-6* dominant marker (pRF4 or pCes1943) and 10 to 20 μg/ml of the purified cosmid. Microinjections with mixtures of

"complex" arrays contained approximately 100 μg/ml of restricted N2 DNA and 1 μg/ml each of the test cosmid and dominant marker. Transgenic lines were generated by injecting young adults (animals that had begun to lay eggs) of N2 or *mut-2(r459) sem-4 (n1378)/++*. Rescue of *mut-2(r459)* was determined as survival of transgenic animals for several generations at 25°C.

RNA interference (RNAi)

The expressed sequence tagged (EST) clone, yk363e3, was obtained from the publicly available EST database curated by the lab of Yuji Kohara (http://www.ddbj.nig.ac.jp/). This insert is a 2.27 cDNA derived from the sequence represented by the candidate ORF, H06O01.2. Antisense and sense RNA was prepared from this clone using T7 RNA polymerase of Novagen's Transcription kit (Novagen) and T3 RNA polymerase (Boehringer Mannheim).

The plasmid containing the cDNA was linearized in two separate 40 μl reactions; *NotI* and *ApaI* (Promega). The restriction reactions were incubated at 37°C overnight, heat denatured at 65°C for 15 minutes and ethanol precipitated by adding one tenth volume 3 M NaOAc and 2.5 volumes ethanol. The DNA pellet was rinsed in 70 % ethanol, air dried briefly and resuspended in 20 μl of 0.1 X $T_{10}E_1$. One μl was analyzed by gel electrophoresis.

One 100 μl transcription reaction contained *NotI* digested plasmid, T7 RNA polymerase, 1 mM rNTP's (0.25 each), 10 mM DTT in the manufacturer's 1X reaction buffer. The second transcription reaction contained the *ApaI* digested plasmid and T3

RNA polymerase. The reactions were incubated at 37°C for two hours and then treated with 5 µl of DNAseI at 37°C for twenty minutes. This incubation was followed by an organic extraction twice with equal volumes of phenol:chloroform:iso-amyl alcohol (25:24:1) and once with chloroform:iso-amyl alcohol (24:1) followed by ethanol precipitation at -20°C overnight. The transcripts was pelleted at 14,000 rpm, washed with 70 % ethanol, air dried very briefly and resuspended in RNAse-free sterile de-ionized water. One µl of each was analyzed by gel electrophoresis.

Roughly equivalent amounts (about 0.25 µg/µl for a total of 0.5 µg/µl) of the transcripts were mixed, de-salted with 100-200 µl of RNAse-free sterile de-ionized water and concentrated through a 0.22 micron spin column (Costar). Aliquots of this dsRNA mix was used to microinject both of the distal gonad arms of WT L4 or young adult hermaphrodites. Injected animals were allowed to recover overnight and then plated singly to fresh plates every 12 hours for 3 days. Dead eggs were scored as well as any viable larvae were scored for an increase in males.

While *mut-2* mutator strains were later determined to be resistant to RNAi, this method now forms the basis for an assay of the presence of *mut-2* in experiments conducted by Hiroaki Tabara (pers comm). Briefly, the assay involves injecting putative *mut-2* mutator strains with dsRNA prepared from *pos-1*. For most worms, RNAi with *pos-1* is lethal to embryos early in development and results in dead eggs. *mut-2* mutator strains and other RNAi resistant worms will produce approximately 100 % viable eggs when injected with this or other embryonic lethal derived dsRNAs (*pie-1*).

RESULTS

Overview of strategies for identifying *mut-2*

To understand how *mut-2* carries out its various roles in the germ line of *C. elegans*, I wanted to identify a molecular clone of the gene. With a clone in hand we could design and apply a wide range of approaches to investigate where, how and when the encoded protein functions to maintain genome integrity in the worm.

My strategy for identifying the *mut-2* clone exploited the well characterized genetic map of *C. elegans* and its completely cloned and sequenced genome. In the section that follows, I describe genetic and molecular techniques I used to map the mutant allele to a chromosomal position relative to known genetic markers and physical polymorphisms. Typically, the next step in such an effort in *C. elegans* is to demonstrate "rescue" of the mutant phenotype. This can be achieved by molecular genetic techniques involving the reintroduction of candidate genomic fragments containing the putative gene and assaying for restoration of wild type function. Fragments that cover the region of interest are readily available as inserts in cosmid or Yeast Artificial Chromosome (YAC) clones. Each of these can be tested for the ability to restore wild type function following its introduction into a mutant background (rescue). In turn, single open reading frames (ORFs) derived from these clones can be introduced and tested for rescue. Identification of molecular lesions associated with mutant alleles provides yet more compelling

evidence. The combination of single ORF rescue and associated molecular lesions constitutes the currently accepted standard for gene identification in *C. elegans*. My project follows this general strategy and at the same time serves to highlight many of the factors that can complicate an otherwise straightforward approach.

mut-2 is an elusive gene. The only known mutant allele of this gene, *r459*, causes enhanced germ line activity of several transposons (Mut phenotype) as well as increased X chromosome non-disjunction, resulting in a high frequency of male offspring (Him phenotype). Both phenotypes are semi-dominant making complementation analysis with potential alleles uninterpretable. Moreover, both phenotypes represent enhanced frequencies of events in populations. Hence many large populations must be screened to acquire data representing significant frequencies. I was able to use the Mut and Him phenotypes for mapping *mut-2* to a small genetic interval, but they proved too cumbersome for standard rescue approaches. To clone *mut-2* by rescue, I sought a recessive phenotype that clearly affected the morphology or viability of individual animals.

I characterized a temperature sensitive (ts) sterile phenotype of *mut-2(r459)* that was previously associated with its BO parent. I determined that this ts sterility was a consequence of *mut-2(r459)*. Rescue with a series of genetic duplications demonstrated that this phenotype was recessive. Subsequent injections into mutant strains with cosmid-containing arrays were initially unsuccessful but the challenges of microinjection were eventually met. The end result of the work described below was to demonstrate rescue of the ts sterile phenotype with a single cosmid, H06O01, after incorporating modifications

to the standard techniques. Attempts to repeat and confirm single cosmid rescue and rescue with single ORFs derived from this cosmid are in progress.

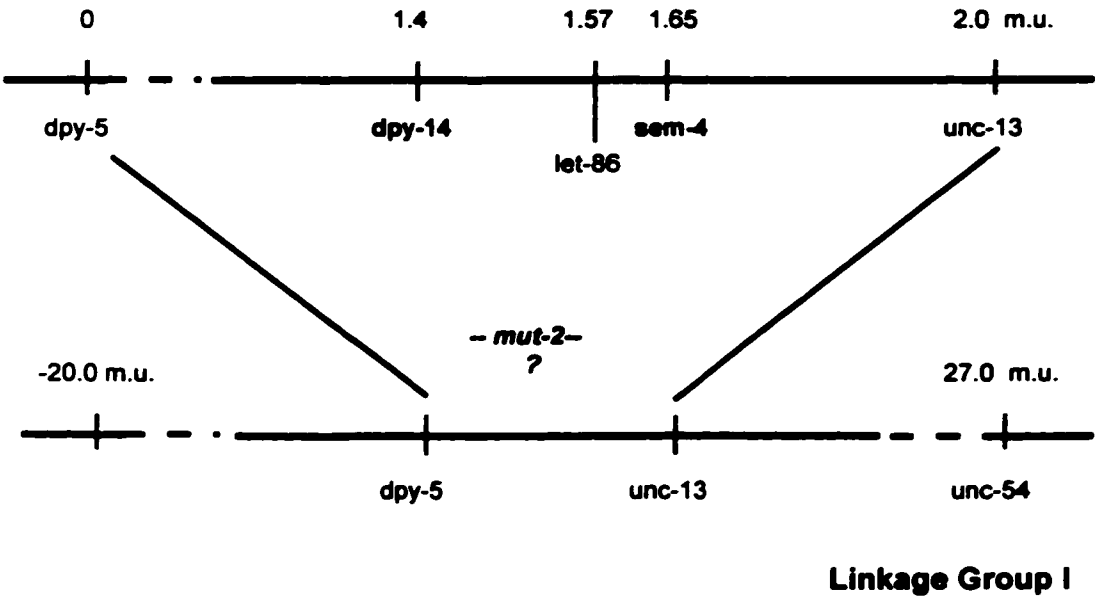
The characteristics of *mut-2* outlined above highlight a role for *mut-2* in chromosome segregation, chromatin behavior and gametogenesis. Evidence from other labs suggests additional roles in germ line gene silencing and sensitivity to gene inactivation techniques using interference by dsRNA, implying a role in RNA metabolism (Tabara et al., 1999). These characteristics are important clues to the normal function of *mut-2*. At the same time, they present significant technical hurdles for the identification strategies outlined above.

The first step: confirming the position of *mut-2* on LG I:

The first step to cloning the *mut*-2 gene is to find it. Preliminary mapping evidence linked *mut*-2 to Linkage Group I (LGI) to the right of *dpy-5* and near *unc-13* (Figure 4) (Finney and Horvitz, 1987). To confirm and refine this result I crossed a *dpy*-5 strain by a *mut-2 unc-54::Tc1* strain, picked Dpy Unc F2 progeny and asked if they were also Mut. Consistent with Finney's result, 3 of 59 Dpy Unc recombinants were Mut, placing *mut-2* to the right of and near *dpy-5*. In addition to confirming its position on LG I, this cross generated a strain with a visible marker (*dpy-5)* closely linked to *mut-2*. This provided an easy phenotype (dumpy, short, fat worms) for following *mut-2* through subsequent mapping experiments.

The genetic assay I used to detect the presence of the mutator allele, *mut-2(r459)*, involved two steps. First, I screened for the Him phenotype associated with this allele and then for the Mut phenotype. Using the above cross as an example, the Him

Figure 4. The *dpy-5 unc-13* region of Linkage Group I (LG I).



Positions of relevant genetic markers used for subsequent mapping crosses are shown. Positions are given in map units (mu.). (Adapted from Finney and Horvitz, 1987)

phenotype of each independent Dpy Unc recombinant line was determined by establishing several clonal populations of Dpy Unc animals and counting males. Populations with a frequency of males greater than 0.4 % were considered to be "Him" (compared to frequencies $\leq$ 0.1% males that occur in natural populations) and indicative of the *mut-2* mutator.

Each Dpy Unc line was tested for the presence of *mut-2* using the "Gold Standard" for *mut-2* activity: germ line activation of transposons in a reporter strain. The reporter strains I used are paralyzed due to transposon insertions (Tc4 or Tc5) into the *unc*-22 coding region. *unc*-22 codes for a very large muscle protein known as "twitchin" (Benian et al., 1989). Transposon insertions disrupt the normal protein and render it inactive resulting in worms that exhibit a characteristic twitching. Excision of the transposon can result in "reversion" and restoration of normal movement.

Each Dpy Unc recombinant was crossed with two reporter strains: *unc-22(r764)*::Tc4 and *unc-22(r644)*::Tc5. Subsequent populations derived from the progeny were screened for reversion of the twitchers to non-twitchers indicating excision of the transposon. The reversion frequency was $\geq 1 \times 10^{-3}$ for 18/19 independent Dpy Mut Unc lines (*dpy-5 mut-2; unc-22::Tc4*) from the Tc4 reporter cross. The reversion frequency was also $\geq 1 \times 10^{-3}$ for 22/23 lines of the Tc5 reporter cross (*dpy-5 mut-2; unc-22::Tc5*). These frequencies indicate that this strain: *dpy-5(e61) mut-2(r459)* is Mut. In nature, spontaneous reversion occurs at frequencies of $<1 \times 10^{-6}$. In the presence of the *mut-2* mutator, reversion occurs at a frequency of $\geq 1 \times 10^{-4}$ (Collins et al., 1987). Furthermore,

both Tc4 and Tc5 are only active in a *mut-2* background, therefore observation of excision events constitutes a sensitive assay for the presence of *mut-2* mutator.

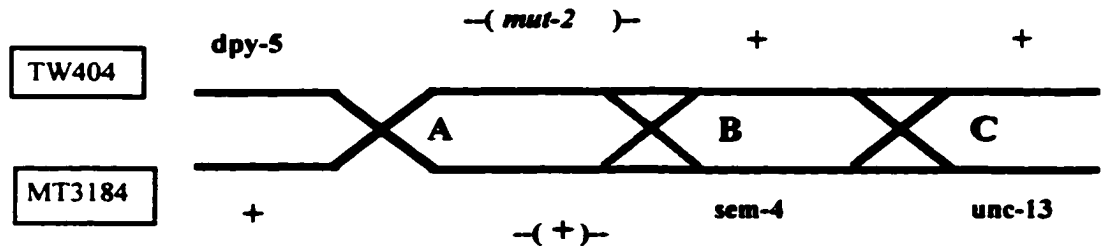
Interval Mapping - Refining the position of *mut-2*

To map *mut-2* with higher resolution, I employed an interval mapping approach that relied on scoring for Him and reversion rates among clonal populations. This approach did not depend on precise recombination frequencies but simply asked, "What is the frequency of males among recombinants? What is the frequency of reversion among recombinants? I selected convenient genetic markers in the *dpy-5 unc-13* region of LG I and mapped *mut-2* relative to each (Figure 4). I initiated this mapping strategy with a *mut-2* strain designated TW404 [genotype: *dpy-5(e61) mut-2(r459)*]. In this background the *mut-2* mutator allele is linked to the visible marker *dpy-5(e61).*

First, I mapped *mut-2* relative to the gene *sem-4* (Table 3). To do this, I crossed TW404 to MT3184 (*sem-4 unc-13*). MT3184 worms don't lay eggs due to a sex muscle defect (Sem = Egl), and they are paralyzed, kinky and uncoordinated, due to a muscle mutation (Unc). I identified independent lines of Dpy Egl Unc and Dpy non-Egl Unc recombinants representing cross-over events between the parental chromosomes. Using the Him criteria and the *unc-22::*TcX reversion assay described above, I determined that two of 20 independent Dpy Egl Unc lines were Him and Mut. These results placed *mut-2* to the left of *sem-4.*

To apply the transposition assay, the paralytic Unc gene needed to be crossed out because it is incompatible with the reversion assay of twitchers to non-twitchers. The resulting Dpy Egl animals were crossed to the reporter strains (genotypes: *unc*-22::TcX)

Table 3. Cross of TW404 (*dpy-5 mut-2)* X MT3184 (*sem-4 unc-13)* to position *mut-2(r459).*



Line	Cross over Class	Phenotype	% Males	Total Males
3A	C	Dpy Mut Unc	0.4	4/1071
4B	C	"	0.7	8/1079
Q	C	"	0.4	5/1173
1I	B	Dpy Mut Egl Unc	0.5	7/1316
8C	B	"	0.5	6/1092
1D	A	Dpy Egl Unc	<0.06	0/1549
1B	A	"	0.09	5/5322
QF	A	"	<0.07	0/1419
2A	A	"	0.04	1/2409
6B	A	"	<0.01	0/973
B	A	"	0.09	2/2096
I	A	"	<0.06	0/1609
J	A	"	<0.07	0/1449
S	A	"	0.03	1/3324
T	A	"	<0.06	0/1813
36	A	"	<0.05	0/2100
41	A	"	<0.04	0/2732
45	A	"	<0.04	0/1916
46	A	"	<0.04	0/2658
47	A	"	<0.04	0/2492
65	A	"	0.06	2/3424
78	A	"	<0.05	0/2096

and Dpy twitchers were picked. Populations were screened for reversion to non-twitcher phenotypes and as expected, the two strains that were Him demonstrated a high frequency of reversion ($\geq 1 \times 10^{-3}$) while no revertants were observed in the non-Him strains.

