ABSTRACT

Identification of Phenotypes in *Caenorabhditis elegans*
on the Basis of Sequence Similarity

Sushil Batra, Ph.D.

Mentors: Erich J. Baker, Ph.D., Myeongwoo Lee, Ph.D.

In biomedical research, *Caenorabhditis elegans* is an ideal choice as experimental organism due to striking similarity with human genome and its distinct features such as short life span, small reproductive cycle, simple body plan, easily observable mutant phenotypes and ease of cultivation in laboratory. The 97 megabase genomic sequence of *C. elegan*s comprises approximately 19,920 genes, of which about 2807 genes (14% of total genome) are uniquely associated with one or more RNAi phenotypes. The challenge to assign phenotypes to remaining 86% genes has incited development of new rapid techniques and computational tools.

Objective of this project was to identify phenotypes in *C. elegans* on the basis of sequence similarity using bioinformatics techniques. To find similarity in genes, we used BLAST as computational tool and predicted the phenotypes. Bi-directional pair wise BLAST was performed on 2,807 unique genes (associated with known phenotypes) against 19,920 genes. As a result, 141 new genes (with unknown phenotype) were obtained which share high sequence similarity with known RNAi phenotype genes of 16 categories.

In the present work, putative genes associated with two phenotypes, Ste (37 genes) and Unc (29 genes), were studied by RNA interference (RNAi) in laboratory. The outcome of these experiments assigned sterility phenotype to 8 new genes and uncoordinated phenotype to 12 new genes which were not linked with any phenotype in previous studies. These observations were further verified by silencing the response using reverse transcriptase polymerase chain reaction (RT-PCR) for Ste genes. Thus, bioinformatics techniques were successfully utilized in identification of phenotypes on the basis of sequence similarity with a relatively high success rate of 22% and 41% for sterility and uncoordinated phenotypes respectively. High success rate of this bioinformatics technique will allow researchers to focus their efforts on identifying particular phenotypes of interest and understanding various biological processes and elucidating the pathogenesis of diseases.

Identification of Phenotypes in *Caenorhabditis elegans*
on the Basis of Sequence Similarity

by

Sushil Batra

A Dissertation

Approved by the Institute of Biomedical Studies

_____
Robert R. Kane, Ph.D., Director

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

Approved by the Dissertation Committee

_____
Erich J. Baker, Ph.D., Co-Chairperson

_____
Myeongwoo Lee, Ph.D., Co-Chairperson

_____
Christopher M. Kearney, Ph.D.

_____
Sang-Chul Nam, Ph.D.

_____
Bryan Gibbon, Ph.D.

Accepted by the Graduate School
May 2009

_____
J. Larry Lyon, Ph.D., Dean

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Abbreviation	Full Form
µg	Microgram
µl	Microliter
% w/v	Percentage weight by volume
% w/w	Percentage weight by weight
°C	Degree Celsius
ADK	Altschul-Dembo-Karlin statistics
BLAST	Basic Local Alignment Search Tool
BLASTn	Nucleotide-nucleotide BLAST
BLASTp	Protein-protein BLAST
BLASTx	Translated query vs. protein database BLAST
BME	Beta mercaptoethanol
C. elegans	*Caenorabhditis elegans*
$CaCl_2$	Calcium chloride
cDNA	Complementary deoxyribonucleic acid
DI	Deionized water
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DPC4	Deleted in Pancreatic Cancer, locus 4
dsRNA	Double stranded RNA
DTT	Dithiothreitol

ABBREVIATIONS (Continued)

Abbreviation	Full Form
E. coli	*Escherichia coli*
E-value	Expect value (BLAST cut off)
FOXO	Forkhead transcription factor, sub-group O
gm	Gram
IPTG	Isopropyl-beta-D-thiogalactopyranoside
KCl	Potassium chloride
KH_2PO_4	Potassium phosphate monobasic
L	Liter
L1 stage	Larval stage 1
L2 stage	Larval stage 2
L2d stage	Dauer larva
L3 stage	Larval stage 3
L4 stage	Larval stage 4
LB Broth	Lysogeny Broth
M	Molar
mg	Milligram
mg/ml	Milligram per milliliter
$MgSO_4$	Magnesium sulfate
OMIM	Online Mendelian Inheritance in Man
ORF	Open reading frame
PCR	Polymerase chain reaction

ABBREVIATIONS (Continued)

Abbreviation	Full Form
PKD-1	Polycystin-1
PKD-2	Polycystin-2
P-value	Probability value
RdRP	RNA-dependent RNA polymerase
RING	Really interesting new gene
RISC	RNA induced silencing complex
RNA	Ribonucleic acid
RNAi	RNA mediated interference
RNase	Ribonuclease
rpm	Rotations per minute
RT buffer	Reverse transcriptase buffer
RT-PCR	Reverse transcription PCR
siRNA	Small interfering RNA
SMA	Spinal muscular atrophy
SMN	Survival motor neuron
SSII RT	SuperScript™ II reverse transcriptase
Ste	Sterility phenotype
Taq	Taq DNA polymerase
TBLASTn	Protein query vs. translated database BLAST
TBLASTx	Translated query vs. translated database BLAST
Unc	Uncoordinated movement

ACKNOWLEDGMENTS

Inscription of dissertation is a very special occasion in my life. In this unique moment, I express my profound gratitude to God and all the people whose contribution in assorted ways assisted me to accomplish this work successfully. All these wonderful people deserve a special mention in my humble acknowledgments.

At this outset, I deem it a great privilege to express revered gratitude and regards to my supervisors Dr. Erich Baker and Dr. Myeongwoo Lee. It is a great privilege to work with two advisors with entirely different areas of expertise. Dr. Baker introduced me to the field of bioinformatics and Dr. Lee provided invaluable guidance to verify the outcomes of bioinformatics research in laboratory using various biological and biomedical research methods. Their wholehearted and persistent help, excellent guidance, encouragement and moral support throughout this work was the actual driving force behind its successful completion.

I would like to thank to other members of my graduate committee, Dr. Chris Kearney, Dr. Sang-Chul Nam, and Dr. Bryan Gibbon for their continuous support and critical review of my research work. I express my heartfelt and respectful thanks to Dr. Kearney who not only allowed me to use his lab facilities for my research but also offered profound suggestions. Continuous encouragement and wonderful guidance by Dr. Nam and Dr. Gibbon are also highly acknowledged. Thank you all for your willingness to serve and guidance.

Dr. Diana Wycuff deserves a special mention for providing an extended help and I am incredibly grateful to her for very constructive technical discussions during the project. I also express my sincere thanks to Statistics department of Baylor University which helped me in analyzing the data.

I am greatly indebted to the Institute of Biomedical Studies, Baylor University for giving me once in lifetime opportunity to do my Ph.D. and supporting financially in the form of stipends as "Teaching assistant" and "Research assistant". During Ph.D., I also got an opportunity to present my research work in some of the prestigious conferences in *C. elegans* field. Travel assistance from the department in the form of "Graduate student travel award" is highly endorsed.

Many thanks go in particular to Dr. Bob Kane who has been a continuous source of encouragement for me and made my stay very comfortable and homely in Baylor University.

The gratis samples of *C. elegans* strains from Caenorhabditis Genetics Center, University of Minnesota (which is funded by the NIH National Center for Research Resources) are highly acknowledged.

My dear colleagues, Jonathan, Sharath, Xiaojian, Haisu, and Erich, deserve a special admiration for their kind help, fruitful discussions and for creating a cordial and lively atmosphere in laboratory. The unremitting help from Linda is also highly admired.

In the last, my loving family deserves a very special mention whose love, inspiration and unconditional support have brought me to this stage. Words are not enough to express my gratitude for my family. I owe a very special thanks to my wife; she made my life much easier with her persistent help and encouragement.

DEDICATION

To my family and friends,

Thanks for being there

CHAPTER ONE

Introduction

In post genomic era, researchers have characterized the functions of many genes discovered by multiple sequencing projects. Within a span of few years, scientists have gained a significant amount of information about the genetic makeup of organisms. Although sequencing of many genomes has been completed, still there exists a wide information gap between the known genome sequence and its protein function. Complete genome sequence of many organisms provides a new platform to understand the basic genetic makeup of human beings and the effects of mutations in genes on diseases or other variations.

Functional genomics is the branch of genomics which enables us to use the vast amount of data produced by genomic objects to study gene, its protein products and their interactions (Arenz 2003). *Haemophilus influenzae Rd* was the first genome sequence published in 1995 (Fleischmann *et al*.1995). Since then, genomes of more than 180 organisms have been sequenced. A major breakthrough was sequencing of human genome published in 2001 by two independent initiatives (Lander 2001; Venter *et al*. 2001). Human genome sequencing project resulted in 30,000 genes which was a very small number than anticipated. However, due to alternative splicing and other post transcriptional modifications, the total number of proteins encoded by those 30,000 genes is approximately 100,000 (Arenz 2003).

Genome sequence from a single organism does not disclose much by itself if not studied further. Hence, scientists have been comparing different sequenced genomes to other species or subspecies to get more comprehensive information. As a result of this comparison, many similarities have been found between genome of human and other species. In 1998, whole genome of *C. elegans* was sequenced and a striking similarity was observed with human genome. Due to this remarkable similarity, researchers have extensively studied the *C. elegans* genome. Complete genome of *C. elegans* consists of 19,920 genes whereas only 2807 (14 % of total genome) genes are reported to be associated with RNAi phenotypes.

Although there is a significant progress in finding the *C. elegans* genes' functions but still extensive work needs to be done. Assigning phenotypes to remaining 86% genome in laboratory is an enormous task and will require large amount of time and money. Hence, researchers have developed new techniques such as Basic Local Alignment Search Tool, BLAST (Altschul *et al*.1990; Altschul *et al*.1997) to fasten the process of finding functions of remaining genes based on the sequence similarity.

In this study, we attempted to assign RNAi phenotypes to new genes from remaining 86% of *C. elegans* genome on the basis of sequence similarity with known RNAi phenotypes genes using BLAST as a bioinformatics tool. As an outcome of this study, we assigned functions to some new genes and moved a bit closer to understand the biology of *C. elegans*.

Objectives

One fundamental paradigm in computational biology is function prediction by sequence similarity (Sjolander 2004). As per this paradigm, if the gene with unknown function has a high sequence similarity with a gene having known function then the function of unknown gene is inferred based on the known. These predictions are used to gain first-order approximation of the molecular function and to prioritize experimental investigation.

Based on this paradigm, main objective of this project was to identify RNAi phenotypes in *C. elegans* on the basis of sequence similarity using bioinformatics techniques. The project was envisaged with following specific aims and objectives:

1. Identification of RNAi phenotypes by systematical screening of *C. elegans* genes on the basis of sequence similarity with the help of Basic Local Alignment Search Tool (BLAST) as computational tool.
2. Association of sterility (Ste) and uncoordinated movement (Unc) phenotypes in new genes identified by BLAST using RNA interference.

Similarities in Human and C. elegans Genome

After sequencing of whole genome of *C. elegans*, it was observed that about 60% of its genes have similarity to a human gene (Harris *et al*. 2004). Also, approximately 40% of human disease causing genes have homologues in the *C. elegans* genome (Culetto and Sattelle 2000).