Recovery of two independent Dpy Egl Unc lines placed *mut-2* to the left of and closest to *sem*-4, within a 1.65 map unit (mu.) interval of *dpy-5*. If *mut-2* mapped to the right of *sem-4*, a cross over between the *dpy-5* and *sem-4* of the parental chromosomes would exclude *mut*-2. However, if *mut-2* mapped between *dpy-5* and *sem*-4, then a percentage of recombinants would exhibit the Him phenotype as I observed.

To define the left border of the *mut-2* interval, I next mapped it relative to *dpy-14* at 1.4 mu. To work with *dpy-14*, *dpy-5* was crossed out to obtain the *mut-2* strain TW409 (genotype: *mut-2 sem-4 unc-13*). TW409 was then crossed with CB188 (genotype: *dpy-14*) and Dpy Egl Unc recombinants were picked and scored for Him and Mut. The results demonstrate that *mut-2* maps to the right of *dpy-14* and hence between *dpy-14* and *sem-4* (Table 4).

From three separate mapping crosses, viable Dpy Egl Unc populations were established and scored for the frequency of males. Egl status was confirmed by a test cross described below. The viability of Dpy Egl Unc mutants is very poor therefore only the triple mutant reaching a total minimum population of 200 worms was scored for Him. By these criteria, 19/89 (21%) of the Dpy Egl Unc independent recombinants recovered were scored for Him.

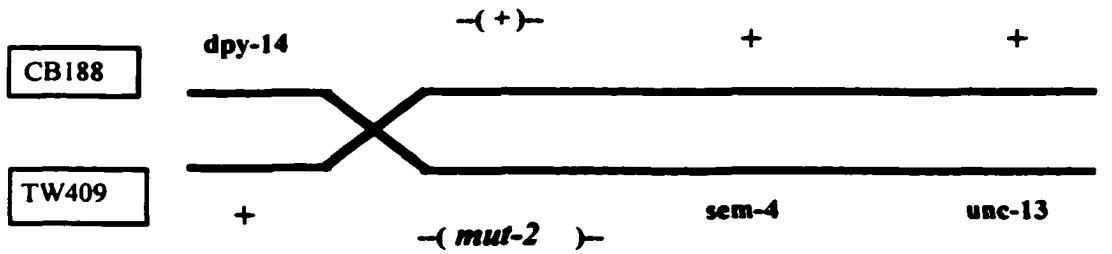
Of the lines assayed for males, several demonstrated frequencies greater than 0.4%. However, rather than a clear difference between Him and non-Him strains, I

observed a range of frequencies. If *mut*-2 mapped to the left of *dpy-14*, none of the recombinants would have been Him. Because I observed several strains with frequencies of males $\geq$ 0.4%, this mapping cross clearly placed *mut*-2 to the right of *dpy*-14, within a 0.25 map unit interval (Figure 4).

The previous mapping cross (TW404 X MT3184) established that *mut-2* maps to the left of *sem-4*, therefore only Dpy Egl Unc recombinants (from TW409 X CB188) were selected for culturing and scoring. Because *dpy-14* animals will sometimes exhibit an Egl phenotype, I devised a cross between potential Dpy Egl Unc recombinant and TW410 (*mut-2 sem-4/++*) heterozygous males to verify the Egl status of each line. If the recombinant was not Egl, then all the F1 progeny of a test cross would appear wild type and lay eggs normally. These recombinants were discarded. If the recombinant was Dpy Egl Unc, then half of the F1 progeny from the test cross would result in bags of worms (Egl) and the lines from which these were derived were retained for scoring the Him phenotype (Figure 5).

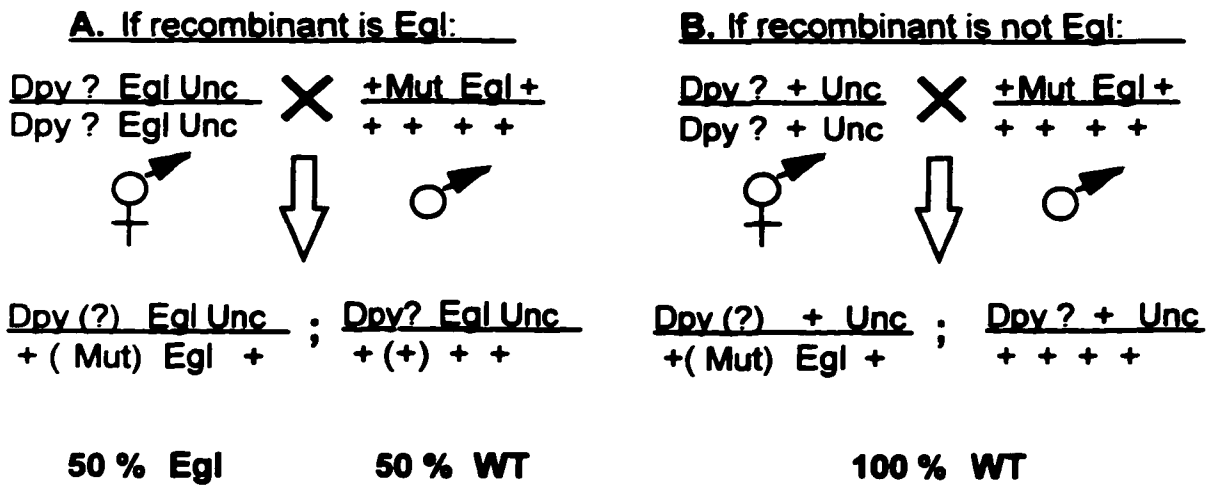
The viability of the desired recombinant class (Dpy Egl Unc) was very poor. As a result, the *unc-22*::TcX reporter cross was difficult to interpret and impractical. However, because the Him phenotype faithfully represents the status of *mut-2* mutator strains, this frequency was used to classify the recombinants. If *mut-2* mapped to the left of *dpy*-14, a cross over between the genetic markers of the parental chromosomes would

Table 4. Frequency of males for independent Dpy Egl lines derived from the mapping cross below.



Independent Line	Phenotype	% males	Total Males
52	Dpy Egl	1.8	4/226
89	"	1.5	6/395
66	"	1	2/200
51	"	0.9	5/539
26	"	0.37	8/2168
21	"	0.33	15/4553
5	"	0.32	5/1551
30	"	0.3	5/1494
38	"	0.3	3/984
44	"	0.29	6/2056
63	"	0.26	4/1500
33	"	0.19	11/5561
47	"	0.19	7/3522
19	"	0.17	11/6288
13	"	0.15	1/659
71	"	0.14	1/700
2	"	0.10	3/2883
11	"	0.09	6/6198
22	"	0.06	1/6184

Figure 5. Test cross devised to verify status of Sem of Dpy Unc recombinants.



The phenotype of *dpy-14* alone sometimes appears to be Dpy and Egl, therefore to verify the presence of an Egl phenotype associated with *sem-4*, I crossed the F3 Dpy Unc recombinant progeny of CB188 X TW409 by males that are heterozygous for *mut-2 sem-4*. If 50% of the resulting progeny are Egl then the genotype of the Dpy Unc recombinant hermaphrodite must have been *dpy-14 (mut-2?) sem-4 unc-13*.

exclude *mut*-2 and none of the recombinants would be Him. However, if *mut-2* resides within the interval between *dpy-14* and *sem*-4, then a percentage of recombinants would exhibit the Him phenotype. The results placed *mut-2* between *dpy-14* and *sem-4.*

Finally, comparison of the frequencies of males for recombinants derived from TW404 X MT3184 (Table 3) and those derived from TW409 X CB188 (Table 4) reveals a lack of distinction between the Him and non-Him class, the reason for which is not understood. It may be a direct result of the genetic background and its influence on expression of the Him and Mut phenotypes. The viability of the triple mutant described was very poor and it was not possible to generate consistently large populations for scoring purposes. While this illustrates the difficulty of using phenotypes associated with populations, it still allowed me to place *mut-2* in a small interval between two known phenotypic markers.

Physical Mapping - The *mut-2* interval has complete cosmid coverage

Once I had defined the genetic interval, I wanted to correlate it to the physical map and search the sequence for potential *mut-2* candidates. The complete genetic sequence for the *dpy-14 sem-4* interval became publicly available in June 1998 through the efforts of the *C. elegans* Sequencing Consortium (1998). By searching the database (ACeDB) I determined that *dpy*-14, defining the left border, is rescued by the cosmid, T21G5. The gene encoding *sem-4* at the right end of the interval is positioned at the end of cosmid VF15C11 and has been cloned. H06O01 is the cosmid just to the left of the *sem-4* "border" on VF15C11. With this information I established that the *dpy-14 sem-4* interval is spanned by 5 overlapping cosmids: T21G5 (31 kb), C48B6 (42 kb), T10B11

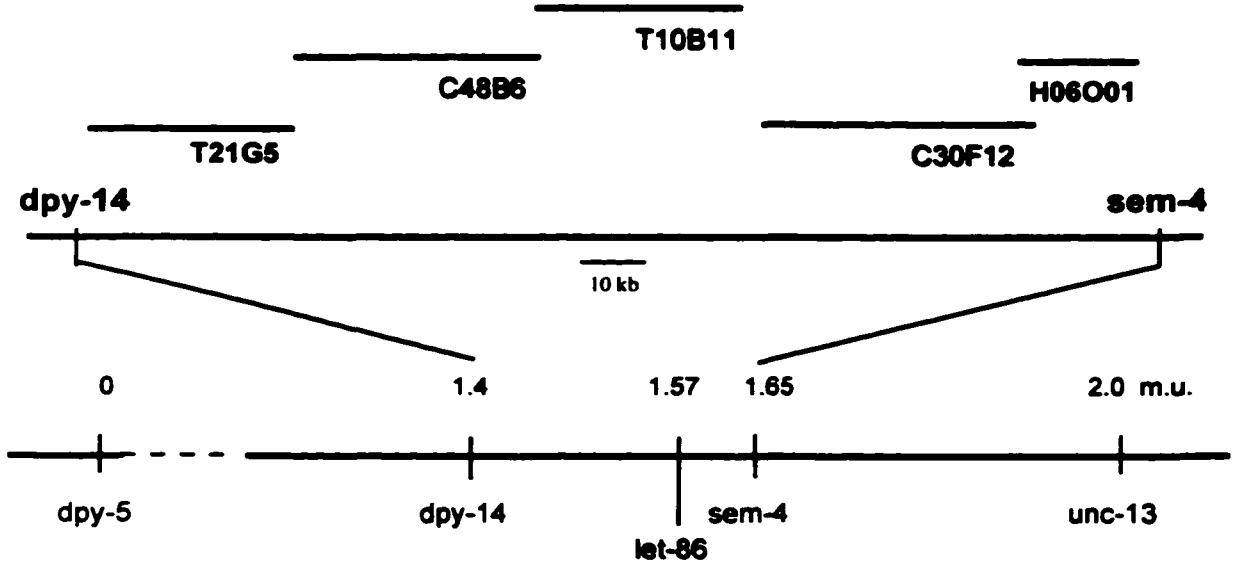
(36.5 kb), C30F12 (45 kb) and H06O01 (18 kb) and measures approximately 160 kb of contiguous sequence (Figure 6).

According to the analytical software tool, GENEFINDER (Green and Hillier, 1998) , there are 33 predicted open reading frames (ORFs) within this interval (Table 5). Three genes were previously characterized genetically and subsequently mapped and sequenced: *glh-1* (Roussell and Bennett, 1993), *smg-1* (Hodgkin et al., 1989) and isocitrate dehydrogenase (IDH - C30F12.7; ACeDB). About fourteen predicted ORFs have significant homology to genes represented in the current databases. The remainder are ORFs with weak or no homology to known genes in the current databases and represent novel, "pioneer" genes.

By carefully reviewing the available information and comparing it to what I knew of the characteristics of *mut-2 (*Him and germ line activation of transposons), I identified several ORFs as good candidates to test for rescue. Features that I looked for included motifs indicative of nucleic acid (DNA/RNA):protein interactions (i.e. KH or DNA binding domains), helicase activity/transcriptional modulators (i.e. SWI/SNF superfamily members), or chromosome recombination, repair or segregation activity (i.e. RAD/REC/MEI genes). Several ORFs fitting these different categories with varying degrees of homology (weak to very similar) were spread throughout the interval.

Following the typical approach to clone candidate genes in *C. elegans*, my next step would have been to test each of the five cosmids for the ability to rescue the *mut-2* mutant by DNA transformation methods. Once a rescuing cosmid was identified, smaller fragments could be used to identify the rescuing ORF. However, the *mut-2* phenotypes

Figure 6. The *mut-2* interval has complete cosmid coverage.



The *dpy-14 sem-4* interval is spanned by 5 overlapping cosmids: T21G5 (31 kb), C48B6 (42 kb), T10B11 (36.5 kb), C30F12 (45 kb) and H06O01 (18 kb) and measures approximately 160 kb of contiguous sequence.

Table 5. Predicted Open Reading Frames of the *dpy-14* to *sem-4* interval and their homologies to sequences from current databases*.

cosmid	ORF	identity/similarity/homology *
T21G5	3	*glh-1* ;RNA helicase
	2	unknown
	4	"
	1	tyrosine-protein kinase
	5	unknown
C48B6	4	"
	5	"
	6	*smg-1*; kinase
	7	unknown
	3	"
	2	ribosomal protein
	8	unknown
	9	"
T10B11	g	"
	3	RNA binding motif
	2	unknown
	4	"
	5	"
	6	"
	7	"
	1	"
	8	"
	9	"
C30F12	1	"
	2	SL1-transpliced
	3	downstream operon w/ .4
	4	unknown
	7	isocitrate dehydrogenase
	6	seven transmembrane receptor
	5	unknown
H06O01	4	protein tyrosine phosphatase
	3	retinal binding-like protein
	2	chromodomain-helicase-DNA binding protein
	1	protein disulfide isomerase ER 60 like protein

*from http:://www.sanger.ac.uk/projects/C_elegans/webace_front_end.shtml (8/99)

do not lend themselves to standard rescue approaches; they are characteristics of populations and not individuals. Therefore I sought to map *mut-2* more precisely to further reduce the list of candidate cosmids and ORFs. Lacking genetic markers, I looked for physical markers in the region.

Search for Sequence Level Polymorphisms

I sought to identify sequence level polymorphisms between the wild type strains, N2 and BO, within the 160 kb *mut-2* interval. The rationale for pursuing this effort was based on the following facts related to the ancestry of *mut-2(r459)*. BO and N2, two geographically distinct isolates, are polymorphic at the level of their genomic sequence and are known to be polymorphic for the number and distribution of different transposons. *mut-2(r459)* was derived from a mutagenized recombinant inbred BO strain (TR445: *unc-54(*r323::Tc1)) that is polymorphic at several loci on the right arm of LGI including sP1(aka pPK592), pPK5034, and hP9. These polymorphisms are found in BO but not N2. The mapping crosses described above used N2-derived genetic markers so that with each recombination event, BO sequence on LG I was replaced with N2 sequence.

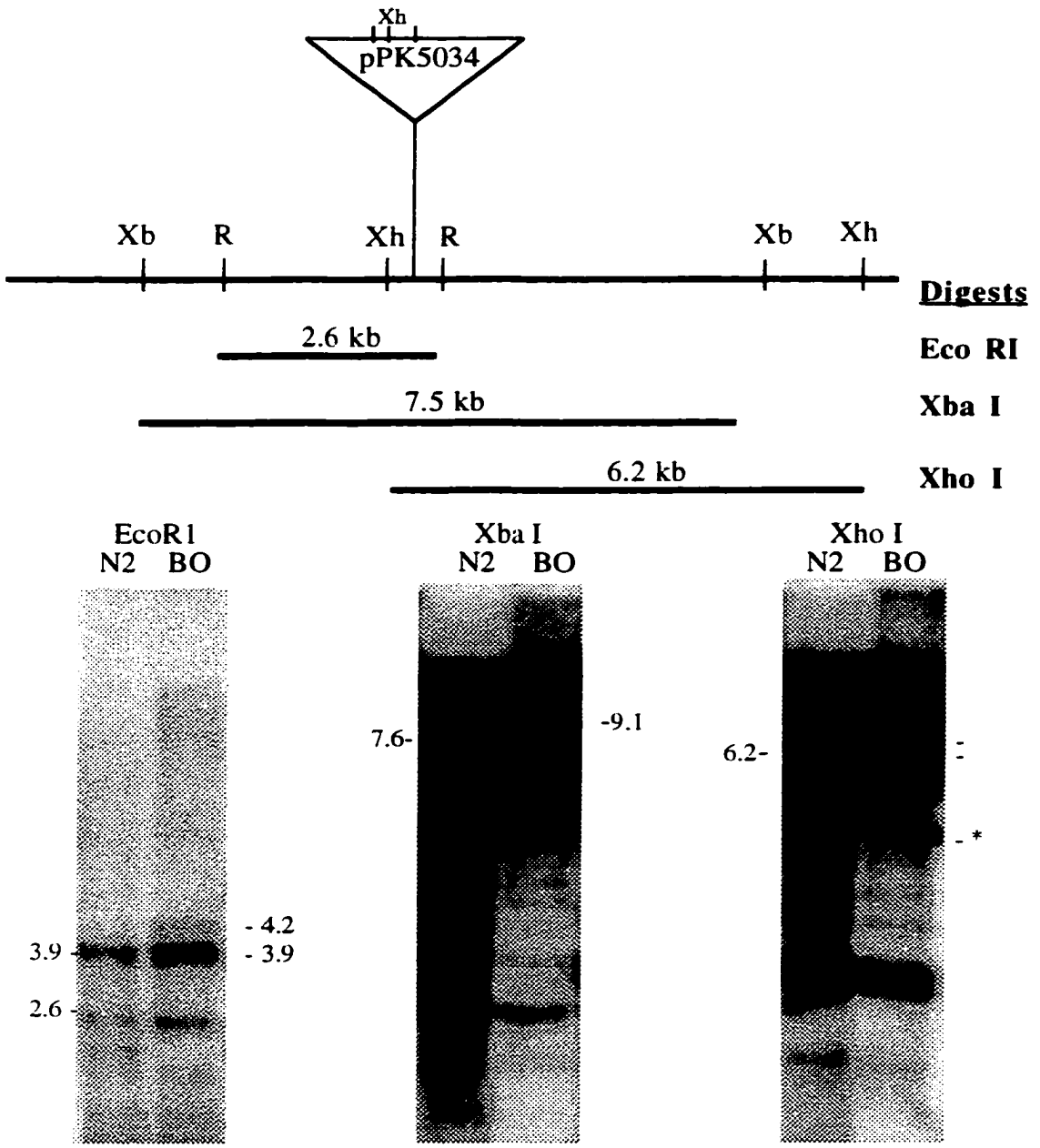
The strategy for identifying transposon-induced polymorphisms proved to be productive. Tc1 polymorphisms reportedly occur between N2 and BO at about 1 per 50 kb to 250 kb from any given gene (Rose and Nelson, 1982; Ruvkun, 1989; Korswagen, 1996). One Tc1 polymorphism had already been characterized at the left most end of the interval (pPK592), therefore I expected to find an additional one or two Tc1 polymorphisms within the *mut-2* interval. Concurrent with the efforts of the Genome

Sequencing Center, I was able to confirm one other Tc1 within this interval (http:://www.sanger.ac.uk/projects/C_elegans/webace_front_end.shtml (8/99)).

To look for transposon-induced polymorphisms, I used Southern blot analyses of genomic DNA from N2 and BO. I probed the DNA, digested with a panel of restriction enzymes (including *Eco*RI, *Xba*I and *Xho*I), using each of the 5 random prime-labeled cosmids. These cosmids represent N2 genomic sequence that span the *mut-2* interval. Band shifts in BO relative to N2 represent the insertion of a transposon. By comparing the observed restriction patterns to the expected restriction patterns (based on the available sequence information for N2), I was able to confirm a recently identified Tc1 polymorphism (pPK5034) (Figure 7). pPK5034 is inserted at base position 4661 of the T10B11 cosmid which is in the central region of the *mut-2* interval. About 60 kb to the left, a previously characterized Tc1 polymorphism, pPK592, occupies base position 12044 of the cosmid, T21G5.

pPK5034 has internal *Xho*I (3 sites @ 503, 555 and 713 bp), but no sites for *Eco*RI or *Xba*I. By probing with radio-labeled T10B11, the banding pattern reveals the polymorphism as a 1.6 kb difference between N2 and BO; a 2.6 kb N2 *Eco*RI band vs a 4.2 kb in BO and a 7.6 kb fragment *Xba*I N2 band vs an 9.1 kb in BO. The *Xho*I banding pattern is complex with respect to the 6.2 kb band observed in N2 and reveals a doublet. This occurs because the element can insert in either orientation at the site of insertion. When BO DNA is digested with *Xho*I, several fragments are generated as compared to the 6.2 kb fragment. Their size will depend on the orientation of the element at the site.

Figure 7. Southern Blot Analyses of N2 and BO genomic DNA.



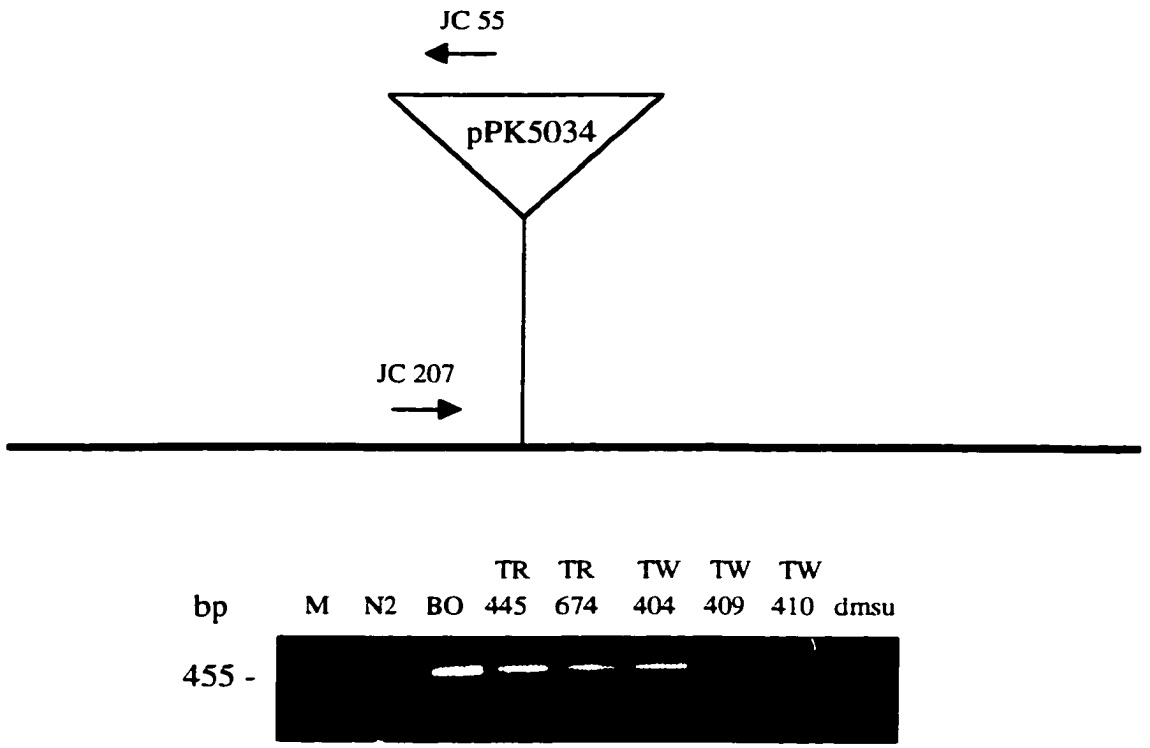
Genomic DNA digested with *Eco*RI, *Xba*I or *Xho*I was probed with random-primed ^{32}P labeled T10B11. Band shifts are the result pPK5034, a 1.6 kb Tc1 element.

The band marked with (*) may represent the remains of the insertion site after an imprecise excision resulting in a deletion (Figure 7).

These Tc1 polymorphic sites would be useful for screening recombinants derived from crosses between N2 and BO strains. Therefore I designed a PCR-based strategy to screen for the Tc1 insertions. Using the available sequence information, I designed genome specific primers for use with Tc1 primers to amplify across junctions at insertion sites for pPK592 and pPK5034. The primer pairs would then be used to screen *mut-2* ancestors and recombinants derived from the mapping crosses described above. In this manner, I would be able to track recombination events and replacement of BO sequence (Tc1 present) with N2 DNA (Tc1 absent).

An example of this PCR assay is shown in Figure 8. Using a Tc1 specific primer (JC55: 5'-GTGGGTATTCCTTGTTCGAAGCCA-3') and genome specific primers (JC207: 5'-CGCGCCATTGTCCTACTCGT-3'; JC208: 5'-ACCGGTTCTTTTGATTCGATGGTT-3') that flanked the predicted insertion site, I amplified a 455 bp fragment representing the junction of the Tc1 IR and flanking sequence from a reaction that contained JC55-JC207 (Figure 8). No product was amplified in the absence of an insertion. N2 is the control for an unoccupied site and BO is the control for the occupied site. Amplification with the primer pair that flanks the unoccupied site (JC207 and JC208) produces a 539 bp fragment and a 2.1 kbp fragment from an occupied site (data not shown).

Figure 8. PCR amplified junction of a Tc1 polymorphic site.



Sequence specific primers amplify across the a Tc1 junction to yield a 455 bp product in strains that are polymorphic for pPK5034. Strains lacking the Tc1 transposon will not amplify. M – molecular weight marker (565 bp): Controls: N2 (unoccupied site) and BO (occupied site); TR445 – *mut-2*(+); TR674 – *mut-2(r459)*; TW404,TW409 and TW410 – *mut-2(r459* derivatives; and dmsu (Dpy Mut Sem Unc) – non-archived *mut-2(r459)* strain.

Unfortunately, the original recombinants of the mapping crosses did not survive cold storage and we could not test this PCR-based method to help define the interval. However, before I was able to generate additional recombinants to test, I made an important breakthrough with another more productive line of investigation: characterization of a temperature sensitive phenotype more amenable to standard rescue techniques.

mut-2(r459) confers sterility at 25°C

As described above, I had mapped *mut-2* to a physically defined 160 kb region. The next step was to identify a molecular clone. While the Him and Mut phenotypes were suitable for the mapping phase, they were not useful for cloning by rescue and I was unable to identify useful physical markers to refine the interval. Therefore I continued to search for phenotypes or traits more amenable to traditional rescue techniques.

The original *mut-2* strain, TR674 and its direct derivative TR679, are inviable at 25°C, a characteristic thought to reflect their BO ancestry since BO strains are known to be temperature sensitive (ts) (Wood et al., 1980). TR674 was isolated following mutagenesis of TR445, a recombinant inbred line. Therefore I tested TR445 to ask whether it survived at 25° C. The strain was viable and proliferated at 25°C. This was an important result because it indicated that the ts sterile phenotype of TR674 was not BO-derived.