Few human disease genes having sequence similarity in *C. elegans* are described in Table 1.1 and complete list is given in Appendix A. First column tells the name of human disease with Online Mendelian Inheritance in Man (OMIM) accession number.

OMIM is the database of human genes and genetic disorders. In the second column, abbreviation of the human protein is given together with its GenBank accession number. Third column describes the *C. elegans* closest open reading frame (ORF) which is based on the probability value or P-value. P-value can be described as the probability of obtaining, by chance, a pairwise sequence comparison of the observed similarity given the length of the query sequence and size of the database searched (Pevsner 2003).

Parkinson disease juvenile 2 which results in neural degeneration, is an autosomal recessive disease caused by mutations in the protein parkin. Human gene has N terminal domain which has 60% similarity to *C. elegans* gene K08E3.7. Also, at the C-terminus, really interesting new gene (RING) finger motif is completely conserved between human and *C. elegans* gene (Culetto and Sattelle 2000; Harris *et al*. 2003).

Polycystic kidney disease is another disorder caused by mutation in two proteins polycystin-1 (PKD-1) and polycystin-2 (PKD-2). It has been found that *C. elegans* gene *lov-1* is orthologue to PKD-1 and PKD-2 (Barr and Sternberg 1999; Ponting *et al*. 1999).

Another *C. elegans* gene R12B2.1 encodes a human homologue of Deleted in Pancreatic Cancer, locus 4 (DPC4), which when mutated leads to pancreatic carcinoma. R12B2.1 is expressed in nerves of head, in the pseudocoelom and also in the posterior gut (Reece-Hoyes *et al*. 2007).

Table 1.1. *C. elegans* homologues of positionally cloned genes mutated in human diseases. Few of them are described here and remaining in Appendix A.

Human Disease (OMIM Accession no.)	Human Gene (GenBank accession no.)	*C. elegans* ORF	BLASTp (P value)
Alzheimer disease, type 3 (104311)	PS1 (L76517)	F35H12.3	2.4×10^{-79}
Alzheimer disease, type 4 (600759)	PS2 (L44577)	F35H12.3	1.4×10^{-81}
Breast and ovarian cancer, early onset (113705)	BRCA1 (U14680)	C36A4.8	7.5×10^{-11}
Cystic fibrosis (219700)	CFTR (M28668)	F21G4.2	1.4×10^{-135}
Duchenne muscular dystrophy (310200)	DMD (M18533)	F38B4.3	1.2×10^{-145}
Hypophosphataemic rickets, X-linked (307800)	XLH (U60475)	F18A12.8	2.6×10^{-91}
Kallmann syndrome (308700)	KAL (M97252)	K03D.10.1	7.1×10^{-34}
Pancreatic carcinoma (260350)	DPC4 (U4437)	R12B2.1	7.7×10^{-77}
Parkinson disease juvenile 2 (600116)	Parkin (AB009973)	K08E3.7	6.1×10^{-42}
Polycystic kidney disease, type 1 (173900)	PKD1 (L33243)	ZK945.9	2.8×10^{-06}
Retinoblastoma (180200)	RB1 (M15400)	C32F10.2	1.3×10^{-10}
Spinal muscular atrophy (253300)	SMN (U18423)	C41G7.1	3.0×10^{-08}

Spinal muscular atrophy (SMA) is an autosomal recessive disorder resulting in muscle weakness and wasting and mutation in survival motor neuron (SMN) is responsible for this disease (Culetto and Sattelle 2000). *C. elegans* gene C41G7.1 is orthologue of human SMN and it is found on chromosome number one (Talbot *et al*. 1997). Both the human and *C. elegans* genes; fusion construct is expressed in the neurons, body wall, vulval muscle cells, hypodermal cells and gut cells (Miguel-Aliaga *et al*. 1999; Culetto and Sattelle 2000).

Thirty three human disease genes are homologues with an RNAi phenotype predicted genes in *C. elegans*. Few of them are described in Table 1.2 and full list is given in Appendix B. First and second column contain *C. elegans* genes name and locus respectively. Third and fourth column contain human disease name and gene name respectively. Fifth is the Expect value (E-value) score obtained by BLAST search, whereas last column contains RNAi phenotype in *C. elegans*. E-value describes the likelihood that a sequence with same score will occur in the database by chance. The smaller the E-value, more significant is the alignment (Pevsner 2003).

Poulin *et al*. (Poulin *et al*. 2004) made a list of 61 human cancer genes and searched for homologues and strikingly 80% have a putative worm homologue. This validates *C. elegans* as a choice of model organism in investigating biological processes as well as elucidating the pathogenesis of diseases.

Due to this striking similarity in between human and *C. elegans* genome researchers have been studying the worm extensively. There is a progress in finding the *C. elegans* genes' functions but still extensive work needs to be done.

Table 1.2. Thirty three human disease gene homologues with an RNAi phenotype (Rubin *et al*. 2000; Wood 2002; Kamath *et al*. 2003). Few of them are described here and remaining in Appendix B.

Predicted Gene	Locus of *C. elegans*	Human Disease	Human Gene	BlastP (E value)	RNAi Phenotype
B0035.5 (Kamath *et al*. 2003)	-	G6PD deficiency	G6PD	1×10^{-176}	Emb, Clr, Gro
B0350.2A (Simmer *et al*. 2003)	*Unc-44*	Hereditary spherocytosis	ANK1	0.00	Slu, Unc
C01G6.8 (Kamath *et al*. 2003)	*cam-1/kin-8*	Insulin-resistant diabetes mellitus	INSR	6×10^{-55}	Unc, Pvl, clear patch
C01G8.5A (Fraser *et al*. 2000)	-	Neurofibromatosis	NF2	1×10^{-123}	Unc, Lvl, Gro
C06A1.1 (Gottschalk *et al*. 2005)	-	Zellweger syndrome	PEX1	3×10^{-67}	Emb, Bmd, Sck, Gro
C07H6.7 (Kamath *et al*. 2003)	*lin-39*	MODY, type IV	IPF1	5×10^{-14}	Egl, Vul, Muv
C17E4.5 (Simmer *et al*. 2003)	-	Oculopharyngeal muscular dystrophy	PABPN1	3×10^{-41}	Emb, Unc, Lva
C29A12.3 (Kamath and Ahringer 2003)	*lig-1*	DNA ligase I deficiency	DNA ligase1	1×10^{-167}	Emb

C. elegans as a Genetic Model

The name *Caenorhabditis elegans* is a blend of Greek and Latin (*Caeno* means recent in Greek; *rhabditis* means rod in Greek; and *elegans* means nice in Latin). In 1974, Sydney Brenner documented the use of *C. elegans* as a model genetic organism (Brenner 1974). Since then thousands of research articles have been published to investigate all aspects of this organism.

C. elegans is found in many parts of the world and it survives by feeding on microbes mainly bacteria. The worm has great potential for genetic, molecular and cellular analysis. Below mentioned features of *C. elegans* make it an increasingly useful experimental organism for investigation of various pathological conditions (Brenner 1974; Byerly *et al*. 1976; Sulston *et al*. 1983; Wood 1988a; Hope 2000).

1. Short life span of approximately 17-18 days.
2. Short reproduction life cycle of about 3 days under normal conditions.
3. Small size of approximately 1.2 mm.
4. Amenability to genetic crosses and produces a large number of progeny per adult.
5. Ease of cultivation in the laboratory.
6. Transparent stages of life cycle, which make it very useful to study various cellular events and processes like fat metabolism, axon growth and embryogenesis with the help of *in vivo* fluorescence markers.
7. Approximately 1000 somatic cells in adult worm which make various tissue types such as muscles, intestinal cells, and neurons etc.
8. Association of worm with easily observable mutant phenotypes (sterility, uncoordinated, dumpy, growth, blister, lethality, multivulva etc).

9. Responsive to various stimuli including touch, temperature change and to the different chemicals and ions.
10. Complete anatomy at the electron microscopic level, the locations and characteristics of all somatic cells, and the complete cell lineage are known in *C. elegans.*

C. elegans life cycle (Figure 1.1) is comprised of the embryonic stage, four larval stages (L1, L2, L3 and L4) and adulthood. The nematode transforms from L1 to L2 stage in approximately 12 hours, L2 to L3 and L3 to L4 each in 8 hours, and from L4 to adult worm in 18 hours. At the end of each cycle, new stage specific cuticle is synthesized and the old one is shed (Cassada and Russell 1975). There is another stage in *C. elegans* life cycle known as dauer larva (L2d). The growth of worm is arrested at this stage. It comes after L2 stage if certain conditions are not favorable to worm such as lack of food, more population density, and high temperatures. The worm can start growing again into L4 stage when the conditions become favorable (Riddle 1988). This dauer stage is non-aging stage because its duration does not cause any affect post dauer life span.

C. elegans has two sexes (Figure 1.2) hermaphrodite (XX) and male (XO) which are essential for genetic crosses. Hermaphrodites have 959 somatic cells and males have 1031 cells (Podbilewicz 2006). Hermaphrodites produce both oocytes and sperm and can produce off springs by self-fertilization. Adult hermaphrodites have embryos in their uterus and eggs are laid through a vulva.