To determine if the ts sterile phenotype co-segregated with *mut*-2, I tested the strains I had constructed from my mapping crosses (Table 6). The *mut-2* mutator bearing

descendants of TR674, including TR679, TW404 and TW410, were viable at 20°C but sterile at 25°C. The non-*mut*-2 bearing derivatives of TR679 and subsequent recombination crosses were viable 25°C. An additional piece of evidence that the ts phenotype is linked to the *mut-2* mutator is revealed by the mapping data. Construction of TW404 (*dpy-5 mut-2),* to position *mut-2* relative to *dpy-.5* and construction of TW410 (*mut-2 sem-4),* to position it relative to *sem*-4, forced crossovers on either side of *mut*-2, therefore the DNA flanking *dpy-14 sem-4* interval is derived from N2. This provides further support that the ts phenotype is conferred by *mut-2(r459).* More importantly, this provided me with an entry point for devising strategies to clone the *mut-2* gene.

Cursory analysis of *mut-2* animals raised at 25°C revealed that the sterile defect involved gametogenesis. The gonads appeared normal but exhibited a dearth of sperm with variable morphology (data not shown). Oocytes were normal in number and showed no gross abnormalities. DAPI staining revealed no obvious chromosomal defects. To determine whether the oocytes were also subtly affected, I tested their fertility. To do this, I cross fertilized *mut-2* hermaphrodites raised at 25°C with normal sperm from WT males and looked for progeny that were viable and fertile. These crosses failed to produce viable animals indicative of an additional defect in oogenesis. This evidence implicates *mut-2* in gametogenesis and further supports a more global role for *mut-2* in worm viability through a subtle effect on genome integrity.

Table 6. The temperature dependent phenotypes of *mut-2(+)* and *mut-2(r459)* strains at 25°C.

			Sterile ?	
Strain	**Genotype**	**Mut ?**	**25°C**	**20°C**
N2	+/+	No	No	No
BO	+/+	No	Yes	No
TR445	*unc-54(r323::Tc1)*	No	No	No
TR674	*mut-2(r459) unc-54(r323::Tc1)*	**Yes**	**Yes**	No
TW404	*dpy-5 mut-2(r459)*	**Yes**	**Yes**	No
TW410	*mut-2(r459) sem-4*	**Yes**	**Yes**	No

N2 and BO are the common wild type laboratory strains. BO is temperature sensitive and sterile at 25°C. TR445 is a recombinant inbred line of BO but is viable at 25°C. TR674 is the EMS mutagenized strain derived from TR445 and is also sterile at 25°C. TW404 and TW410 are recombinant strains derived from TR674 and the appropriate crosses with N2 strains bearing convenient phenotypic markers. These results demonstrate that the temperature sensitive sterile phenotype is not BO derived and that it co-segregates with the r459 allele of *mut-2*.

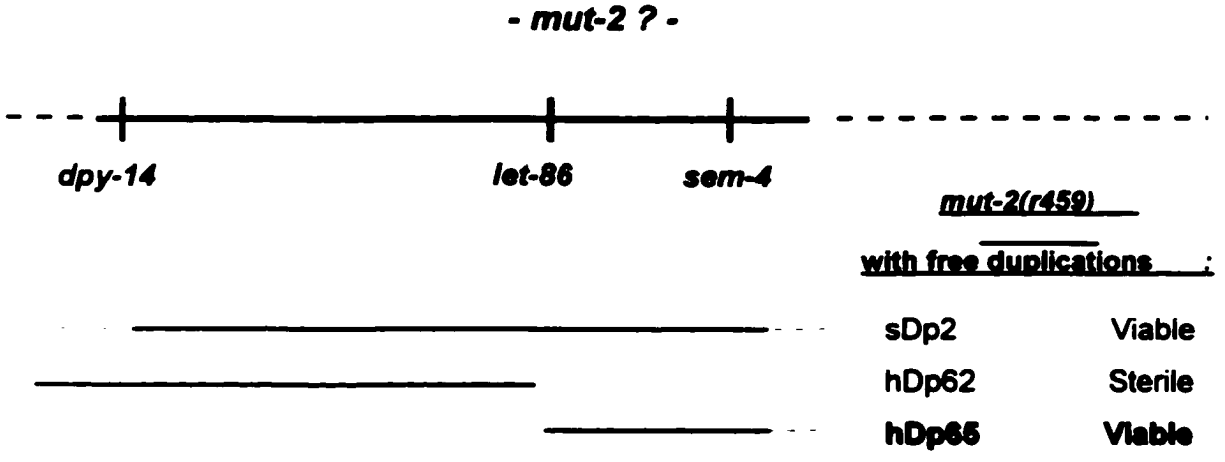
Rescue of ts sterility with genetic duplications

The ts sterile trait had potential value as a phenotype suitable for cloning strategies. The best known phenotypes, Mut and Him, were both semi-dominant so I needed to test whether the ts sterility was dominant (not useful) or recessive (suitable for cloning by rescue).

To do this I examined the ability to complement (rescue) this phenotype with a "free duplication". In *C. elegans*, strains are available that carry specific types of genetic rearrangements used to "balance" or maintain mutations as heterozygotes that are otherwise lethal or sterile. One type of genetic rearrangement includes chromosomal duplications. These large genetic fragments are products of radiation mutagenesis, translocations or rare chromosomal rearrangements. They are maintained as free extrachromosomal elements in addition to a full genomic complement. Duplications segregate in a non-Mendelian fashion and are occasionally lost during mitosis and meiosis (Edgley et al., 1995). The endpoint of duplications are defined by the genetic markers they do or do not rescue. Using standard genetic crosses, duplications can be moved into the desired mutant backgrounds and assayed for the ability to restore wild type function (rescue). I selected several duplications that spanned all or part of the *mut-2* interval and tested their ability to rescue the ts sterile phenotype.

sDp2 is a large free duplication of LG I that spans the entire *dpy-14 sem-4* interval (Figure 9). I crossed it into a *mut-2* background generating the strain TW417

Figure 9. The free duplications, sDp2 (TW417) and hDp65 (TW422), rescue the temperatures sensitive sterile phenotype of *mut-2*, but hDp62 (TW423) does not.



(genotype: *mut-2 sem*-4; *sDp2 (I;f)*) and asked whether it survived at 25°C. The duplication rescued the ts sterile phenotype and was completely viable at 25°C. As expected for free duplications, sDp2 was occasionally lost, allowing me to recover the original *mut-2* strain which I then re-tested at 25°C. In the absence of the duplication, the strain was again Egl and sterile at 25°C. This demonstrated that the ts sterile phenotype was recessive and that it would be a suitable phenotype for standard rescue techniques.

Next, I chose two smaller free duplications, hDp62 and hDp65, that bisect the *mut-2* region. I constructed the *mut-2* strains, TW422 (*mut-2 sem-4 unc-13; hDp65 (I;f)*) and TW423 (*dpy-14 mut-2 sem-4 unc-13; hDp62(I;f)*), carrying each of these genetic rearrangements. The *mut-2* strains into which the duplications were introduced were chosen according to the linked markers that would allow easiest recognition of animals that retained or lost the duplication. TW422 lays eggs and has normal movement but becomes Egl and Unc upon loss of the duplication. TW423 is Egl and Unc but becomes Dpy Egl Unc upon loss of its duplication.

To ask which of these duplications rescued the ts sterile phenotype, I tested each strain at 25°C. TW422 was viable at 25°C indicating rescue. TW423 was sterile at 25°C. This was an important result because it further supports the ability to rescue the ts phenotype with a defined genomic fragment. I also compared the viability of TW422 and TW423 to KR1755 and KR1758 in which the duplications are normally maintained. The parental strains carrying the duplications were viable at 25°C.

As before, animals that had lost the duplication were recovered from the rescued strain, (TW422) and re-tested at 25°C to ensure that rescue was conferred by the

duplication. Interestingly, I noted that the duplications were lost at a very low frequency when in the *mut-2* background ($\leq$ 5%) as compared to the stock strains (KR1755 and KR1758) where they was lost at a much higher frequency (~ 50%) (Table 7). Whether regulating the stability of duplications is another feature of *mut-2* is a very interesting question and remains to be investigated.

Rescue of the ts sterile phenotype by hDp65 represents a critical result because it provided a new means to reduce the *mut-2* interval. The endpoints of the two smaller duplications are defined genetically by the gene *let-86* which resides at 1.57 mu. within the *mut*-2 interval. hDp62 rescues *let-86* but hDp65 does not. By determining the physical location of *let-86,* I could define the endpoint of hDp62. More importantly, I would be able to shift focus from five cosmids to two or three at the right end of the interval.

Rescue of *let-86* defines the duplication endpoints and reduces the physical interval to two cosmids:

As mentioned, the duplication endpoints that lie within the *mut-2* interval are defined genetically; their precise physical locations are unknown. Therefore, I wanted to define the location of *let-86* to in turn define the physical endpoints of hDp62 and hDp65.

let-86 is a larval lethal gene of unknown cause that is rescued by hDp62

Table 7. **Duplications are lost at a reduced frequency in the *mut-2(r459)* background.**

Strain	Genotype	Phenotype with Dup'n	Phenotype without Dup'n	Duplication Loss
KR1755	*dpy-14 unc-13; hDp65*	Dpy	Dpy Unc	49%
KR1758	*dpy-5 dpy-14; hDp62*	WT	Dpy Dpy	48%
TW422	*mut-2 sem-4 unc-13; hDp65*	WT	Egl Unc	4.9%
TW423	*dpy-14 mut-2 sem-4 unc-13; hDp62*	Egl Unc	Dpy Egl Unc	4.4%

but not hDp65 (Figure 9). The lethal gene is linked to *dpy-5* and *unc-13* and maintained as a wild type heterozygote (BC244: genotype: *dpy-14 let-86+unc-13/ ++unc-15+)*. Self fertilization of BC244 yields WT heterozygotes (50%), Unc (25%) and Dpy Unc (25%) progeny. The Dpy Unc progeny will not mature beyond the L1 larval stage and die.

To rescue *let-86* and determine the relative positions of the duplications endpoints, I injected each of 3 different cosmids: T10B11, C30F12 or H06O01 along with the dominant marker, rol-6(su1006) into N2 to create strains transgenic for these cosmids. To see which cosmid would rescue *let-86,* I genetically introduced each into the mutant background by the appropriate cross and asked whether the relevant animals survived passed the L1 larval stage. The Dpy Uncs that survived would indicate that *let-86* (and the endpoint of hDp62) mapped to the region spanned by this cosmid.

Constructing and recognizing transgenic BC244 animals is complicated by the fact that Dpy and Unc each mask the roller phenotype. Therefore, to identify transgenic lines, progeny of the crosses could be scored for the presence of WT rollers and their sibling Dpy Uncs followed for viability. The crosses were performed as follows: heterozygous males (genotypes: *dpy-14 let-86+unc-13/++++* or *++unc-15+/++++)* were crossed with hermaphrodites from N2 transgenic strains carrying T10B11, C30F12 or H06O01 along with the dominant marker, rol-6*(su1006)* and the progeny examined as described above. I identified several independent lines segregating WT rollers (indicating the presence of arrays containing the marker and cosmid) and Dpy Uncs (*let-86)*. The Dpy Uncs were picked to separate plates and scored for viability.

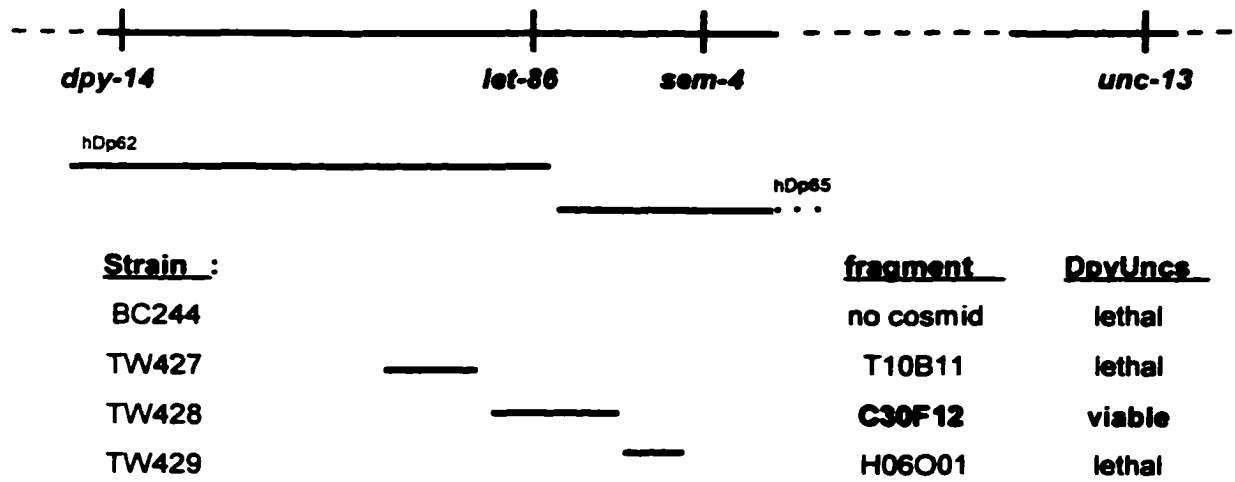
The Dpy Uncs transgenic for C30F12 matured beyond the L1 stage demonstrating rescue (Figure 10). Dpy Uncs transgenic for T10B11 and H06O01 died as L1 larvae. This defined *let-86*'s relative position within C30F12 among 7 predicted ORFs. Of these ORF's, five are novel, one has homology to a seven transmembrane receptor (C30F12.6) and another is isocitrate dehydrogenase (C30F12.7) (Table 5). Rescue with C30F12 also positions the right end of hDp62 within this cosmid. Therefore, hDp65, the duplication that rescued the *mut-2* ts sterile phenotype, must lie to the right with its endpoint in the same general region of C30F12 or H06O01 (Figure 10).

Cosmid Rescue of *mut-2*:

The rescue of *let-86* and the mapping results described above establish that *mut-2* is located between *let-86* and *sem-4*. This reduces the total number of cosmids to be tested for *mut-2* from five to two and the total number of ORFs to be screened from 33 to less than 11. Furthermore rescue of the ts sterile phenotype with sDp2 and hDp65 demonstrates that this phenotype is suitable for standard transformation rescue techniques.

To determine which of the two cosmids, C30F12 or H06O01, rescues the ts sterile phenotype, I microinjected each directly into *mut-2* strains and screened for viability at 25°C. However after recovering only one transgenic line (that was inviable at 25°) from injections of 200 or more individuals of TW410, I revisited my strategy. The following analysis provides an important and cautionary note when determining the expected success rate for microinjection strategies.

Figure 10. C30F12 rescues the *let-86* gene.



Strain :	fragment	DpyUncs
BC244	no cosmid	lethal
TW427	T10B11	lethal
TW428	C30F12	viable
TW429	H06O01	lethal

Rescue of *let-86* with the cosmid, C30F12, helps define the relative position of the endpoint of hDp62. hDp62 does not rescue *mut-2* but hDp65 does. Therefore *let-86* defines the left border of the *mut-2* interval.

Injecting directly into the mutant strain and demonstrating rescue would be ideal. However, successful injections rely on healthy recipient worms that generate large broods (200-300 progeny/hermaphrodite) of which only 1-10% will carry the transgene (Fire and Waterston, 1989; Mello and Fire, 1995). A small fraction of these F1 progeny will transmit the array at frequencies of 5-95% (Stinchcomb et al., 1985). It was clear that the combination of a low brood size and poor viability associated with *mut-2* strains worked against recovering transgenic lines.

In addition, the linked phenotypic markers for TW404 (Dpy Mut) and TW410 (Mut Egl) made these strains poor candidates for rescue by injection. TW404 is not a good candidate because the Dpy phenotype masks the roller phenotype. TW410 is not a good candidate because its brood is severely reduced (30-50 progeny/hermaphrodite).

Therefore, my revised strategy was to construct transgenic strains in an N2 background then introduce *mut-2* into the transgenic background by the appropriate genetic crosses. These crosses are complicated by the fact that the hermaphrodite must be the transgenic animal because male rollers mate poorly if at all, while hermaphrodite rollers can be cross-fertilized. While *mut-2* males occur in greater frequency, they are also sterile. Therefore *mut-2* must be introduced by crossing the transgenic hermaphrodites with heterozygous males (*mut-2 sem-4/++*). As a result, three generations are required before the homozygous *mut-2* transgenic line is established. This proved to introduce an additional complication, allowing sufficient time for transgene silencing to occur (see below).

Using the revised strategy described, two independent N2 lines transgenic for H06O01 plus the rol-6 marker were constructed by Kim Reese (Seydoux lab of Johns Hopkins School of Medicine) and were crossed with heterozygous TW410 males (*mut-2 sem-4/++*). Transgenic lines were established from progeny that rolled and never laid eggs (Egl). These transgenic lines were then tested at 25°C. Both lines were found to be sterile. This would suggest that *mut-2* does not reside in the region represented by H06O01. Alternatively, *mut-2(+)* is represented but expression from the array was insufficient for rescue.

To test the latter explanation, I considered evidence related to DNA transformation strategies and expression from transgenes. Injected DNA is assembled into extrachromosomal arrays from which expression of the marker and candidate gene can be driven. One caveat of this rescue method is that recipient germ lines of most strains exhibit the ability to silence gene expression from transgenic arrays by a mechanism that is not well understood and currently under intense investigation (Kelly et al., 1997). Evidence suggests that the degree of germ line transgene silencing depends on the sequence complexity of the microinjected arrays. Arrays containing many copies of a transgene, and therefore highly repetitive, are readily silenced. To overcome this hurdle, arrays of non-repetitive genomic DNA mixed with few copies of the transgene and marker are often used although progressive silencing after passage through many generations is still observed (Kelly and Fire, 1998).

My initial attempts at cosmid rescue of *mut-2* could have failed for several reasons:

(1) The cosmids were injected into N2 animals and assembled as simple, highly repetitive arrays containing many copies of the cosmid plus the dominant marker, (*rol-6[su1006]*). Because they are simple arrays, they are recognized by the host germ line as substrates for the silencing machinery.

(2) After transgenic lines were established, arrays were present for several generations before transgenic lines homozygous for *mut-2* were established and screened. For reasons that are unclear, many transgenes, whether introduced as simple or complex arrays, are progressively silenced over several generations (Kelly, 1998).

(3) Finally, it has been observed that many, but not all, germ line specific transgenes, in addition to many endogenous genes, are subjected to silencing, again for reasons (transgenes) or mechanisms (endogenous genes) that are unclear (Seydoux, 1996; Kelly, 1998) Given that *mut-2* plays multiple roles in the germ line, it seems likely that it is normally expressed in the germ line and hence, that *mut-2* bearing transgenic arrays would be subject to silencing.

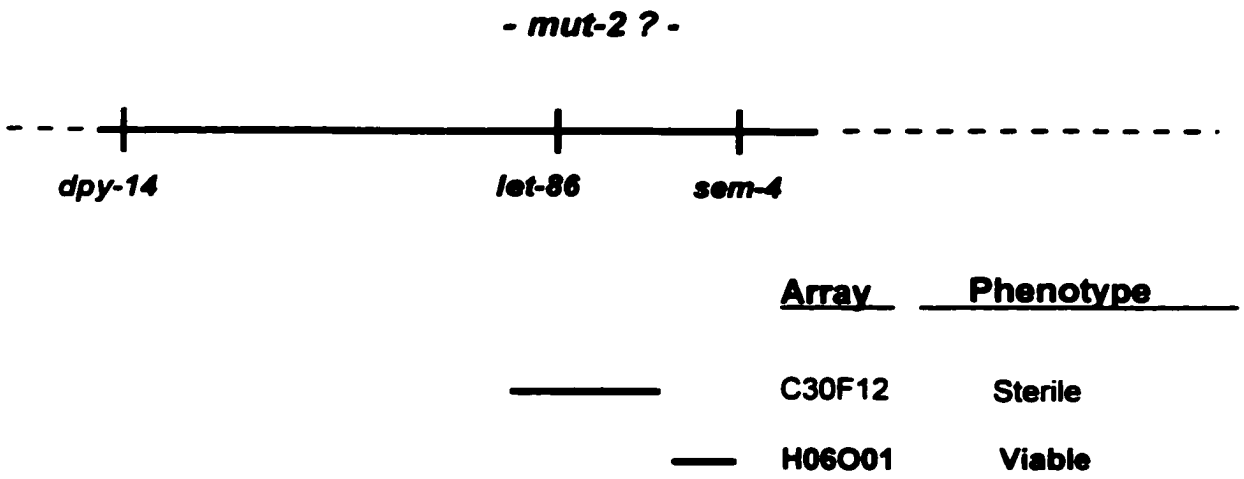
To test the possibility that rescue was not achieved because of the repetitive composition of the array, a complex mix of H06O01 plus *rol-6* marker (@1ng/µl each) diluted in fragmented genomic N2 DNA (@100ng/µl) was prepared and injected into N2. Transgenic lines segregating rollers were established. Arrays were then crossed into the *mut-2* background as described above. Independent lines were established and tested at 25°C. Unfortunately, all lines (including independent lines with H06O01-containing complex arrays and C30F12-containing complex arrays) were ts sterile indicating that

these arrays did not rescue *mut-2* function or that repetitive arrays alone were not causing transgene silencing.

To test the possibility that passage through several generations contributed to progressive silencing from the transgenic array, I injected the complex DNA mix directly into TW410 heterozygous hermaphrodites (*mut-2 sem-4/++*). These animals are relatively healthy and have a brood size similar to wild type (200-300/hermaphrodite). By picking the progeny of these injected heterozygotes, I bypassed the several passages required to recover homozygous *mut-2* roller progeny by the process described above. Animals that were transgenic (rollers) were picked immediately to 25°C and transmitting *mut-2* lines followed and scored for viability. Using this strategy, I isolated three independent lines transgenic for C30F12 and one transgenic line for H06O01. All three of the C30F12 transgenic *mut-2* lines were sterile at 25°C. The critically important result was that the transgenic H06O01 line was viable at 25°C (Figure 11).

To ensure that the ts rescue was conferred by the H06O01 transgenic array and not by some other rare event such as recombination, several rollers were "cured" of the array and re-tested for sterility at 25°C. To do this, I picked 10-20 animals from the 3rd or 4th generation of the rescued line at 25°C and established populations at 15°C. Animals that no longer segregated rollers after 2 to 3 generations at 15°C had lost the array and were considered to be "cured". Populations of cured animals were re-tested for viability at 25°C. The "cured" lines were sterile at 25°C indicating that viability was dependent on the presence of the H06O01 containing array.

Figure 11. H06O01 rescues the temperature sensitive sterile phenotype of *mut-2* when introduced as a complex array.



Three independent *mut-2 sem-4* roller lines carrying the C30F12 plus a *rol-6* marker in a complex array were tested at 25°C. All three were determined to be sterile. A single transgenic line injected with H06O01 plus a *rol-6* marker in a complex array was viable at 25°C.

These results strongly suggest that the *mut-2* gene resides in the region represented by H06O01. Experiments are on-going to repeat and confirm this result. Additional injections are underway using restriction fragments derived from this cosmid to identify the smallest rescuing fragment and to identify the candidate ORF by transformation rescue.

Candidate genes:

By rescuing the ts phenotype with H06O01, I have reduced the search for the *mut-2* gene to a relatively small cosmid of 18 kb that contains four predicted ORFs. These four ORFs all exhibit significant similarity to known genes in the current databases. H06O01.1 has 46% identity to the human protein disulfide isomerase ER 60, a protein involved in post-translational folding. H06O01.3 has 30% identity with a retinal-binding protein and appears to be a novel hydrophobic ligand binding protein. H06O01.4 has similarity to a Drosophila protein tyrosine phosphatase. H06O01.2 has 44% identity to the mouse chromodomain-helicase-DNA binding protein, CHD1. Given the characteristics of the *mut-2* mutator with respect to its effects on chromatin behavior, the CHD1-like gene seems to be the best candidate for the *mut-2* gene.

CHD1 is a member of the SNF/SWI superfamily of transcriptional regulators and belongs to a sub-family of proteins thought to exert control via chromatin re-modeling. CHD1 is unique among members of this superfamily because it combines three distinct and highly conserved motifs within its ORF. It has a chromodomain motif implicated in chromatin remodeling, the highly conserved helicase motif and invariant DEAH box associated with transcriptional regulation, and a variable but conserved domain associated

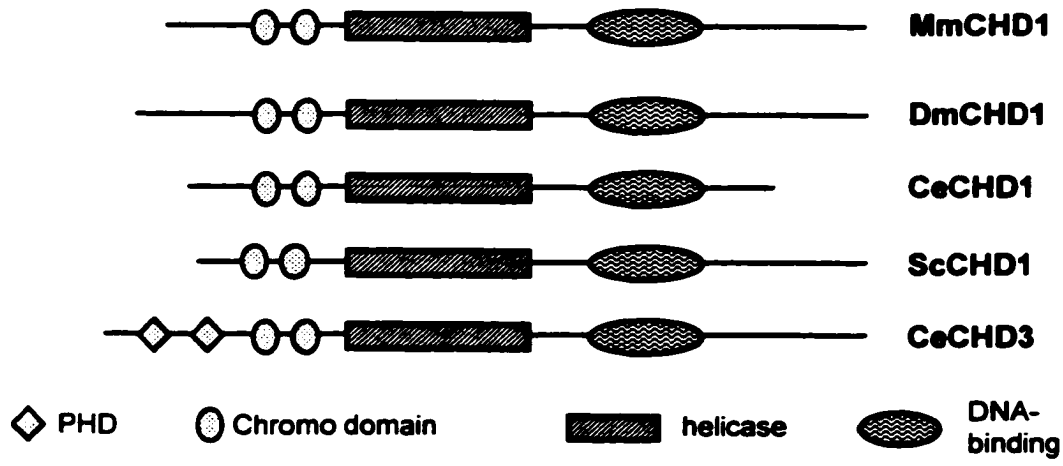
with DNA binding. There are several related homologues in mammals (MmCHD1/HsCHD1, HsCHD2, -3, and -4), three in the worm (H06O01.2 (CeCHD1), T14G8.7 (CeCHD2, and *cec*-1), two in *Drosophila* (DmCHD1 and -2) and a single gene each in fission and budding yeast (ScCHD1 and SpCHD1(not depicted)) (Figure 12; reviewed in Woodage et al, 1997). The normal function of this gene is not well understood.

Project Status

In summary, I localized *mut-2* to a small genetic interval employing standard genetic methods, recombination mapping and population assays for the *mut-2* mutator phenotypes: Him and enhanced reversion frequencies. I defined the *mut-2* interval as a 160 kb region between *dpy-14* and *sem-4* on LGI that is spanned by five cosmids. Lacking both convenient physical markers and phenotypes amenable to standard transformation rescue (the Him and Mut phenotypes), I identified a recessive temperature dependent phenotype associated with *mut-2(r459)* and determined that the ts phenotype was independent of its BO ancestry. I subsequently restored viability at 25°C by introducing hDp65, a small free duplication. Microinjection with C30F12 rescued *let-86* but not the ts phenotype, defining the left endpoint of hDp65. Microinjection of H06O01 rescued the *mut-2* mutator dependent ts phenotype . Of the four predicted ORF's, H060O1.2, which encodes a chromodomain-helicase-DNA binding-like gene, appears to be the most likely candidate for *mut-2*.

To determine whether H06O01.2 represents *mut-2*, it is desirable to identify a molecular lesion that is associated with *mut-2(r459)*. Therefore, I initiated an effort to

Figure 12. Summary of the Chromodomain-helicase-DNA binding domain (CHD) gene family members.