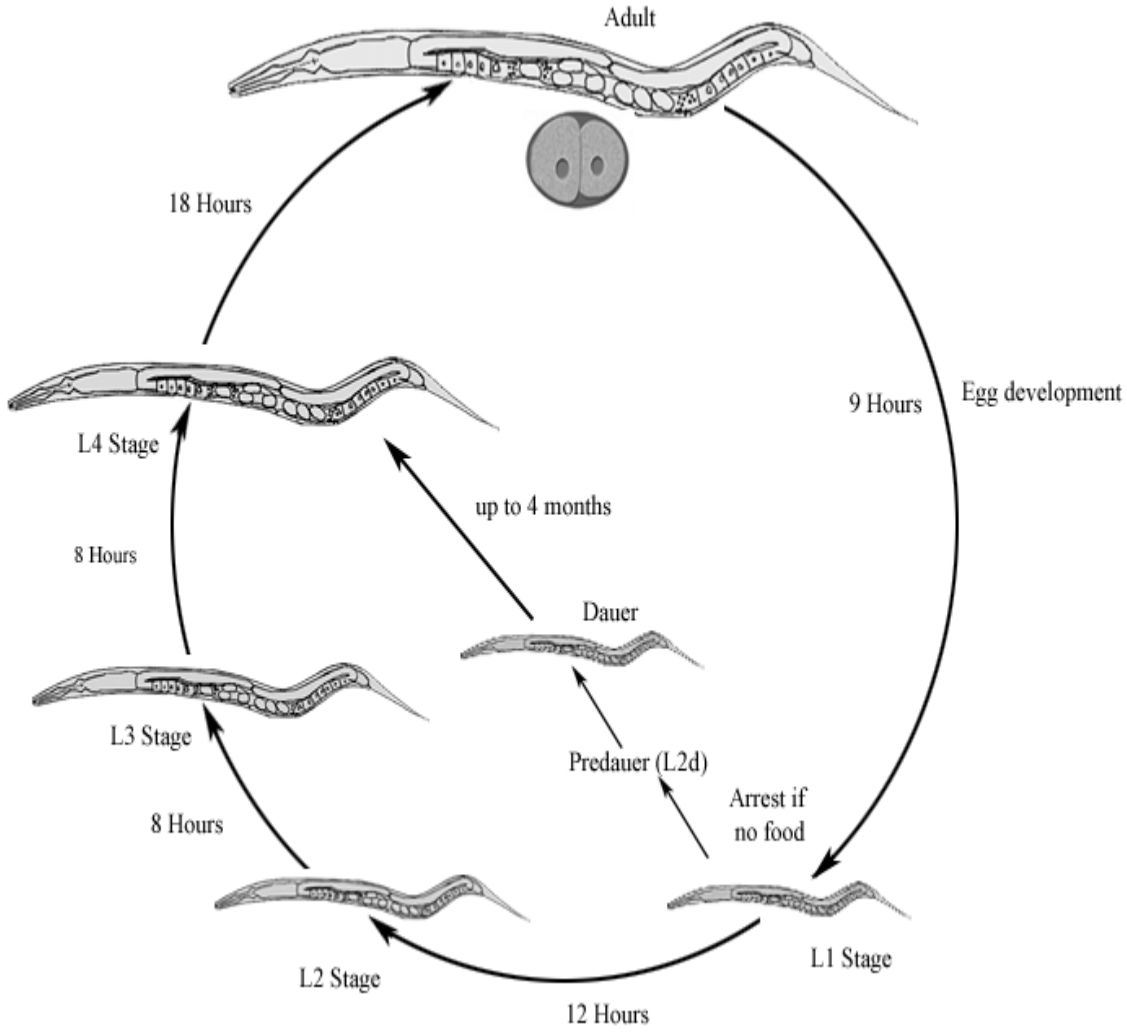

Figure 1.1. *C. elegans* life cycle. Life cycle includes a) embryonic stage (egg development), b) four main larval stages (L1, L2, L3 and L4), and c) adult worm. Another stage, called dauer stage appears only in adverse conditions.

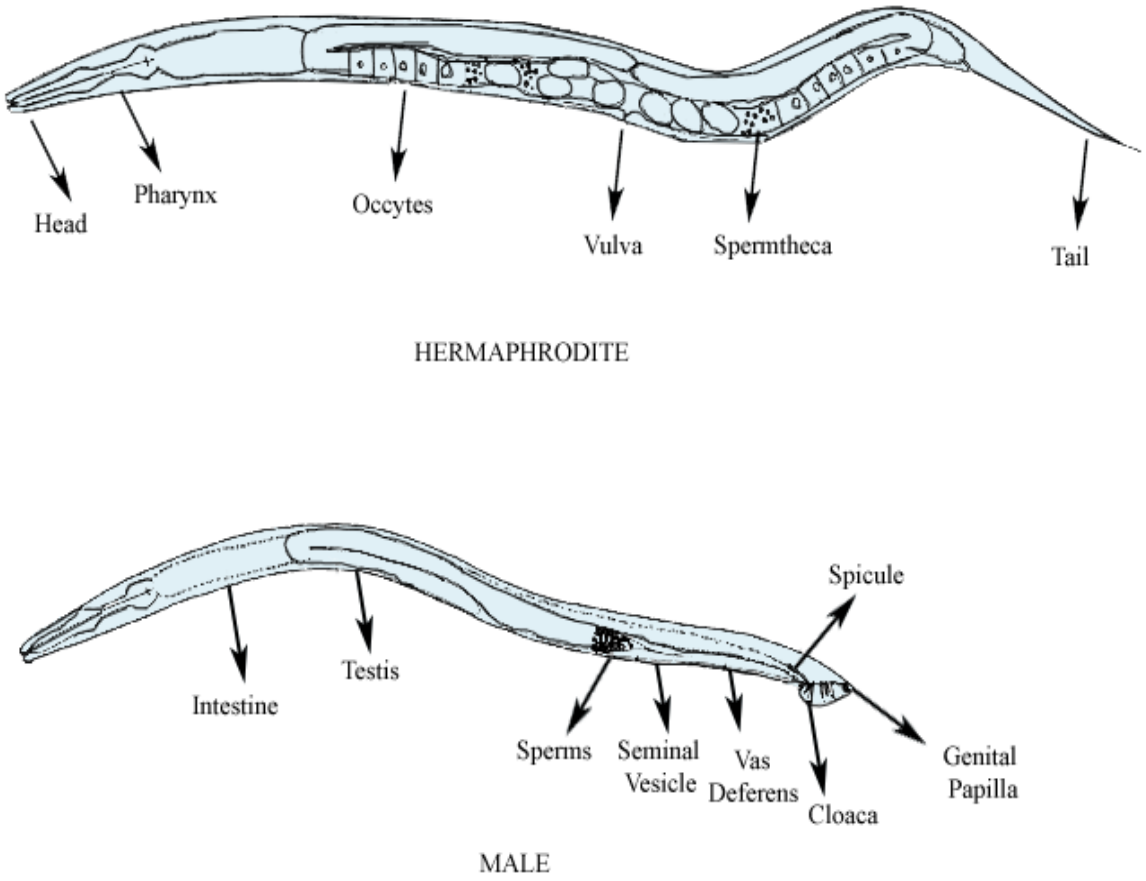

Figure 1.2. Schematic diagram of *C. elegans* hermaphrodite and male.

Males can fertilize hermaphrodites and they are very rare (0.1-0.2%) in population. Self-fertilization facilitates homozygous worms to reproduce genetically identical progeny and mating with males help in the isolation and maintenance of mutant strains (Anderson 1995; Jorgensen and Mango 2002). A hermaphrodite can lay up to 300 fertilized eggs from self-fertilization because of the limited number of sperm. If mating occurs, the progeny number can go beyond 1200 (Hodgkin 1988). Males are thinner and smaller than hermaphrodites and have specific structure in their tails which is necessary for mating.

The nervous system of *C. elegans* is composed of 302 neurons and 56 glial cells make up 37% of the somatic cells in a hermaphrodite (Hobert *et al*. 2005). Their main function is to distribute nerves along the body wall muscles and case them to act, resulting in a sinusoidal movement. This nervous system is responsible for variety of behaviors that can be observed very easily such as movement towards food and other attractive odorants and away from undesirable chemicals.

Non-mammalian model organisms are being used in research to unravel the functions of genes or to find out novel therapeutic entry points (Kaletta and Hengartner 2006). However, due to all above mentioned reasons *C. elegans* has been used extensively to understand the underlying mechanisms of human diseases and it is one of the cost effective and fastest animal models to do biomedical research.

There are few examples of key biomedical discoveries enabled by *C. elegans* research. For example in diabetes type 2, genetic studies in *C. elegans* helped in identifying negative regulators of the insulin signaling pathway. One of the gene, *daf-16*, encodes the *C. elegans* orthologue of the forkhead transcription factor, sub-group O (FOXO) (Ogg *et al*.1997). Few years later FOXO loss of function was found to rescue the diabetic phenotype of insulin-resistant mice (Nakae *et al*. 2002). Another good example would be Alzheimer's disease (AD). In 1993, first presenilin was discovered in *C. elegans* (Sundaram and Greenwald 1993). Few years later it was found that human presenilin-1 gene was associated with familial AD (Levitan and Greenwald 1995; Sherrington *et al*. 1995). Researchers have expressed human presenilin in *C. elegans* which rescued neuronal deficiencies of *C. elegans sel-12* presenilin mutants (Levitan *et al*. 1996; Wittenburg *et al*. 2000).

The biomedical discoveries show the remarkable functional conservation between *C. elegans* and human genomes. Due to this striking similarity with human genes and its simplicity, researchers have been trying to find the functions of each and every gene in *C. elegans*. Many techniques have been used to find the genes' functions in different species like knock out mice, antisense oligomers, aptamers and ribozymes but these approaches have been superseded by RNA interference (RNAi) (Arenz 2003).

RNA Interference (RNAi)

Before introduction of RNA interference (RNAi), forward genetic approaches were used to identify gene functions in *C. elegans*. RNA interference, a reverse genetics tool, has made *C. elegans* more useful model organism in studying the functions of genes. RNAi is a mechanism to silence the target gene by introduction of double-stranded RNA (dsRNA) which degrades the homologous endogenous mRNA. RNAi is now a well established technique to find the functions of genes. Figure 1.3 describes a working model of RNAi. RNase–III processing enzyme-DICER processes the dsRNA into small interfering (siRNA) duplexes which are approximately 21 nucleotides in length. One strand of this duplex assembles into the RNA-induced silencing complex (RISC) which acts on its target by mRNA cleavage (Hannon 2004).

There is another source of siRNA production other than the exogenous dsRNA and which is cleaved by DICER (Sijen *et al*. 2001) as shown in Figure 1.3. A second group of siRNA has biological characteristics that indicate the use of an RNA-dependent RNA polymerase (RdRP) enzyme involved in RNA virus replication by synthesizing complementary RNA as a template (Arenz 2003).

RdRP is assumed to convert the single-stranded mRNA to dsRNA using the antisense strands of primary siRNA as primers (Lipardi *et al*. 2001; Sijen *et al*. 2001). However, RdRPs have not been found in flies and humans which suggest that there are different mechanisms of RNAi in different species (Arenz 2003).

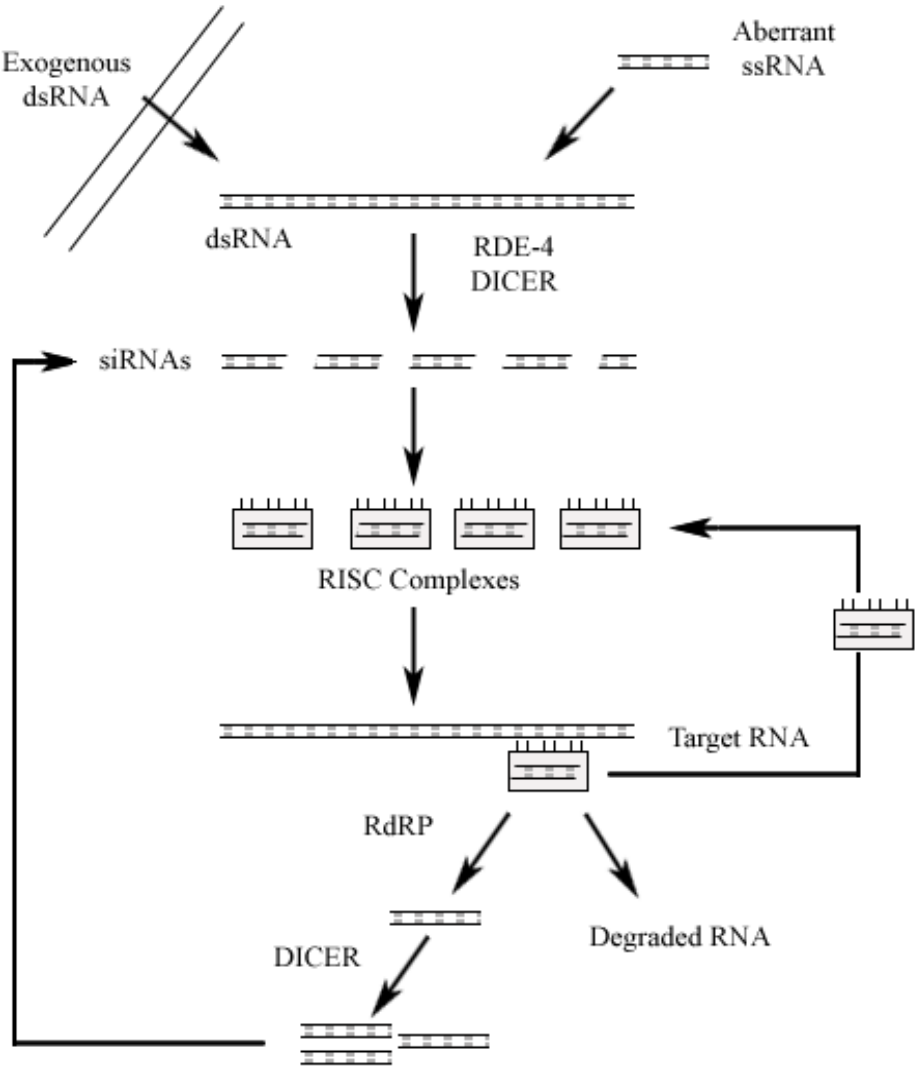

Figure 1.3. Working model of RNAi showing two sources of siRNA. One is from exogenous dsRNA which is cleaved by DICER and other source is through RdRP (Sijen *et al*. 2001).