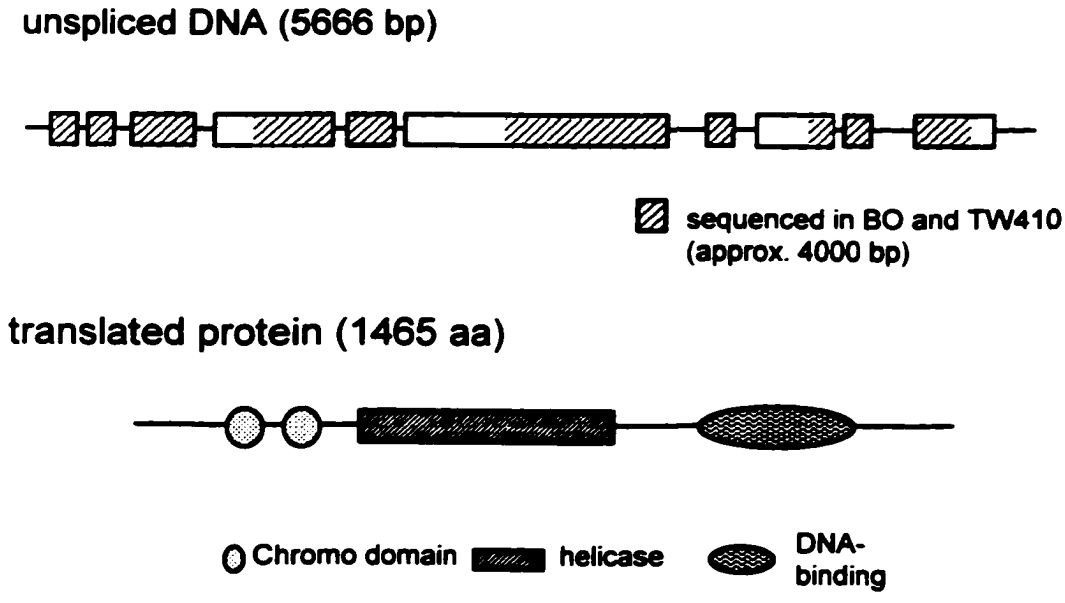


The conserved protein domains are indicated below the alignment. (Taken from Woodage et al, 1997)

sequence the coding region of H06O01.2. I concentrated on the conserved motifs (the chromodomain, the helicase and the DNA binding domains) and designed PCR primers to amplify the regions containing these motifs using BO and TW410 genomic DNA as template. I found that the conserved regions are identical in sequence to that of N2. To date, I have sequenced approximately 4000 kb of the 5666 kb of genomic sequence for H06O01.2 (Figure 13). Sequencing the remainder of the H06O01.2 ORF continues as molecular lesions are just as likely to be associated with non-conserved coding or non-coding sequences. It is also not unusual for lesions to affect regulatory sequences found at some distance upstream or downstream from the immediate gene. Identification of regulatory sequences and associated lesions are beyond the scope of this project but an important effort to be pursued.

To provide further evidence that the worm homologue of CHD1 is represented by *mut-2*, I attempted to produce a knock out using dsRNAi. By this method, I can ectopically interfere with gene function to reproduce a null phenotype similar to what would be seen for the genetic null. The mechanism behind RNAi is not understood but is gene specific, sensitive and reproducible for most but not all genes (Fire et al., 1998). I obtained an expressed sequenced tag (EST) cDNA clone of H06O01.2 (yk363e3) from the EST database (http://www.ddbj.nig.ac.jp/) and synthesized sense and anti-sense RNA from the cDNA insert. The two strands were mixed, annealed and injected as dsRNA into the gonads of N2 L4 hermaphrodites. The resulting progeny exhibited no obvious phenotype.

Figure 13. The genomic structure of H06O01.2 and its predicted protein.



My RNAi results agree with those of others who are investigating the role of chromodomain proteins in meiosis and chromosome segregation in *C. elegans* and the parasitic worm, *Ascaris* (Fritz Mueller, pers. comm). Thomas von Zelewsky generated antibodies specific for H06O01.2 to examine the CHD1 expression pattern. Preliminary results show ubiquitous expression localized to the nucleus of the somatic cells. In addition, expression is found in the oocytes. These results are consistent with the suggestion that H06O01.2 represents the best candidate for *mut-2* and only await identification of the associated molecular lesion for verification.

DISCUSSION

Summary of past and present work

The *mut-2* mutator, was previously identified in a screen for genetic loci that regulate or repress germ line activity of the transposon, Tc1. In addition to causing enhanced germ line transposition, the *mut-2* mutator results in a higher frequency of male progeny, a feature known as the Him phenotype. Males arise spontaneously in natural populations as a result of rare X chromosome non-disjunction events during meiosis. Therefore this phenotype indicates that *mut-2* plays a role in chromosome segregation during meiosis.

Subsequently, three new families of transposons, Tc3, Tc4 and Tc5, were identified in the *mut-2* mutator background. These families are present in most if not all strains of *C. elegans* but are active only in the presence of the *mut-2* mutator (Parham, 1994). These initial characteristics of the *mut-2* mutator: enhanced germ line transposition of multiple transposons families and increased X chromosome non-disjunction, provided the first clues that this gene played a genome-wide role in the general functional and physical behavior of the genetic material, chromatin.

During the course of this work I mapped *mut-2* to a small genetic interval between *let-86* and *sem-4*. I extended previous characterization of the *mut-2* mutator and provided evidence that supports a role for *mut-2* in chromatin stability and gametogenesis. *mut-2* mutator strains consistently generate a higher frequency of males due to X chromosome

non-disjunction and they are sterile at 25°C. The ts sterile phenotype proved to be recessive and I was able to demonstrate rescue with a genetically defined duplication, hDp65, and then by transformation rescue with a single cosmid H06O01. Of the four predicted ORF's represented by H06O01, the CHD1-like gene (H06O01.2) appears to be the most likely candidate.

Mapping efforts localize *mut-2* to a small interval

The work that I have described lends support to the idea that genome-wide mechanisms exist to maintain chromatin stability. By characterizing *mut-2,* I hoped to provide insight into such mechanisms. To do this, I set out to clone the *mut-2* gene. This first phase involved mapping the position of *mut-2* to a small genetic interval on the right arm of LG I between the genetic markers, *dpy-14* and *sem-4,* using the Him and Mut phenotypes. With completion of the entire *C. elegans* genome sequence by the Worm Sequencing Consortium (1998), I was able to correlate this genetic interval to a physical sequence of about 160 kb. This region is spanned by five overlapping cosmid clones. Sequence analysis of this region revealed 33 predicted genes (ORF's). These ORF's fall into three groups: those that have been cloned, those with significant homology to known genes; and those with no known homology to genes within existing databases.

Using the *C. elegans* database, ACeDB, I looked for candidate genes that fit the known roles of *mut*-2. That is genes with homology to germ line specific genes that might contain motifs of DNA recombination or repair, helicases or protein:nucleic acid interactions. I identified several candidates distributed across the five cosmids including *glh-1* (T21G5.3), a germ-line specific RNA helicase; an ORF with weak homology to KH

domains implying protein:RNA interactions; an ORF with weak homology to the Walker motif of bacterial mismatch repair genes; and an ORF with significant homology to CHD1 (H06O01.2), a member of the SWI/SNF superfamily of transcriptional regulators. These, and other less obvious potential candidates among the novel genes, are distributed across the 160 kb *mut-2* interval. Therefore I needed to identify an aspect of the *mut-2* mutator phenotype that would allow me to use standard DNA transformation techniques to rescue a defect and identify a r cosmid, and ultimately a single ORF, that rescued this defect. Both the Him and Mut phenotypes were useful for general mapping strategies, but too cumbersome for fine mapping or transformation rescue approaches. Therefore I investigated other features of *mut-2* mutants to identify a phenotype more amenable to rescue.

Extending the characterization of *mut-2*

I characterized a temperature sensitive sterile phenotype conferred by *mut-2(r459)*. I was able to attribute sterility to *mut-2(r459)* by showing that the *mut-2(+)* parent of TR674 is viable at 25°C, while TR674 and all its r459-bearing derivatives are sterile at this temperature. The basis of this sterility appears to be a defect in gametogenesis; both sperm ad oocytes are abnormal. With this ts sterile phenotype in hand and after several experimental trials to optimize my microinjection strategy, I demonstrated rescue of the *mut-2* ts phenotype with the cosmid, H06O01.

H06O01 contains four ORFs. Of these, H06O01.2 seems to represent the best candidate for the *mut-2* gene. It is 33% similar to CHD1, a member of the SWI/SNF superfamily of transcriptional regulators that function in a complex to re-organize

chromatin (Delmas et al., 1993; Pearce et al., 1992; Woodage et al., 1997)(Delmas et al., 1993; Pearce et al., 1992; Woodage et al., 1997). The *C. elegans* homologue retains three motifs which distinguishes the CHD1 protein: the chromodomains (chromodomain and shadow chromodomain) implicated in chromatin modification, the helicase and DEAH motif typical of ATP-dependent chromatin remodeling complexes and a sequence selective DNA-binding domain (Figure 12).

Evidence for *mut-2* as the *C. elegans* homologue for CHD1

Confirmation that CHD1 represents the *mut-2* gene awaits two results: rescue of the ts phenotype with the CHD1 ORF and identification of the molecular lesion responsible for *mut-2(r459)*. However, several lines of evidence support this idea. This evidence includes: 1) enhanced transposition of multiple transposon families, 2) increased X chromosome non-disjunction (Him) (Collins, 1987) and 3) ts sterility. Genome-wide regulation of transposition suggests general control at the level of transposase transcription or access to transposon ends. Defects in chromosome segregation suggest a role in chromosome behavior during meiosis or the associated recombination machinery or access by this machinery to chromatin. The temperature sensitive sterile defect suggests a role in some aspect of gametogenesis or germ line chromatin stability.

Observations that mutators exhibit germ line de-silencing further supports a role in transcriptional repression (Kelly, 1998). Finally, evidence that the *mut-2* mutator is resistant to interference by dsRNA suggests that this gene normally plays a role in the mechanisms that respond to these RNA intermediates by silencing the source of transcription (Tabara, 1999).

Role of *mut-2* in transposon control

In *C. elegans, mut-2* regulates multiple families of transposons including Tc1, Tc3, Tc4 and Tc5. With respect to germline transposition, several studies demonstrated that the only requirement for mobilization of a transposon is its own functional transposase (van Luenen et al., 1994; Vos et al., 1996; Vos and Plasterk, 1994). Transposases are family specific yet *mut-2* regulates transposition of multiple transposon families. This genome wide regulation was an important indicator of a more wide-spread role for *mut-2* in the maintenance of chromatin integrity.

There are several possible explanations for how *mut-2* regulates transposition. MUT-2 may act at the level of a transposase transcription complex, either directly interacting with a promoter complex or indirectly with chromatin to limit access of a promoter complex. It may regulate the stability of the transposase or the stability of the transposase complex composed of an excised transposon and its target sequence. Alternatively, it may exert a regulatory effect by modifying chromatin structure and limiting access of the transposase to the transposon ends.

One intriguing model is that *mut-2* is a component of a chromatin modifying complex that recruits and assembles silencing complexes a regions of chromatin that are then conformationally altered and transcriptionally "shut down". Similar mechanisms for generating large transcriptionally silent domains of chromatin have been described in yeast (telomeric and mating type silencing) and *Drosophila* (Polycomb Group (PcG) repression of homeotic genes and heterochromatic position effect variegation (PEV)) (reviewed by Pirotta, 1997). Recruiting complexes to silence extended regions of

chromatin might explain how *mut-2* is able to prevent the general transcription of transposases (and resulting transposition). If most transposons are located in heterochromatic, transcriptionally inactive regions, this provides a reasonable explanation of how they are silenced. Perhaps when transposons mobilize to euchromatic, transcriptionally active regions, they might carry some inherent feature (i.e. their repetitive nature) that allows silencing complexes to target them although this explanation remains to be tested.

Role of *mut-2* in chromosome segregation

In the gonad, during the process of germ line development, germ cells undergo meiosis. Homologues pair and then disjoin in meiosis I. Critical to this process is the formation of DSB intermediates that promote recombination events. These events may be resolved in two ways. They may result in an exchange of genetic information between two homologues. Or, they may be resolved so that no physical exchange occurs. These latter non-recombinant cross-overs are important because they form the chiasmata that provide the mechanical strain needed for proper disjunction of homologues during the first meiotic division. *mut-2* may play a role in homologue pairing, disjoining or in the associated recombination events.

In *C. elegans*, mutations that affect X chromosome segregation increase the frequency of male progeny (XO) and are known as the *him* genes. Mutations of the autosomal gene, *him-8*, specifically affect segregation of the X chromosome. Other *him* mutations, such as *him-5*, affect segregation of both the X chromosome and the autosomes. Analysis of *him-8* mutants revealed that while segregation is affected,

recombination on the X chromosome still occurs normally (Albertson et al., 1997). The difference is in the frequency and distribution of the cross-over events that initiate recombination along the X chromosome. The mechanism by which *him-8* controls segregation does not affect the recombination machinery but rather appears to affect access of this machinery to the chromatin. Therefore, if DSB's and cross-overs occur but are inappropriately positioned and processed, they may not be resolved properly and result in a reduction or lack of chiasmata. This would lead to a greater frequency of non-disjunction events (Him).

In *mut-2* mutator mutants, the Him phenotype observed is relatively mild ($\geq$ 0.4% males) compared to *him-8* (37% males) but it provides an additional clue to the normal function of *mut-2*. If *mut-2* is part of a complex that modifies chromatin to control transcription, the position of such a complex may be an important "decoration" of the chromatin that serves the dual roles of limiting access of certain modifying components to chromatin and protecting it. Evidence suggests that sites of meiotic homologue recognition and recombination in *C. elegans* are not random (Barnes et al., 1995; Zetka and Rose, 1995). This could be the result of structural modifiers that regulate access of the appropriate machinery to chromatin. Mutations that affect the stability of the chromatin modification complexes or their ability to find their normal targets may permit DSB promoting factors to randomly access "unprotected" chromatin and initiate events that lead to recombination without associated chiasmata formation. Whether *mut-2* affects recombination on the X chromosome (and/or the autosomes) has not been

addressed. Understanding this may shed light on how *mut-2* functions in homologue paring or dis-junction or in processes that permit or promote recombination.

mut-2 and the segregation of extrachromosomal elements

My preliminary observations regarding free duplications in the *mut-2* mutator background are also related to the issue of chromosome segregation. I noted that chromosomal duplications were lost at a much lower frequency in a mutator background (about 5% loss) than in the genetic backgrounds in which they are maintained (approximately 50% loss). Retention could be related to a selective advantage conferred by the duplications in this mutator background. In this case, most of the progeny would be expected to carry the duplication. Alternatively, it could imply a role for *mut-2* in ensuring segregation of true chromosomes but limiting segregation of extrachromosomal elements. How the *mut-2* mutator affects stability, retention and possibly recombination of extrachromosomal elements is another interesting line of investigation that would shed light on the role of *mut-2* in chromatin behavior.

A role in gametogenesis

In *C. elegans*, as the two-armed somatic gonad matures, gametes proliferate mitotically and then enter a meiotic phase. In each gonad arm of adult hermaphrodites, the first approximately 40-50 pre-meiotic germ cells develop into a population of about 150 spermatocytes. The remaining gametes develop as oocytes (Schedl, 1997). Therefore the mature hermaphrodite gonad contains several hundred sperm at the tip proximal to the vulva and about a thousand oocytes in spatially and temporally

progressive stages of meiotic development originating from the distal tip and extending towards the proximal tip.