RNAi was first used in *C. elegans* (Fire *et al*. 1998) but it has also been shown effective in other organisms such as fungi, protozoa, mammalian cells, and planarians and it is known by different names in various species (Table 1.3).

Table 1.3. Gene silencing mechanism across taxa.

Species	Organism	Phenomenon
Fungi	*Neurospora*	Quelling (Cogoni and Macino 1997, 1999)
Plants	*Arabidopsis Nicotiana*	Post transcriptional gene silencing, Co-suppression (Dehio and Schell 1994; Elmayan *et al*. 1998)
	Petunia	Transcriptional gene silencing (Furner *et al*. 1998; Mittelsten Scheid *et al*. 1998)
Invertebrates	*Paramecium*	Homology dependent gene silencing (Ruiz *et al*. 1998)
	Hydra	RNAi (Lohmann *et al*. 1999)
	T. brucei	RNAi (Ngo *et al*. 1998)
	Planaria	RNAi (Sanchez Alvarado and Newmark 1999)
	Drosophila	RNAi, Co-suppression (Kennerdell and Carthew 1998; Misquitta and Paterson 1999)
	C. elegans	RNAi, Gene silencing (Fire *et al*. 1998)
Vertebrates	Zebrafish	RNAi (Wargelius *et al*. 1999)
	Mouse	RNAi (Wianny and Zernicka-Goetz 2000)

Methods of Administration of RNAi

There are mainly three methods to carry out RNAi in *C. elegans* as mentioned below. All three methods can efficiently produce gene knock downs and have their own advantages and disadvantages.

1) Injection (Fire *et al*. 1998)
2) Soaking (Tabara *et al*. 1998)
3) Feeding (Timmons and Fire 1998)

For all the three methods, a plasmid having the target gene or your favorite gene (YFG) needs to be constructed (Figure 1.4). The insert can be cloned into the mostly used L4440 vector (Timmons and Fire 1998), which contains T7 promoter regions on both sides of the cloning site or T3/T7 promoters that surround a multiple cloning site. If the plasmid has only T7 promoter sites, it can be transformed into HT115, which is an RNase III deficient strain of *Escherichia coli* (*E. coli*) bacteria (Timmons *et al*. 2001) that can induce T7 polymerase in the presence of isopropyl-beta-D-thiogalactopyranoside (IPTG) or lactose (Timmons *et al*. 2001). Due to genome-wide RNAi studies, many clones of *C. elegans* genes in the L4440 vector exist in the RNAi libraries that are available for purchase (Fraser *et al*. 2000; Kamath *et al*. 2003). All three methods have been used in genome wide RNAi phenotype analysis in *C. elegans.*

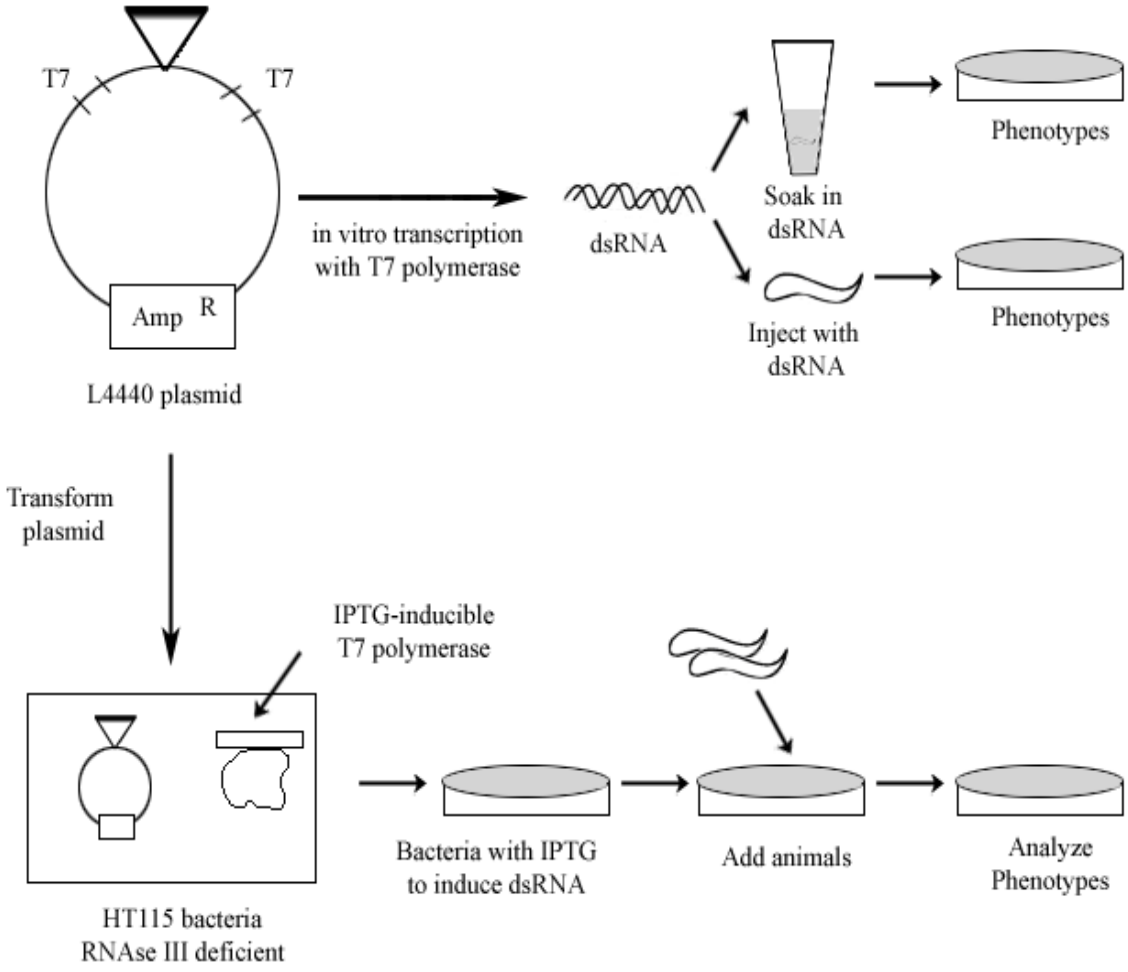

Figure 1.4. Schematic diagram of methods used for RNAi in *C. elegans.* A plasmid containing T7 promoters with target gene is constructed. Top part of the figure show the use of dsRNA as in soaking and injecting the animals. Bottom part shows how animals can be fed bacteria producing dsRNA (Corsi 2006).

RNAi by Injection

Initially, RNAi in *C. elegans* was performed by injection method. In this method, dsRNA is produced *in vitro* and is injected in to young adult hermaphrodites and their progeny is examined for a phenotype. It can be used at any site in the *C. elegans*. Injection method is more labor-intensive than soaking or feeding methods. Nevertheless, (Gonczy *et al*. 2000) analyzed 96% of the genes on Chromosome III of *C. elegans* required for cell division by using injection method and identified 133 genes required for different cellular processes in early embryos.

RNAi by Soaking

In this method, worms are soaked in a dsRNA solution and then they or their progeny are scored for phenotypes. This method was first developed by (Tabara *et al*. 1998) and then improved by Maeda *et al* (Maeda *et al*. 2001). This method is useful for large number of worms or for high throughput screening (Maeda *et al*. 2001). This method can be used at any stage of the worms but more dsRNA are required in this method than for injection. In this method, L1 larvae are soaked in the dsRNA solution, the RNAi effect starts later in development. Therefore, it can help in analyzing the post embryonic function of the genes which is not feasible with injection method as injection is difficult in larvae of *C. elegans*.

RNAi by Feeding

In feeding method, bacteria producing desired dsRNA are fed to worms and subsequently they or their progeny can be scored for phenotypes. Feeding method is easy to use on a large number of worms and less expensive than other two methods.

RNAi by feeding has similar efficiency in generating phenotypes as compared to injection or soaking but penetrance for several kinds of phenotypes is more variable. Worms of any stage can be used in the feeding method. A bacterial library has been constructed of approximately 86% of 19,000 predicted genes in *C. elegans* for easy use of feeding method (Kamath and Ahringer 2003). A comparison of all three methods is given in Table 1.4.

Table 1.4. Comparison of three methods of RNAi.

RNAi	Injection	Soaking	Feeding
Method	dsRNA produced *in vitro* and injected into adult hermaphrodites	Worms soaked in a dsRNA solution and then they or their progeny is scored for phenotypes	Bacteria producing desired dsRNA are fed to worms and subsequently they or their progeny can be scored for phenotypes
Site/Stage	Used at any site in *C. elegans*	Used at any stage of the worms	Used at any stage of the worms
Advantage	Strong potency	Useful for large number of worms or for high throughput screening	Easy to use on a large number of worms and less expensive
Disadvantage	Time consuming and labor intensive	More dsRNA required than injection method	Slightly more variable results

RNAi phenotypes are categorized into four main classes such as maternal, embryonic, general postembryonic and specific post embryonic phenotypes described in Table 1.5.

Table 1.5. Categorization of RNAi phenotypes (Wormbase.org 2005, release WS159).

Class	Examples of Phenotypes	
Maternal Phenotype	Partly or Completely Sterile	
Embryonic Phenotypes	Embryonic lethal (Emb)	
General Post Embryonic Phenotypes	Growth defect Morphology defect Movement defect Reproductive defect	
Specific Post Embryonic Phenotypes	Blistered (Bli) Egg laying defect (Egl) Hyper Active (Hya) Molt defect (Mlt) Protruding vulva (Pvu) Sterile (Ste) Body morphology defect (Bmd) Feminization of XX and XO animals (Fem) Larval lethal (Let) MultiVulva (Muv) Roller (Rol) Sterile progeny (Stp) Clear (Clr) Feminization of germline (Fog)	Long body (Lon) Paralyzed (Prl) Exploded (Rup) Uncoordinated (Unc) Dauer Formation (Daf) Slow growth (Gro) Larval arrest (Lva) Sick (Sck) Vulvaless (Vul) Dumpy (Dpy) High incidence of male progeny (Him) Small (Sma)

The two widely occurring phenotypes in *C. elegans* are uncoordinated movement and sterility. As the name suggests, in sterility phenotype *C. elegans* fail to reproduce due to various reasons and uncoordinated movement is a phenotype where the worms have abnormal style of moving and bending their body.

Sterility

Sterility is the common problem worldwide. In some countries people suffering from sterility feel isolated from their community (Adekunle 2002). Sterility can cause clinical depression, stress and other psychological problems. In the United States alone 12 % of women have difficulties in naturally conceiving (Dye 2005). Sterility in humans can be caused by female factors such as ovarian, uterine and vaginal diseases or by male factors like pretesticular, testicular, and post testicular. It can also be caused by environmental and genetic factors. Due to high similarity between human and *C. elegans* genome it is worth studying the sterile genes in *C. elegans*. For example, by studying sterility genes useful information regarding the fertility of *C. elegans* has been found. The importance of IP3 receptor, calcium channel in regulating the contraction of *C. elegans* gonad was also observed (X. Xu 2005).

There could be many reasons for sterility phenotype (as described in Table 1.6). L'Hernault *et al.* and SchedlT *et al.* (L'Hernault. 1997; SchedlT 1997) have categorized the causes of sterility into four main reasons.

Table 1.6. Criteria for categorization of sterility in *C. elegans*.

Types of Sterility	Criteria
Sheath cell contraction defects	Rate of contractions, initiation of contractions, strength of contractions differ significantly from wild-type
Spermathecal defects	Initiation of spermathecal widening differs significantly from wild-type, premature closing of spermatheca, non-functioning
Oocyte developmental defects	Oocyte production, maturation events differ significantly from wild-type
Sperm production defects	Sperm production, sperm location during fertilization differ significantly from wild-type

C. elegans Gonad

There are two U shaped tubular arms in the adult hermaphrodite. Each tubular arm connects the proximal end to a spermatheca (Figure 1.5). The U-shaped tubes function as the ovo-testis which produces oocytes as well as sperms. In the distal gonad, oocyte precursor cells proliferate and undergo series of mitotic and meiotic maturation. In the proximal gonad, oocytes line up in a single row and are encapsulated by sheath cells.

C. elegans has five pairs of sheath cells. First and second pairs encapsulate the distal gonad and help to maintain its structural integrity. The third, fourth and fifth pair are localized in the proximal gonad (Rose *et al*. 1997) and these pairs maintain its shape and contraction (Hall *et al*. 1999; Hubbard and Greenstein 2000). The proximal gonad produces contractile forces along the distal to proximal axis, which helps to propel oocytes into spermatheca during ovulation (Strome 1986).

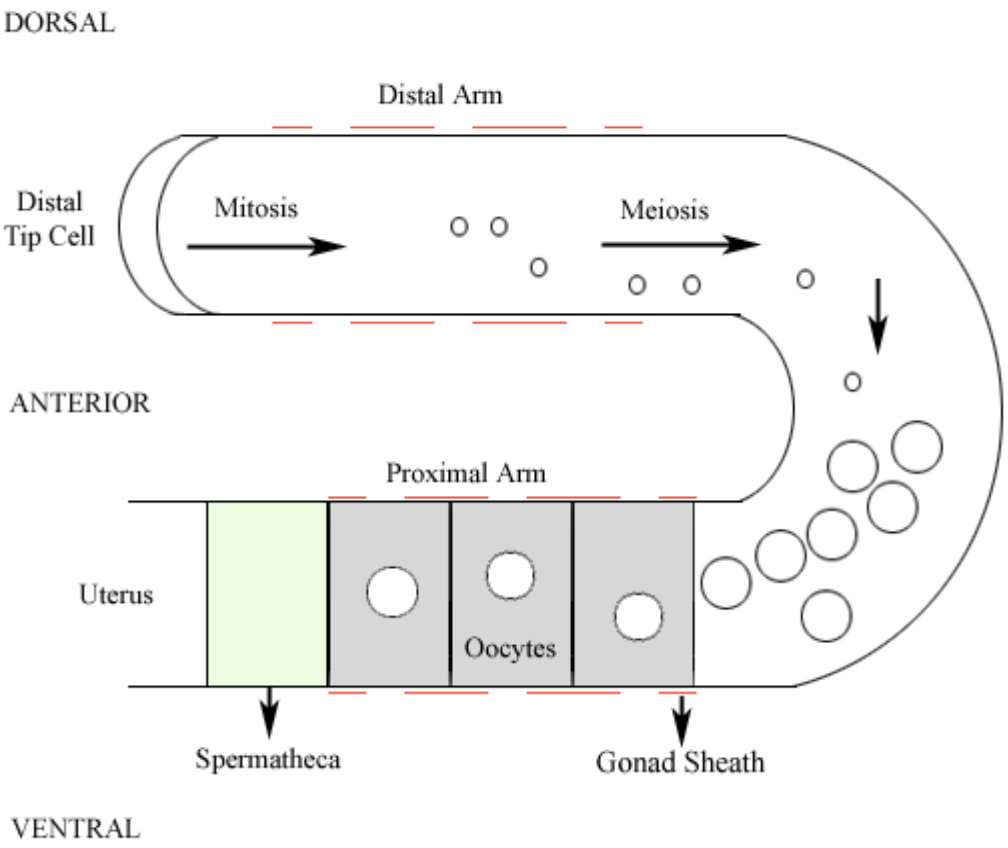

Figure 1.5. Gonad structure of *C. elegans*. The figure represents a posterior U-shaped gonad arm. Near the distal tip cell, germ cells undergo mitosis than become meiotic. In the proximal arm, matured oocytes are arranged in a row. One by one all the oocytes undergo maturation and advance to spermatheca for fertilization.

As described earlier (Table 1.6) there could be various causes of sterility phenotype in *C. elegans*. One such cause of sterility is shown in Figure 1.6 where failure of normal oocytes to undergo meiotic division occur and subsequent DNA replication leads to the formation of Emo oocytes in the gonad arm (Iwasaki *et al*. 1996).

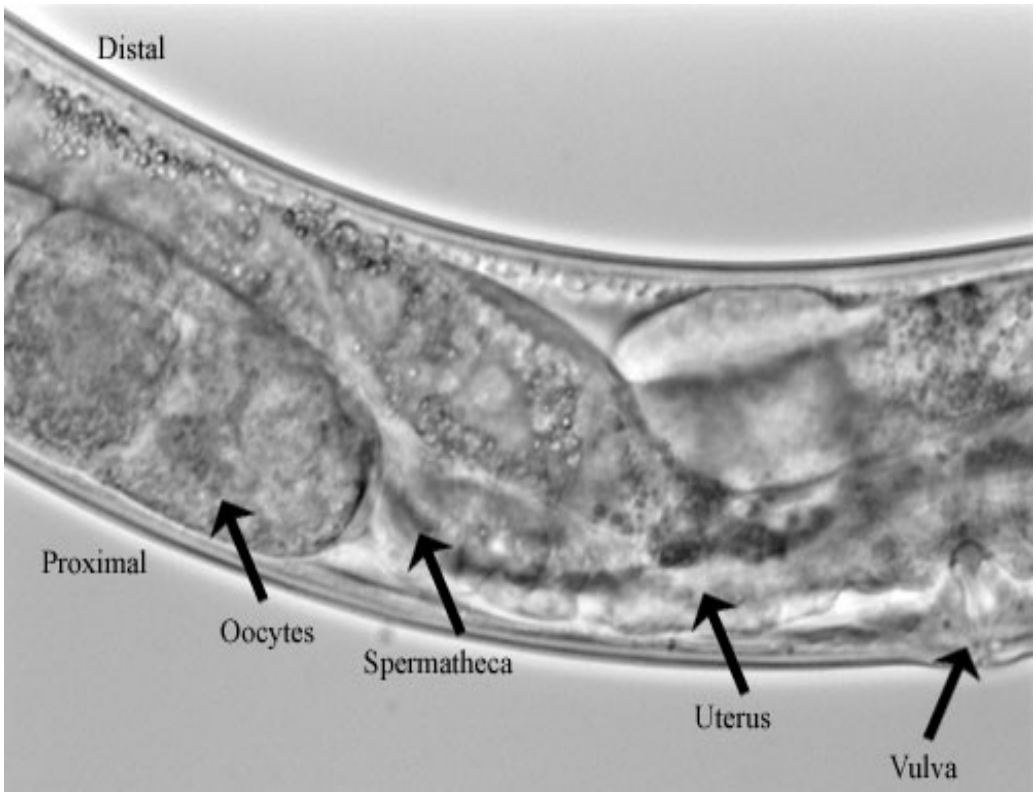

Figure 1.6. Sterility phenotype in *C. elegans*. In this picture, eggs are not moved into spermatheca either due to defects in sheath cell contraction or spermathecal defects.

Uncoordinated Movement

Previous studies have shown that this is one of the largest classes of post embryonic phenotype (Fraser *et al*. 2000; Kamath *et al*. 2003). Bargmann *et al*. (Bargmann 1998) summarized that Unc phenotypes arise from defects in the development or function of the neuromuscular system (Table 1.7).

Table 1.7. Genes that affect nervous system function.
PLC, phospholipase C; PDE, phosphodiesterase; ChAT, choline acetyltransferase; AChE, acetylcholinesterase; VChAT, vesicular ChAT.

Neurotransmitter Receptors, Ion channels, Transporters and Pathways	Approximate Number of *C. elegans* Genes
Voltage-regulated calcium channels (Lee RY 1997 ; Goodman *et al*. 1998)	5 α1, 2 α2, 2β
Potassium channels (Wei *et al*. 1996)	20 6TM (10 Shak/Shaw/Shab/ Shal, 3 KQT, 2 EAG, 2 SLO, 4 SK), 3 2TM (IRK), 40+ TWIK
Chloride channels (CLC) (Caterina *et al*. 1997; Colbert *et al*. 1997)	6
Classic neurotransmitter synthesis, degradation (Johnson *et al*. 1981; Arpagaus *et al*. 1994; McIntire *et al*. 1997)	1 each ChAT, GAD, tyrosine hydroxylase, etc., 4 AChE
Neurotransmitter transporters (Stretton Aow 1985; Walrond *et al*. 1985)	1 each VChAT, VMAT, one outlier 12 GABA transporters/amino acid permeases, 6 EAAT transporters
Neuropeptides (Stretton *et al*. 1991; Cowden *et al*. 1993)	15 FMRF-amide related, 15 other
Neurotransmitter release/exocytosis (Maruyama and Brenner 1991; Gengyo-Ando *et al*. 1993; Jorgensen *et al*. 1995; Iwasaki *et al*. 1997)	1 highly conserved each and 3 to 7 additional candidate synaptobrevin, syntaxin, synaptotagmin, *unc-18*, SNAP25; latrophilin; 1 rab3
Ligand-gated ion channels (Walrond and Stretton 1985; Cully *et al*. 1994; Treinin and Chalfie 1995)	10 excitatory glutamate receptors, 42 acetylcholine receptors, 37 GABA-A and inhibitory glutamate receptors, including outliers
PDZ domain proteins (Rongo *et al*. 1998)	30+

Table 1.7 Continued.