To gain a better understanding of the role of *mut-2* in gamete formation, I characterized the ts sterile phenotype in some detail. Eric Lambie (Dartmouth College), examined the gonads of wild type and *mut-2* animals raised at 25°C by Nomarski microscopy and revealed that in *mut-2(r459)* hermaphrodites raised at 25°C, sperm were reduced in number and morphologically abnormal. Therefore it was not surprising that the hermaphrodites were sterile. Oocytes appeared normal in number and gross morphology. However, they were not competent for fertilization when crossed with WT sperm at 25°C, indicating that they are defective as well. Collectively, these results indicate that *mut-2* is involved in some aspect of gametogenesis in general. Because the sperm are chronologically the oldest gametes populating the gonad, the effects of *mut-2(r459)* may result in more obvious degeneration and morphological abnormalities. These defects may be mild at 20°C leading to fertile progeny but a reduced brood size and exacerbated at higher temperatures, resulting in 100% sterility.

In *C. elegans*, many mutations that result in sterility have been described. Some of the best studied mutations include the *glp* mutants that exhibit abnormal to no germ line proliferation, the *gld* mutants with defects in germ line differentiation and the *mes* mutants which lack maternally supplied products necessary for germ line proliferation (maternal effect sterile). Among these sterile mutant classes, it is intriguing to note that the *mes* genes share the germ line de-silencing phenotype exhibited by *mut-2* (Kelly, 1998).

The *mes* genes are interesting also in part because of their "grandchildless" phenotype. Heterozygous hermaphrodites provide the necessary WT maternal factors to support germ line development in their homozygous mutant progeny but the mutant hermaphrodite progeny are sterile (no grandchildren) because of a lack of maternally supplied factors necessary to support post embryonic germ line development.

In *C. elegans,* of 6 *mes* genes, *mes-2* and *mes-6* have been cloned and are similar to the PcG of genes first characterized in *Drosophila* (Korf et al., 1998). The PcG genes and their homologues encode transcriptional repressors of homeotic genes that are required for proper patterning and development in a wide variety of organisms (reviewed by Pirotta, 1997). PcG proteins are thought to assemble into complexes through protein:protein interactions via motifs characteristic of each PcG protein. Examples of these motifs include the chromodomain of Pc and HP1 (heterochromatin protein 1), SET domain of *mes-2* and its Drosophila homologue, *Enhancer of zeste (E(z))* and the WD40 repeats of *mes-6* and its fly homologue, *extra sex combs (esc)* . Studies suggest that heteromeric PcG complexes localize to chromatin response elements from which they direct or extend their repressive influence (Chan, 1994). The mechanism of repression is not known but might involve modifications to the higher order structure of chromatin. How the *mes* genes affect chromatin structure and germ line viability is not clear, however, it is likely that mutations affecting these developmentally specific regulators lead to inappropriately timed gene expression and death to the developing gametes (Korf et al., 1998).

While *mut-2* is not known to be a maternal effect gene, it clearly affects germ line viability. If MUT-2 is part of a general transcriptional repressor, the *mut-2* mutator may permit inappropriate gene expression that leads to gamete cell death and degeneration of the germ line, similar to the *mes* mutants. One could argue that the germ line de-silencing phenotype observed by Kelly and co-workers (1998) in *mes* and *mut-2* mutants lends support to such a role as a transcriptional regulator. Speculating further, one could propose that *mut-2* may function by maintaining chromatin in a repressed state like the *mes* genes.

Finally, the observation that mutator strains are resistant to interference by dsRNA suggests that *mut-2* may play a role in a germ line process that targets dsRNA intermediates. The mechanisms that result in knocking out gene expression are unknown. It is thought that dsRNA processing may be part of a naturally occurring defense mechanism in many organisms that recognize these intermediates as harbingers of retroviral infection. Once "infection" is perceived, the interference machinery is somehow mobilized to the source of the dsRNA. Identification of *mut-7* as an RNAse D homologue that targets dsRNA has lead to a model whereby *mut-7* participates by targeting endogenous (read-through transcripts of transposons) or ectopically introduced RNA intermediates. The interference machinery may then be mobilized to the source (the transposon) or the endogenous gene and effectively prevent further transcription. *mut-2* may fit into this model for interference as a component of the silencing apparatus that converts the chromatin where the transposon resides to transcriptionally silent "real estate".

Conclusion

mut-2 is an important participant in maintaining chromatin stability in the germ line. The collective results of the work I have described and the work of others serve to extend the phenotypes attributed to *mut-2* and provide additional clues to its normal role. These features predict that *mut-2* plays a role in transposition regulation, meiotic chromatin behavior, gametogenesis, and germ line gene silencing. Whether it functions as a structural component that modifies chromatin or as a regulatory component of a transcriptional apparatus awaits cloning of the *mut-2* gene.

Based on the mutator phenotypes observed and the best candidate ORF on the rescuing cosmid (H06O01.2 - CHD1), *mut-2* probably functions by interacting with components of a larger regulatory complex. The likely role of this complex is to regulate general transcription as part of a coordinated developmental plan in the germ line that integrates multiple processes during gametogenesis. These processes are likely to include maintenance of a chromatin structure that permits proper meiotic homologue segregation, general transcriptional repression and silencing of aberrant expression. The mechanism remains to be elucidated but is likely to involve accessibility to the chromatin of the different regulatory complexes that direct each of these processes in the germ line.

In *mut-2* mutator strains, aspects of these processes were first glimpsed as enhanced germ line transposition of multiple families and X chromosome non-disjunction. Further characterization of *mut-2* revealed temperature sensitive defects in

gametogenesis (this work). Others found that the normally quiescent germ line became de-silenced for gene expression in mutants affecting fertility including *mut-2* (Kelly, 1998). Still others found that the *mut-2* germ line was unresponsive to interference by dsRNA (Tabara, 1999). Therefore, the regulation of transposons in the germ line has emerged an important model for investigating the dynamic processes required to maintain chromatin integrity and plasticity.

BIBLIOGRAPHY

Albertson, D. G., Rose, A. M., and Villeneuve, A. M. (1997). Chromosome Organization, Mitosis, and Meiosis. In *C. elegans* II, D. L. Riddle, T. Blumenthal, B. J. Meyer and J. R. Priess, eds. (Cold Spring Harbor: Cold Spring Harbor Press), pp. 47-78.

Almeida, C. (1996). Targeted Gene Replacement to Mimic Familial Hypertrophic Cardiomyopathy in *Caenorhabditis elegans*. In Biochemistry (Durham, NH: Universtiy of New Hampshire), pp. 83.

Baker, T. A., and Luo, L. (1994). Identification of residues in the Mu transposase essential for catalysis. Proc Natl Acad Sci U S A *91*, 6654-8.

Barnes, T. M., Kohara, Y., Coulson, A., and Hekimi, S. (1995). Meiotic recombination, noncoding DNA and genomic organization in Caenorhabditis elegans. Genetics *141*, 159-79.

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

Benian, G. M., Tang, X., and Tinley, T. L. (1996). Twitchin and related giant Ig superfamily members of C. elegans and other invertebrates. Adv Biophys *33*, 183-98.

Berg, D. E., and Howe, M. M. (1989). Mobile DNA (Washington, D.C.: American Society for Microbiology).

Bingham, P. M., Levis, R., and Rubin, G. M. (1981). Cloning of DNA sequences from the white locus of D. melanogaster by a novel and general method. Cell *25*, 693-704.

Boeke, J. D., Garfinkel, D. J., Styles, C. A., and Fink, G. R. (1985). Ty elements transpose through an RNA intermediate. Cell *40*, 491-500.

Bradley, D., Carpenter, R., Sommer, H., Hartley, N., and Coen, E. (1993). Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the plena locus of Antirrhinum. Cell *72*, 85-95.

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

Britten, R. J. (1997). Mobile elements inserted in the distant past have taken on important functions. Gene *205*, 177-82.

Church, G. M., and Gilbert, W. (1984). Genomic sequencing. Proc Natl Acad Sci U S A *81*, 1991-5.

Collins, J., Forbes, E., and Anderson, P. (1989). The Tc3 family of transposable elements in *C.elegans*. Genetics *121*, 47-55.

Collins, J., Saari, B., and Anderson, P. (1987). Activation of a transposable element in the germ line but not the soma of *Caenorhabditis elegans*. Nature *328*, 726-728.

Collins, J. J., and Anderson, P. (1994). The Tc5 family of transposable elements in Caenorhabditis elegans. Genetics *137*, 771-81.

Consortium, S., and Sequencing, T. C. e. (1998). Genome Sequence of the Nematode *C. elegans*: A platform for investigating Biology. Science *282*, 2012-2018.

Delmas, V., Stokes, D. G., and Perry, R. P. (1993). A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain. Proc Natl Acad Sci U S A *90*, 2414-8.

Dimitri, P., and Junakovic, N. (1999). Revising the selfish DNA hypothesis: new evidence on accumulation of transposable elements in heterochromatin. Trends Genet *15*, 123-4.

Doak, T. G., Doerder, F. P., Jahn, C. L., and Herrick, G. (1994). A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. Proc Natl Acad Sci U S A *91*, 942-6.

Dreyfus, D. H., and Emmons, S. W. (1991). A transposon-related palindromic repetitive sequence from C. elegans. Nucleic Acids Res *19*, 1871-7.

Edgley, M. L., Baillie, D. L., Riddle, D. L., and Rose, A. M. (1995). Genetic balancers. Methods Cell Biol *48*, 147-84.

Egilmez, N. K., Ebert, R. H., 2nd, and Shmookler Reis, R. J. (1995). Strain evolution in Caenorhabditis elegans: transposable elements as markers of interstrain evolutionary history. J Mol Evol *40*, 372-81.

Eide, D., and Anderson, P. (1985b). The gene structures of spontaneous mutations affecting a *C. elegans* myosin heavy chain gene. Genetics *109*, 67-79.

Eide, D., and Anderson, P. (1988). Insertion and Excision of *C. elegans* transposable element Tc1. Mol. and Cell. Biol. *8*, 737-746.

Eide, D., and Anderson, P. (1985). Transposition of Tc1 in the nematode Caenorhabditis elegans. Proc Natl Acad Sci U S A *82*, 1756-60.

Emerson, R. A. (1929). The frequency of somatic mutation in variegated pericarp of maize. Genetics *14*, 488-511.

Emerson, R. A. (1917). Genetical studies of variegated pericarp maize. Genetics *2*, 1-35.

Emerson, R. A. (1914). The inheritance of recurring somatic variation in variegated ears of maize. Am. Nat. *48*, 87-115.

Emmons, S. W., Klass, M. R., and Hirsh, D. (1979). Analysis of the constancy of DNA sequences during development and evolution of the nematode Caenorhabditis elegans. Proc Natl Acad Sci U S A *76*, 1333-7.

Emmons, S. W., Yesner, L., Ruan, K. S., and Katzenberg, D. (1983). Evidence for a transposon in Caenorhabditis elegans. Cell *32*, 55-65.

Engels, W. R. (1983). The P family of transposable elements in Drosophila. Annu Rev Genet *17*, 315-44.

Fanti, L., Giovinazzo, G., Berloco, M., and Pimpinelli, S. (1998). The heterochromatin protein 1 prevents telomere fusions in Drosophila. Mol Cell *2*, 527-38.

Fedoroff, N. (1989). Maize Transposable Elements. In Mobile DNA, D. E. Berg and M. M. Howe, eds. (Washington, D.C.: American Society for Microbiology), pp. 375-412.

Fedoroff, N., Schlappi, M., and Raina, R. (1995). Epigenetic regulation of the maize Spm transposon. Bioessays *17*, 291-7.

Files, J. G., Carr, S., and Hirsh, D. (1983). Actin gene family of Caenorhabditis elegans. J Mol Biol *164*, 355-75.

Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. Trends in Genetics *5*, 103-107.

Finney, M., and Horvitz, H. R. (1987). Mutator Mapped: Worm Breeder's Gazettee).

Finney, M., Ruvkun, G., and Horvitz, H. R. (1988). The C. elegans cell lineage and differentiation gene unc-86 encodes a protein with a homeodomain and extended similarity to transcription factors. Cell *55*, 757-69.

Fire, A., and Waterston, R. H. (1989). Proper expression of mysoin genes in transgenic nematodes. EMBO J *8*, 3419-3428.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans [see comments]. Nature *391*, 806-11.

Fridell, R. A., Pret, A. M., and Searles, L. L. (1990). A retrotransposon 412 insertion within an exon of the Drosophila melanogaster vermilion gene is spliced from the precursor RNA. Genes Dev *4*, 559-66.

Garfinkel, D. J., Boeke, J. D., and Fink, G. R. (1985). Ty element transposition: reverse transcriptase and virus-like particles. Cell *42*, 507-17.

Gloor, G. B., Nassif, N. A., Johnson-Schlitz, D. M., Preston, C. R., and Engels, W. R. (1991). Targeted gene replacement in Drosophila via P element-induced gap repair. Science *253*, 1110-1117.

Green, and Hillier. (1998). GeneFinder.

Greene, B., Walko, R., and Hake, S. (1994). Mutator insertions in an intron of the maize knotted1 gene result in dominant suppressible mutations. Genetics *138*, 1275-85.

Greenwald, I. (1985). lin-12, a nematode homeotic gene, is homologous to a set of mammalian proteins that includes epidermal growth factor. Cell *43*, 583-90.

Grindley, N. D. F., and Leschziner, A. E. (1995). DNA transposition: from black box to color monitor. Cell *83*, 1063-66.

Hart, C. M., Fischer, B., Neuhaus, J. M., and Meins, F., Jr. (1992). Regulated inactivation of homologous gene expression in transgenic Nicotiana sylvestris plants containing a defense-related tobacco chitinase gene. Mol Gen Genet *235*, 179-88.

Hodgkin, J., Papp, A., Pulak, R., Ambros, V., and Anderson, P. (1989). A new kind of informational suppression in the nematode Caenorhabditis elegans. Genetics *123*, 301-13.

Hodgkin, J. A., and Brenner, S. (1977). Mutations causing transformation of sexual phenotype in the nematode Caenorhabditis elegans. Genetics *86*, 275-87.

Hodgkin, J. A., Horvitz, H. R., and Brenner, S. (1979). Non-disjunction mutants of *C. elegans*. Genetics *91*, 67-94.

Izant, J. G., and Weintraub, H. (1984). Inhibition of thymidine kinase gene expression by anti-sense RNA: a molecular approach to genetic analysis. Cell *36*, 1007-15.