Neurotransmitter Receptors, Ion channels, Transporters and Pathways	Approximate number of *C. elegans* genes
Heterotrimeric G proteins and targets (JE Mendel 1995; Segalat *et al*. 1995; Brundage *et al*. 1996)	20 Gα subunits, 2 Gβ subunits, 2 Gγ, 12 RGS regulators, 3 adenylyl cyclases, 8 PLC, 4 cGMP PDE
G protein-coupled receptors (Sonnhammer and Durbin 1997)	18 class A amine receptors, 50 class A peptide receptors, 4 class B peptide receptors, 4 metabotropic glutamate receptors, 3 GABA-B receptors
G protein-coupled orphan receptors (chemoreceptors) (Troemel *et al*. 1995; Yu *et al*. 1997)	700 *str* (ODR-10 related)*/stl/srd,* related groups; 150 *sra/srb/sre;* 40 *srg;* 80 class A orphan receptors
Innexin/gap junction proteins (Chalfie *et al*. 1985; Starich *et al*. 1996; Barnes and Hekimi 1997)	24 genes
Degenerin/mechanosensory proteins (Driscoll M 1991)	22 genes
Stomatin/*mec-2*-like regulatory proteins (Huang *et al*. 1995; Barnes TM 1996)	9 genes
Receptor guanylyl cyclases (Yu *et al*. 1997)	26 genes, also 5 soluble guanylyl cyclase genes
Cyclic nucleotide-regulated channels (Coburn and Bargmann 1996; Komatsu *et al*. 1996)	6 genes (plus two *eag*/*erg*, K^+ channels)
TRP-related channels (Perez-Reyes *et al*. 1998)	11 genes
CREB and regulatory pathways (Bailey *et al*. 1996; Wen JY 1997)	1 CREB, 300+ protein kinases: 2 Ca/CAM kinase, 2 protein kinase A, cGMP-dependent protein kinase

Ackley *et al* (Ackley *et al*. 2003) defined that defects in muscle cell function contribute to the observed movement defects. However, in UNC-52, staining with myosin specific antisera demonstrated no defects in the organization of thick filaments in mutant animals. Furthermore, defects in the organization of nervous system were observed, mostly synaptic. So it was concluded that uncoordinated phenotype happens due to defects in neuromuscular system. It was demonstrated that wild type animals flex around mid body and avoid over bending where as mutants with Unc phenotypes show over bending and less number of thrashing as compared to wild type (Figure 1.7).

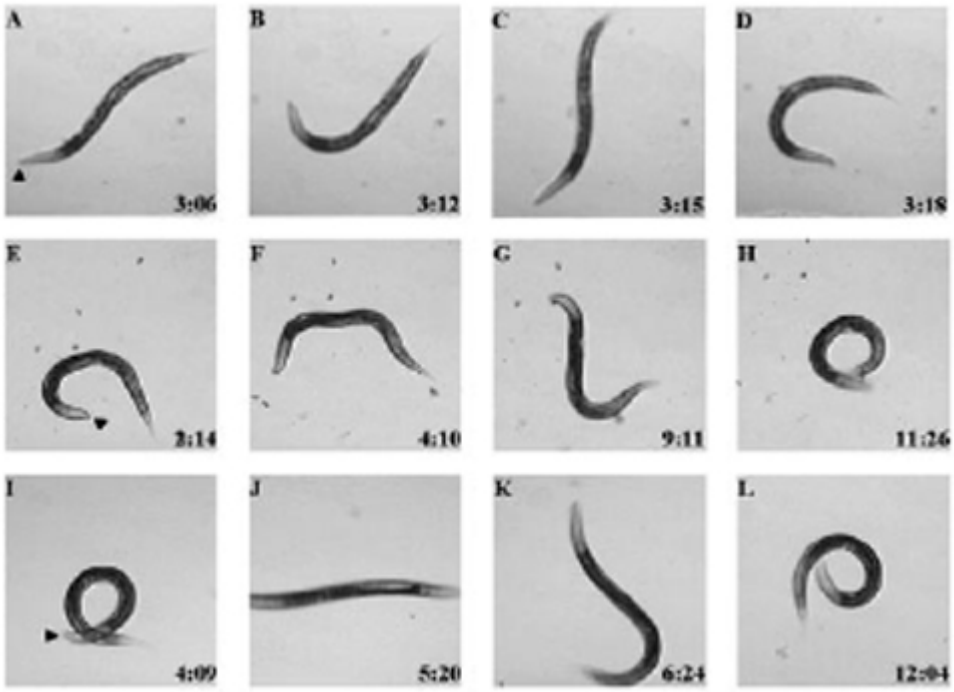

Figure 1.7. Thrashing movie stills. Ackley *et al*. (Ackley *et al*. 2003) demonstrating the uncoordinated movement of mutants with Unc phenotype. Pictures A to D show, wild type animals flex around the mid region of the body and make C-like structure. Pictures E to L show uncoordinated movement like less number of thrashing, over bending such that head and tail cross to each other.

Previous Genome Wide RNAi Screens in C. elegans

Many genome wide RNAi screenings performed on *C. elegans* (Fraser *et al*. 2000; Maeda *et al*. 2001; Kamath and Ahringer 2003; Kamath *et al*. 2003; Hansen *et al*. 2005) have increased our understanding of different functions of genes and their molecular and biochemical pathways. Some studies have shown a striking similarity of *C. elegans* and human genome (Culetto and Sattelle 2000; Harris *et al*. 2003). Cancer causing genes in humans were found to have sequence similarity with *C. elegans* genes (Poulin *et al*. 2004). Therefore, it is necessary to know the function of all genes in *C. elegans* and it could be helpful in identifying new genes' functions in other species too.

Genome-wide RNAi screenings of *C. elegans* have used different strains such as *N2* (Fraser *et al*. 2000), *rrf-3* (Simmer *et al*. 2002), *eri-1* (Kennedy *et al*. 2004) or *eri-1; lin 15B* (Wang *et al*. 2005). Mutation of *rrf-3* which is a RNA-directed RNA polymerase (RdRP) has been shown to result in increased sensitivity to RNAi (Simmer *et al*. 2002; Simmer *et al*. 2003). Using *rrf-3*, they identified phenotypes for 625 clones, which were not associated with any phenotype in previous screens with *N2* (Fraser *et al*. 2000, Kamath *et al*. 2003). Because of high sensitivity to RNAi, we have also used *rrf-3* strain in our RNAi study.

Two kinds of RNAi analyses have been done previously. One is to screen the genes involved in specific processes (Gonczy *et al*. 2000; Piano *et al*. 2000). Another is the genome wide screening to find the functions of large number of genes (Fraser *et al*. 2000; Kamath *et al*. 2003; Simmer *et al*. 2003). Major genomic RNAi studies done by various researchers and a comparison of their experimental approaches are briefly described in Table 1.8.

Table 1.8. Major RNAi studies – strategies and results.

Study	Experimental Strategy and Result
(Piano *et al*. 2000)	One of the first major genomic RNAi study using injection method and a sample size of 10 young adult *N2* worms. This study was focused on the expression of genes in the ovary and 81 new genes were identified with essential role in embryogenesis.
(Fraser *et al*. 2000)	This study is based on feeding method and targeted 90% of predicted genes on the chromosome I of *C. elegans*. *N2* worms and sample size of 10-15 worms were used in this study. Major highlight of this study was to assign functions to13.9% of the genes analyzed, thereby increasing the number of sequenced genes with known phenotypes on chromosome I from 70 to 378
(Gonczy *et al*. 2000)	This study had also used the injection method with a sample size of 6 worms (*N2*) and found 133 genes on chromosome III which are necessary for distinct cellular processes in early embryos
(Maeda *et al*. 2001)	This study established the use of spermidine in increasing the efficiency of soaking method. In this study, a sample size of 4 worms (*N2*) was used and it was found that the F1 progeny subjected to RNAi by soaking phenocopied the respective mutant phenotypes with high penetrance
(Piano *et al*. 2002)	Injection method was used in this study with a sample size of 10 worms (*N2*). RNAi was performed on 98% of 766 ovary enriched genes and showed at least one detectable phenotype in over half of them
(Kamath and Ahringer 2003)	Major highlight of this study was the construction of RNAi genome library of 86% of the *C. elegans* 19,000 genes. Feeding method was used in this study with a sample size of 3 worms (*N2*).
(Simmer *et al*. 2003)	This study established the *rrf-3* strain as a hypersensitive to RNAi. Feeding method was used with a sample size of 10 worms. This study reported in additional loss-of-function phenotypes for 393 genes, increasing the number by 23%

Figure 1.8 depicts phenotypic genes in *C. elegans* genome which is only 14% of total genome. Different techniques such as BLAST (Altschul *et al*. 1990; Altschul *et al*. 1997), and microarrays have been used to study different aspects of the genes which help in predicting genes functions. BLAST is used to study sequence similarity where as microarrays (Schena *et al*. 1995) are used to study gene expression in the cell or to compare in two different cells. In present study, one of the techniques BLAST is used to find the genes which share sequence similarity to the other known RNAi phenotype genes in *C. elegans* itself.

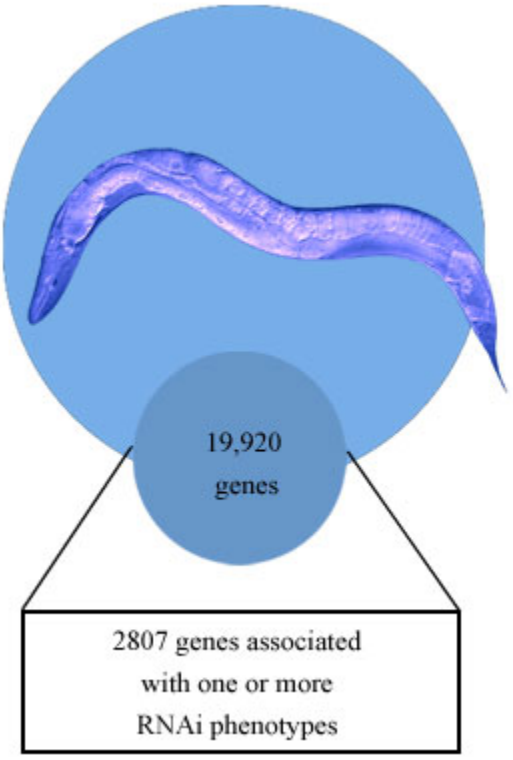

Figure 1.8. Phenotypic and non phenotypic genes in *C. elegans*.

Basic Local Alignment Search Tool (BLAST)

Computational biology or bioinformatics is used to predict the function of any gene by homology. Basically, sequence similarity is searched between a gene and protein against a known database. If a significant similarity is detected between two sequences, then the function of unknown gene is inferred on the known function of the homolog (Sjolander 2004).

BLAST is a tool for comparing protein or DNA sequence to other sequences in various databases (Altschul *et al*. 1990; Altschul *et al*. 1997). BLAST searching reveals the related sequences present in the same or different organisms. Sequence alignment helps to infer homology and function of genes. There is a general acceptance that if two sequences are matching, in part or as a whole, then they are similar or may be identical. Similarity between two genes can indicate similar functions of two sets (Bergeron, Bioinformatics Computing, 2003). Sequence similarities can be measured through global or local alignments. Since similar sequences may share active domains, they may also share common folds and functions. Therefore local alignment can give more appropriate results than global alignment (Zhong-Hui Duan 2006).

Homology is different from similarity and refers to two structures or sequences that evolved from a single ancestral structure or sequence (Gogarten and Olendzenski 1999). To understand more about homology, two more terms are described *viz.* orthology and paralogy (Fitch 1970). Orthology refers to homologous sequences or structures in two different organisms that evolved from the same feature in their last common ancestor but it does not mean that they have retained their ancestral function.

Paralogy refers to the homologous sequences or structures in the same organism, where as similarity is what we can measure from alignment of sequences or structures.

Classification of BLAST

BLAST can be classified in different classes (http://www.ncbi.nlm.nih.gov/BLAST) (Pevsner 2003) based on the comparison of proteins, nucleoside etc. as given below:

- BLASTp – protein-protein blast. It compares a protein query to a database of proteins.
- BLASTn – nucleotide-nucleotide blast. It compares both strands of a DNA query against a DNA database.
- BLASTx – translated query vs. protein database. It translates a DNA sequence to 6 protein sequences and then compares each of 6 proteins to protein database.
- TBLASTx – translated query vs. translated database. It translates DNA from both sides into 6 potential proteins and then performs 36 protein-protein database searches.
- TBLASTn – protein query vs. translated database. It translates every DNA sequence into 6 potential proteins and then compares our protein query against each of those translated proteins.

Pairwise Alignment

In the pairwise alignment, two sequences are placed right next to each other. For many reasons, the comparison of protein sequences is usually more informative than nucleotide sequences. For example many alternations (like third position of codon) do not affect the amino acid sequence.

Further, many amino acids share related biophysical or biochemical properties (e.g. Lysine and arginine are both basic amino acids) (Pevsner 2003).

Bidirectional Pairwise Alignment

In the bi-directional pairwise alignment, BLAST is performed to see whether sequence A shares sequence similarity to sequence B and if this holds true then another BLAST search is conducted to find out if sequence B also shares sequence similarity to sequence A. If sequence B is similar to sequence A then this is a result of bi-directional pairwise BLAST search.

The BLAST search algorithm finds a match between a query and a database sequence and then extends the match in either direction (Altschul *et al*. 1990; Altschul *et al*. 1997). BLAST algorithm can be divided into three different steps.

- In the first step, BLAST compiles a preliminary list of pairwise alignments, called word pairs (W). For example, given the sequences AGATAC and ACATAG and a W = 3, BLAST would identify the matching substring ATA that is common to both sequences. For protein searches the W has typically the size of 3 and for DNA the default size of W is 11 (Pevsner 2003).
- In the second step, BLAST extends the match in both directions until the alignment score goes under the threshold value. It is an important parameter to consider because it specifies how probable is to find the sequence that have biologically meaningful similarity. (http://bioinformatics.utu.fi/courses/alginbinf/similarity.pdf)
- In the third step, gap alignment is performed between the query and database sequence. Statistically significant alignments are displayed to the user.

Another powerful feature of BLAST is the use of Altschul-Dembo-Karlin (ADK) (www.bioalgorithms.info) statistics for estimating the significance of found matches. ADK calculates the probability of existence of equal or greater similarity between two random sequences of the same length. Meaning, given a sequence of length n and a database of length m, ADK statistics can calculate the probability of a match between n and m occurring randomly. This probability guides the choice of BLAST parameters and allows one to evaluate the statistical significance of found matches.

CHAPTER TWO

Materials and Methods

Bioinformatics Analysis

On the basis of sequence similarity, a systematic approach (as depicted schematically in Figure 2.1) was followed for identification of phenotypes in *C. elegans*. A database was made with all the 19,920 genes listed with their phenotypes in the Wormbase (Wormbase.org 2005) release WS159. The worm genome was classified in two different categories, one consisted of all the RNAi phenotypes associated genes and the second consisted of all *C. elegans* genes.

BLASTp program which compares a protein query to a database of proteins was used to analyze these genes. Reciprocal BLAST was performed to find the sequence similarity. E-value cut-off limit was 0.1 in first BLAST analysis. After picking the best hit (above the threshold value) from the first BLAST, reciprocal BLAST was performed to see if best hit sequence is similar to its parent one. If it was a reciprocal hit, then it was recorded.

In high-throughout bioinformatics, this is the most widely used means to determine whether or not sequences with positive sequence relationships may be possible homologs, either paralogs or orthologs. In this method, the notion of an "E value" is less important then the closest similarity neighbor. The initial BLAST of a sequence against a data base is filtered to return any possible sequence match (for example, E values of 0.1 to 0.5).

The top match is used as a seed for an additional BLAST query against the same data base. Since these results may produce alternative splice variants, further steps in our method are to eliminate matches that occur in the same genomic positions.

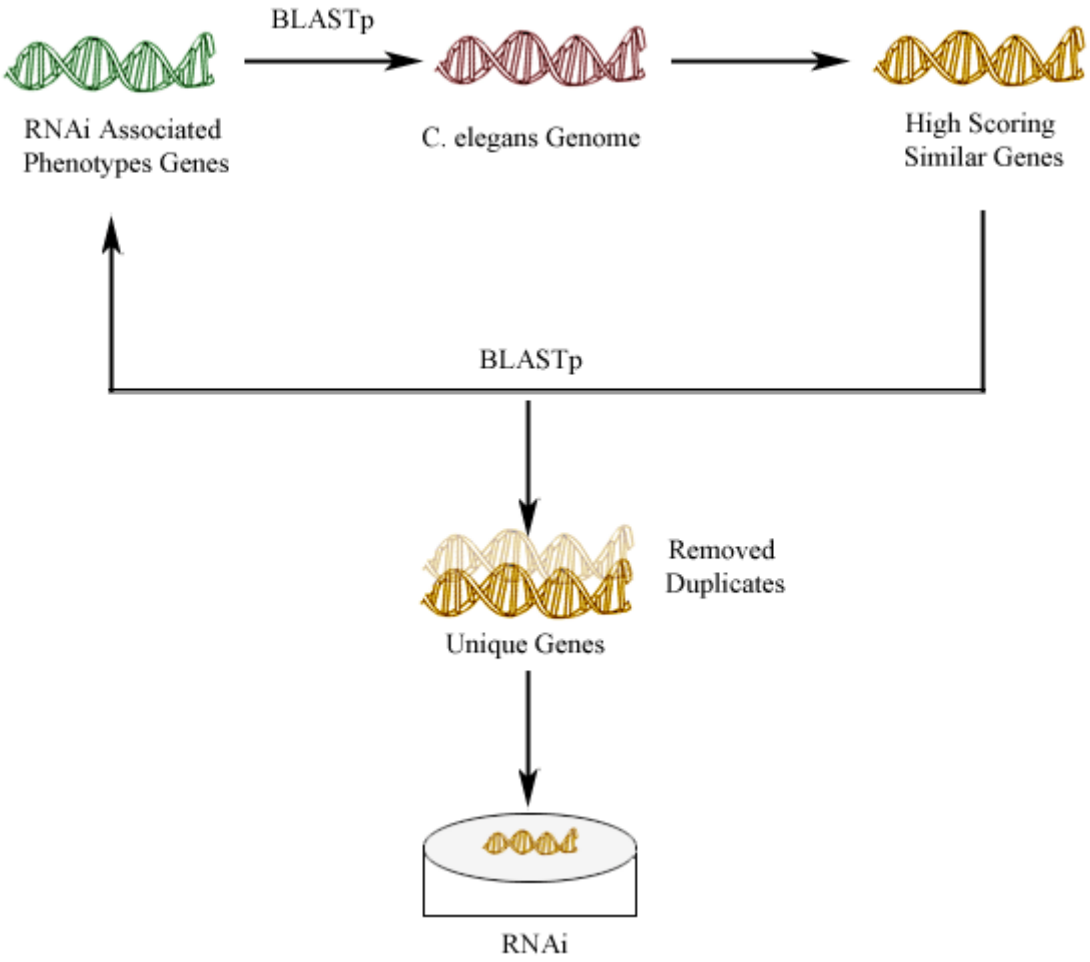

Figure 2.1. High schematic diagram demonstrating the sequence of BLAST experiments for phenotype identification.

Analysis and Identification of RNAi Phenotypes

Preparation of Laboratory Reagents and Plates

M9 buffer. M9 buffer consisted of an aqueous solution of 3 mg/ml of potassium phosphate monobasic, KH_2PO_4 (EMD chemicals, NJ, USA), 6 mg/ml of sodium phosphate basic, Na_2HPO_4 (Fisher Scientific, NJ, USA), 5 mg/ml of sodium chloride, NaCl (EM Science, NJ, USA), 0.12 mg/ml of magnesium sulfate, $MgSO_4$ (EM Science, NJ, USA), and 7.5 mg/ml of agar (Lab Scientific, NJ, USA). The buffer solution was prepared by dissolving all above-mentioned ingredients (except $MgSO_4$) in deionized (DI) water and autoclaved for 15 minutes. 1 ml of 1M $MgSO_4$ solution was added to the autoclaved solution after cooling. M9 buffer thus prepared was ready for experiments and stored at room temperature till further use.

Nematode freezing solution. For preparation of nematode freezing solution, following ingredients were added n 1 liter (L) of DI water - 5.6 ml of 1M sodium hydroxide (NaOH, purchased from EM Science, NJ, USA), NaCl (5.85 mg/ml), KH_2PO_4 (6.8 mg/ml), and 300 gm (equivalent to 240 ml) of glycerol (EM Science, NJ, USA). The resulting solution was then autoclaved for 15 minutes followed by addition of 3 ml of 0.1 M $MgSO_4$ solution. Nematode freezing solution thus prepared was ready for experiments and stored at room temperature till use.

Lysogeny broth (*LB*). LB, required for inoculation was prepared by adding tryptone (10 mg/ml), yeast extract (5 mg/ml), NaCl (10 mg/ml), and 10 ml of 1 M Tris-HCl pH 7.5 buffer in 500 ml of DI water.

All the ingredients were mixed thoroughly and aliquot in 100 ml bottles which were then autoclaved and stored at -20^{o}C. Tryptone and yeast extract were obtained from Becton and Dickinson (MD, USA) and Tris-HCl buffer was purchased from Fisher Scientific (NJ, USA).

Ampicillin stock solution. A 40 mg/ml stock solution of ampicillin (procured from Shelton Scientific, CT, USA) was prepared by dissolving 2 gm of drug in 50 ml of DI water. The stock solution was filtered through a sterile 0.22 micron pore size filter and filtrate was decanted to sterile 50 ml centrifuge tubes. Finally, 4.5 ml of the ampicillin stock solution was aliquot into sterile 15 ml conical vials and stored in -20^{o}C freezer.

LB plates. LB plates were needed for seeding the bacteria. For making LB plates, agar (15 mg/ml) was added during the preparation of LB broth before autoclaving. Autoclaved media was poured into petri dishes (100 mm diameter) and allowed to solidify. The solidified plates were inverted and left on the bench top for overnight and then stored at 4^{o}C till further use in experiments.

LB Plates with ampicillin. LB plates with ampicillin (40 µg/ml) were prepared by adding 0.5 ml of sterile ampicillin stock solution to the media (used in above-mentioned LB plates) before autoclaving. Remaining procedure of preparation of plates was same as that of LB plates.