Jasinskiene, N., Coates, C. J., Benedict, M. Q., Cornel, A. J., Rafferty, C. S., James, A. A., and Collins, F. H. (1998). Stable transformation of the yellow fever mosquito, Aedes aegypti, with the Hermes element from the housefly. Proc Natl Acad Sci U S A *95*, 3743-7.

Kelly, W. G., and Fire, A. (1998). Chromatin silencing and the maintenance of a functional germline in Caenorhabditis elegans. Development *125*, 2451-6.

Kelly, W. G., Xu, S., Montgomery, M. K., and Fire, A. (1997). Distinct requirements for somatic and germline expression of a generally expressed Caernorhabditis elegans gene. Genetics *146*, 227-38.

Kempken, F., Jacobsen, S., and Kuck, U. (1998). Distribution of the fungal transposon Restless: full-length and truncated copies in closely related strains. Fungal Genet Biol *25*, 110-8.

Kennerdell, J. R., and Carthew, R. W. (1998). Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. Cell *95*, 1017-26.

Ketting, R. F., Fischer, S. E. J., and Plasterk, R. H. (1997). Target choice determinants of the Tc1 transposon of Caenorhabditis elegans. Nucleic Acids Res *25*, 4041-7.

Kidwell, M. G., and Lisch, D. (1997). Transposable elements as sources of variation in animals and plants. Proc Natl Acad Sci U S A *94*, 7704-11.

Kim, K., Namgoong, S. Y., Jayaram, M., and Harshey, R. M. (1995). Step-arrest mutants of phage Mu transposase. Implications in DNA- protein assembly, Mu end cleavage, and strand transfer. J Biol Chem *270*, 1472-9.

Kleckner, N. (1996). Meiosis: how could it work? Proc Natl Acad Sci U S A *93*, 8167-74.

Korf, I., Fan, Y., and Strome, S. (1998). The Polycomb group in Caenorhabditis elegans and maternal control of germline development. Development *125*, 2469-78.

Korswagen, H. C., Durbin, R. M., Smits, M. T., and Plasterk, R. H. (1996). Transposon Tc1-derived, sequence-tagged sites in Caenorhabditis elegans as markers for gene mapping. Proc Natl Acad Sci U S A *93*, 14680-5.

Kulkosky, J., Jones, K. S., Katz, R. A., Mack, J. P., and Skalka, A. M. (1992). Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol Cell Biol *12*, 2331-8.

Labrador, M., and Corces, V. G. (1997). Transposable element-host interactions: regulation of insertion and excision. Annu Rev Genet *31*, 381-404.

Li, W., and Shaw, J. E. (1993). A variant Tc4 transposable element in the nematode C. elegans could encode a novel protein. Nucleic Acids Res *21*, 59-67.

Liao, L. W., Rosenzweig, B., and Hirsh, D. (1983). Analysis of a transposable element in Caenorhabditis elegans. Proc Natl Acad Sci U S A *80*, 3585-9.

Lohe, A. R., De Aguiar, D., and Hartl, D. L. (1997). Mutations in the mariner transposase: the D,D(35)E consensus sequence is nonfunctional. Proc Natl Acad Sci U S A *94*, 1293-7.

Loukeris, T. G., Arca, B., Livadaras, I., Dialektaki, G., and Savakis, C. (1995). Introduction of the transposable element Minos into the germ line of Drosophila melanogaster. Proc Natl Acad Sci U S A *92*, 9485-9.

Marin, I., Platarengifo, P., Labrador, M., and Fontdevial, A. (1998). Evolutionary relationships among the members of an ancient class of non-lTR retrotransposons found in the nematode *C. elegans*. Molecular Biology and Evolution *15*, 1390-1402.

McClintock, B. (1951). Chromosome Organization and Genic Expression. Cold Spring Harbor Symp. Quant. Biol. *16*, 13-47.

McClintock, B. (1947). Cytogenetic studies of maize and Neurospora. Carnegie Inst. Wash. Year Book *46*, 146-152.

McClintock, B. (1987). The Discovery and Characterization of Transposable Elements, Volume 17, J. A. Moore, ed. (New York: Garland Publishing, Inc.).

McClintock, B. (1946). Maize Genetics. Carnegie Inst. Wash. Year Book *45*, 176-186.

McClintock, B. (1948). Mutable loci in Maize. Carnegie Inst. Wash. Year Book *47*, 155-169.

Medhora, M., Maruyama, K., and Hartl, D. L. (1991). Molecular and functional analysis of the mariner mutator element Mos1 in Drosophila. Genetics *128*, 311-8.

Mello, C., and Fire, A. (1995). DNA transformation. Methods Cell Biol *48*, 451-82.

Mello, C. C., Kramer, J. M., Stinchcomb, D., and Ambros, V. (1991). Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences. Embo J *10*, 3959-70.

Misra, S., and Rio, D. C. (1990). Cytotype control of Drosophila P element transposition: the 66 kd protein is a repressor of transposase activity. Cell *62*, 269-84.

Mizuuchi, K. (1992). Transpositional recombination: mechanistic insights from studies of mu and other elements. Annu Rev Biochem *61*, 1011-51.

Moerman, D. G., Benian, G. M., and Waterston, R. H. (1986). Molecular cloning of the muscle gene unc-22 in Caenorhabditis elegans by Tc1 transposon tagging. Proc Natl Acad Sci U S A *83*, 2579-83.

Moerman, D. G., and Waterston, R. H. (1984). Spontaneous unstable unc-22 IV mutations in C. elegans var. Bergerac. Genetics *108*, 859-77.

Montgomery, M. K., and Fire, A. (1998). Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression. Trends Genet *14*, 255-8.

Montgomery, M. K., Xu, S., and Fire, A. (1998). RNA as a target of double-stranded RNA-mediated genetic interference in Caenorhabditis elegans. Proc Natl Acad Sci U S A *95*, 15502-7.

Moore, J. K., and Haber, J. E. (1996). Capture of retrotransposon DNA at the sites of chromosomal double- strand breaks [see comments]. Nature *383*, 644-6.

Mori, I., Moerman, D. G., and Waterston, R. H. (1988). Analysis of a mutator activity necessary for germline transposition and excision of Tc1 transposable elements in Caenorhabditis elegans. Genetics *120*, 397-407.

Nellen, W., and Lichtenstein, C. (1993). What makes an mRNA anti-sense-itive? Trends Biochem Sci *18*, 419-23.

Ngo, H., Tschudi, C., Gull, K., and Ullu, E. (1998). Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. Proc Natl Acad Sci U S A *95*, 14687-92.

O'Kane, C. J., and Gehring, W. J. (1987). Detection in situ of genomic regulatory elements in Drosophila. Proc Natl Acad Sci U S A *84*, 9123-7.

Ochman, H., Gerber, A. S., and Hartl, D. L. (1988). Genetic applications of an inverse polymerase chain reaction. Genetics *120*, 621-3.

Pardue, M. L., Danilevskaya, O. N., Lowenhaupt, K., Slot, F., and Traverse, K. L. (1996). Drosophila telomeres: new views on chromosome evolution. Trends Genet *12*, 48-52.

Pearce, J. J., Singh, P. B., and Gaunt, S. J. (1992). The mouse has a Polycomb-like chromobox gene. Development *114*, 921-9.

Pimpinelli, S., Berloco, M., Fanti, L., Dimitri, P., Bonaccorsi, S., Marchetti, E., Caizzi, R., Caggese, C., and Gatti, M. (1995). Transposable elements are stable structural components of Drosophila melanogaster heterochromatin. Proc Natl Acad Sci U S A *92*, 3804-8.

Plasterk, R. H., and Groenen, J. T. (1992). Targeted alterations of the Caenorhabditis elegans genome by transgene instructed DNA double strand break repair following Tc1 excision. Embo J *11*, 287-90.

Plasterk, R. H. A., and van Luenen, H. G. A. M. (1997). Transposons. In *C. elegans* II, D. L. Riddle, T. Blumethal, B. Meyer and J. R. Priess, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 97-116.

Rezsohazy, R., van Luenen, H. G., Durbin, R. M., and Plasterk, R. H. (1997). Tc7, a Tc1-hitch hiking transposon in Caenorhabditis elegans. Nucleic Acids Res *25*, 4048-54.

Robertson, H. M., and Engels, W. R. (1989). Modified P elements that mimic the P cytotype in Drosophila melanogaster. Genetics *123*, 815-24.

Rosenzweig, B., Liao, L. W., and Hirsh, D. (1983). Sequence of the C. elegans transposable element Tc1. Nucleic Acids Res *11*, 4201-9.

Rosenzweig, B., Liao, L. W., and Hirsh, D. (1983). Target sequences for the C. elegans transposable element Tc1. Nucleic Acids Res *11*, 7137-40.

Roussell, D. L., and Bennett, K. L. (1993). glh-1, a germ-line putative RNA helicase from Caenorhabditis, has four zinc fingers. Proc Natl Acad Sci U S A *90*, 9300-4.

Ruan, K., and Emmons, S. W. (1984). Extrachromosomal copies of transposon Tc1 in the nematode Caenorhabditis elegans. Proc Natl Acad Sci U S A *81*, 4018-22.

Rubin, G. M., and Spradling, A. C. (1982). Genetic transformation of Drosophila with transposable element vectors. Science *218*, 348-53.

Rushforth, A. M., and Anderson, P. (1996). Splicing removes the Caenorhabditis elegans transposon Tc1 from most mutant pre-mRNAs. Mol Cell Biol *16*, 422-9.

Rushforth, A. M., Saari, B., and Anderson, P. (1993). Site-selected insertion of the transposon Tc1 into a Caenorhabditis elegans myosin light chain gene. Mol Cell Biol *13*, 902-10.

Ruvolo, V., Hill, J. E., and Levitt, A. (1992). The Tc2 transposon of Caenorhabditis elegans has the structure of a self-regulated element. DNA Cell Biol *11*, 111-22.
SanMiguel, P., Tikhonov, A., Jin, Y. K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P. S., Edwards, K. J., Lee, M., Avramova, Z., and Bennetzen, J. L. (1996). Nested retrotransposons in the intergenic regions of the maize genome [see comments]. Science *274*, 765-8.
Schedl, T. (1997). Developmental Genetics of the Germ Line. In *C. elegans II*, D. L. Riddle, T. Blumethal, B. Meyer and J. R. Priess, eds. (Cold Spring Harbor: Cold Spring Harbor Laboratory Press), pp. 241-269.
Spradling, A. C., and Rubin, G. M. (1982). Transposition of cloned P elements into Drosophila germ line chromosomes. Science *218*, 341-7.
Spradling, A. C., Stern, D. M., Kiss, I., Roote, J., Laverty, T., and Rubin, G. M. (1995). Gene disruptions using P transposable elements: an integral component of the Drosophila genome project. Proc Natl Acad Sci U S A *92*, 10824-30.
Stinchcomb, D. T., Shaw, J. E., Carr, S. H., and Hirsh, D. (1985). Extrachromosomal DNA transformation of Caenorhabditis elegans. Mol Cell Biol *5*, 3484-96.
Sulston, J., and Brenner, S. (1974). The DNA of *C. elegans*. Genetics *77*, 95-104.
Tabara, H., Grishok, A., and Mello, C. C. (1998). RNAi in C. elegans: soaking in the genome sequence. Science *282*, 430-1.
Tabara, H., Sarkissian, M., Timmons, L., Fire, A., and Mello, C. (1999). Genetic analysis of RNA interference in *C. elegans*: International Worm Meeting, Madison, WI).
Teng, S. C., Kim, B., and Gabriel, A. (1996). Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks [see comments]. Nature *383*, 641-4.
Timmons, L., and Fire, A. (1998). Specific interference by ingested dsRNA [letter]. Nature *395*, 854.
Tseng, J. C., Zollman, S., Chain, A. C., and Laski, F. A. (1991). Splicing of the Drosophila P element ORF2-ORF3 intron is inhibited in a human cell extract. Mech Dev *35*, 65-72.
van Luenen, H. G., Colloms, S. D., and Plasterk, R. H. (1994). The mechanism of transposition of Tc3 in C. elegans. Cell *79*, 293-301.
van Luenen, H. G., Colloms, S. D., and Plasterk, R. H. (1993). Mobilization of quiet, endogenous Tc3 transposons of Caenorhabditis elegans by forced expression of Tc3 transposase. Embo J *12*, 2513-20.
van Luenen, H. G., and Plasterk, R. H. (1994). Target site choice of the related transposable elements Tc1 and Tc3 of Caenorhabditis elegans. Nucleic Acids Res *22*, 262-9.
Voinnet, O., Vain, P., Angell, S., and Baulcombe, D. C. (1998). Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic promoterless DNA. Cell *95*, 177-87.
Vos, J. C., De Baere, I., and Plasterk, R. H. (1996). Transposase is the only nematode protein required for in vitro transposition of Tc1. Genes Dev *10*, 755-61.
Vos, J. C., and Plasterk, R. H. (1994). Tc1 transposase of Caenorhabditis elegans is an endonuclease with a bipartite DNA binding domain. Embo J *13*, 6125-32.

Vos, J. C., van Luenen, H. G., and Plasterk, R. H. (1993). Characterization of the Caenorhabditis elegans Tc1 transposase in vivo and in vitro. Genes Dev *7*, 1244-53.

Wood, W. B., Hecht, R., Carr, S., Vanderslice, R., Wolf, N., and Hirsh, D. (1980). Parental effects and phenotypic characterization of mutations that affect early development in Caenorhabditis elegans. Dev Biol *74*, 446-69.

Woodage, T., Basrai, M. A., Baxevanis, A. D., Hieter, P., and Collins, F. S. (1997). Characterization of the CHD family of proteins. Proc Natl Acad Sci U S A *94*, 11472-7.

Youngman, S., van Luenen, H. G., and Plasterk, R. H. (1996). Rte-1, a retrotransposon-like element in Caenorhabditis elegans. FEBS Lett *380*, 1-7.

Yuan, J., Finney, M., Tsung, N., and Horvitz, H. R. (1991). Tc4, a *C.elegans* transposable element with an unusual fold-back structure. PNAS USA *88*, 3334-3338.

Zetka, M., and Rose, A. (1995). The genetics of meiosis in Caenorhabditis elegans. Trends Genet *11*, 27-31.

Zwaal, R. R., Broeks, A., van Meurs, J., Groenen, J. T., and Plasterk, R. H. (1993). Target-selected gene inactivation in Caenorhabditis elegans by using a frozen transposon insertion mutant bank. Proc Natl Acad Sci U S A *90*, 7431-5.