Tetracycline solution. A 15 mg/ml stock solution of tetracycline (procured from Shelton Scientific, CT, USA) was prepared by dissolving 150 mg of drug in 10 ml of 100% ethanol (obtained from VWR international, PA, USA). The stock solution was aliquot into 1.5ml sterile microcentrifuge tubes. Finally, these microcentrifuge tubes were stored in the dark colored bottle in the -20^{o}C freezer.

β-lactose stock solution. A 20% w/v solution of β-lactose was prepared by dissolving 10 gram of lactose powder (Becton and Dickinson, MD, USA) in 50 ml of DI water in a 100 ml bottle or beaker. Lactose powder was dissolved with the help of a magnetic stir bar by keeping the bottle or beaker on hot plate. After dissolving, the solution was filtered through a sterile filter (0.22 micron pore size) placed on the top of corning sterile filter cup. The filtrate was decanted into a sterile 50 ml centrifuge tube and 4.5 ml was aliquot into sterile 15 ml conical vials which were stored at -20^{o}C till further use.

Mediated interference (RNAi) plates. Nematode growth media (NGM) plates with the addition of ampicillin and lactose are normally called as RNAi plates and are used to facilitate the growth of the dsRNA producing bacteria. Preparation of RNAi plates involved multiple steps:

(a) Mixing of agar (17 mg/ml), peptone (2.5 mg/ml), and NaCl (3 mg/ml) in 1L of distilled water,

(b) Autoclaving,

(c) Cooling of molten agar to 55^{o}C,

(d) Adding 1 ml each of 5% cholesterol solution, 1M $MgSO_4$, 1M calcium chloride ($CaCl_2$), 1M KH_2PO_4, 20% w/v lactose solution and stock solution of ampicillin in solution of previous step,

(e) Pouring of approximately 13.5 ml of agar into each plate, and

(f) Cooling of the plates. The RNAi plates thus prepared were stored at 4°C. Peptone obtained from Lab Scientific (NJ, USA), cholesterol from J. T. Baker (NJ, USA), $CaCl_2$ from EM Science (NJ, USA) and other ingredients used from earlier mentioned sources were used in preparation of NGM media.

C. elegans Strains

Researchers have been using different *C. elegans* strains such as *N2* (Fraser *et al*. 2000), *rrf-3* (Simmer *et al*. 2002) *eri-1* (Kennedy *et al*. 2004) and *lin 15B* (Wang *et al*. 2005). However, mutation of *rrf-3* which is a RNA-directed RNA polymerase (RdRP) has been proven to result in increased sensitivity to RNAi (Simmer *et al*. 2002; Simmer *et al*. 2003). Due to its high sensitivity to RNAi, we have also used *rrf-3(pk-1426)* strain in our RNAi study. The *rrf-3* strain was grown and maintained at 20° C with *E. coli* OP50 bacteria. OP50 bacteria are a leaky *E. coli* uracil auxotroph that grows slowly and provides nutrients to *C. elegans* without overgrowing them (Brenner 1974). Strains were procured from Caenorhabditis Genetics Center (CGC), University of Minnesota, USA.

Freezing of strains. For freezing the strains, M9 buffer (3 ml) was added to two large plates full of freshly starved L1 and L2 staged *C. elegans.* Plates were left undisturbed for 3 minutes to make the worms float. M9 buffer with worms was collected with a sterile Pasteur pipette in a sterile 15 ml centrifuge tube.

Sterile 60% sucrose (EM Science, NJ, USA) solution (150 µl) was added to the tube and mixed with gentle shaking. Tube was centrifuged at full speed for 3 minutes. After centrifugation, the supernatant was removed and 500 µl of M9 buffer and 500 µl of nematode freezing media were added to the tube and mixed well by gentle shaking. This worm solution was aliquot into two cryovials, placed in the sponge rack and frozen at -80^{o} C.

RNAi Feeding for Sterility

RNAi by feeding was used to study the Sterility phenotype with a sample size of 25 worms. *E. coli* OP50 bacteria from the UK HGMP resources center (Hinxton, UK) were purchased and stored in -80^{o}C freezer. Desired bacteria were taken out and seeded on LB agar plates and allowed to incubate at 34ºC for 12 hours. A small colony of bacteria was isolated and inoculated at 34ºC for 12 hours with a small amount of LB broth containing ampicillin and tetracycline to inhibit the growth of any other bacteria than desired.

Three RNAi plates were seeded with LB culture broth and the plates were maintained at room temperature for 12 to14 hours. Next day, five hermaphrodites were transferred to a RNAi plate where they laid the eggs. After four hours, adults were destroyed and progeny was left to grow for one day. From this group, 25 worms were transferred to 25 different RNAi seeded with LB culture broth plates.

These plates were stored at 20ºC for two days. Approximately after two days, all the plates were observed under the Nikon SMZ800 stereomicroscope (Nikon Instruments Inc, NY, USA) and scored for sterility phenotype. Sterility required a brood size of less than 10 and was required to be present among at least 10% of worms.

RNAi Feeding for Uncoordinated Movement

RNAi by feeding was also used to study the uncoordinated movement phenotype with a sample size of 15 worms. Desired bacteria were taken out from -80^{o}C freezer, seeded on LB agar plates and inoculated in liquid broth as mentioned in RNAi feeding for sterility. Three RNAi plates were seeded with this LB culture broth and the plates were maintained at room temperature for 12 to 14 hours. Next day, 15 hermaphrodites were transferred to 15 different RNAi plates where they laid the eggs. After four hours, adults were destroyed and progeny was left to grow. These plates were stored at 20^{o}C and thrashing assay was performed on 15 young adult worms. Various steps used in RNAi Feeding for studying sterility and uncoordinated phenotypes are given in Table 2.1.

Table 2.1. RNAi schedule.

Days	Description
Day 1	Chunk the worm stock plates onto NGM seeded OP50 bacteria. Inoculate the given RNAi bacteria in small amount of ampicillin/tetracycline LB broth and place in 34^{o}C incubator for 12-14 hours.
Day 2	Take out inoculation from 34^{o}C and seed 3 RNAi plates. Place these plates in 34^{o}C incubator to dry
Day 3	Place 5 adult *rrf-3* worms onto each seeded plate on day 2. After 4 hours kill adult worms and check for laid eggs and place them in 20^{o}C. Inoculate the given bacteria in large amount of ampicillin/ tetracycline LB broth and place in 34^{o}C incubator for 12-14 hours.
Day 4	Take out inoculation from 34^{o}C and seed 25 plates for sterility and 15 for Unc and place these plates in 34^{o}C incubator to dry.
Day 5	Remove plates seeded on day 4 and plates with eggs from day 3. Transfer 1 worm each on all 25 plates for Ste and on all 15 plates for Unc and place it at 20^{o}C for 48 hours.
Day 6	Waiting period.
Day 7	Check out plates and analyze phenotypes.

Thrashing Assay

Thrashing is a high frequency locomotory behavior that occurs when animals are placed in liquid (Ackley *et al*. 2003). In this behavior there is a coordinated movement in which the animal brings the head and tail toward each other, flexing around the approximate midpoint with a regular amplitude and alternation between ventral and dorsal flexure.

Worms with Unc phenotype exhibit uncoordinated movements such as reductions in the rate of thrashing behavior and overbending (such that head and tail cross to each other) etc. Thrashing movie stills from a previously reported study (Ackley *et al*. 2003) are shown in "Introduction" section (Figure 1.7). This assay was done as described in a previous study (Lee *et al*. 2005) to find the uncoordinated movement phenotype in *C. elegans.* Young adult worms were transferred to 10 µl of M9 buffer and after a short recovery period thrashes were counted for 30 seconds under Nikon SMZ800 stereomicroscope to determine the rate of thrashing. Thrashing assay results were statistically analyzed by Wilcoxon-Mann-Whitney test using JMP (release 6) Statistical Software (SAS institute, Cary, NC, USA).

Control Set of Genes

A script was written in a Perl program (release 5.8.0) on a dual core computer processor with 2 gigahertz (GHz) speed by using open source Linux operating system to generate random set of genes from the pool of 141 genes. These 141 genes sequences were similar to the genes which were not associated with Ste and Unc phenotypes. RNAi for both Unc and Ste was also performed on this control set.

Reverse Transcription Polymerase Chain Reaction (RT-PCR)

Extraction of mRNA From C. elegans

Extraction of intact RNA is one important factor for synthesis of full-length complimentary deoxyribose nucleic acid (cDNA) for RT-PCR. The messenger ribose nucleic acid (mRNA) was extracted from the worms by following the below-mentioned procedure:

1. Three plates of *rrf-3* strain *C. elegans* treated with various RNAi conditions and grown for 48 hours were washed with M9 buffer.
2. Worms were centrifuged and frozen either on dry ice or liquid nitrogen.
3. 1 ml of Trizol and 12.5 µl of beta mercaptoethanol (BME) were added to the worms and vortexed for 30 seconds. Both Trizol and BME were obtained from Sigma-Aldrich (MO, USA).
4. Supernatant was transferred to a new tube and 150 µl of chloroform (obtained from Burdick and Jackson, MI, USA) was added.
5. The tube was centrifuged at 14,000 rotations per minute (rpm) for 10 minutes.
6. Again, supernatant was transferred into a new tube and 0.8 volume of ice-cold isopropanol (obtained from Fisher Scientific, NJ, USA) was added and centrifuged at 14,000 rpm for 10 minutes.
7. RNA pellet formed at the bottom of the tube was washed by 70% of ethanol.
8. Ethanol was completely dried and resuspended in 20 µl of 0.1% diethylpyrocarbonate (DEPC, obtained from EMD Chemicals, NJ, USA).

RT-PCR

Reverse transcription kit (SuperScript™ First-Strand Synthesis System for RT-PCR) was obtained from Invitrogen (CA, USA) and all the ingredients were used as such without further diluting. RT-PCR was performed in multiple steps as mentioned below:

1. In a centrifuge tube, 2 µl of mRNA extracted from *C. elegans*, 1 µl of 10 mM dNTP i.e. deoxyribonucleotide triphosphate, 1 µl of 0.5 µg/ml oligodeoxythymidylic acid (Oligo dT) and 6 µl of DI water were added.
2. The tube was placed in PCR machine (Applied Biosystem, Gene Amp, PCR system 2700) at 65°C for 5 minutes and was placed on ice for 1 minute.
3. 2 µl of 10X RT buffer (consisting of 200 mM Tris-HCl having pH 8.4 and 500 mM KCl), 4 µl of 25 mM magnesium chloride ($MgCl_2$), 2 µl of 0.1 M dithiothreitol (DTT) and 1 µl of RNAseOut™ recombinant ribonuclease inhibitor (40 units/µl) were added to above step and mixed.
4. The tube was centrifuged again and placed at room temperature for 2 minutes.
5. The tube was placed in the PCR machine for 2 minutes at 42°C.
6. After 2 minutes, tube was taken out and 1 µl of SSII RT (SuperScript™ II RT, 50 unit/µl) was added.
7. The tube was placed in the PCR machine for 42°C for 50 minutes and 70°C for 15 minutes.
8. In the last step of preparing cDNA, 1 µl of *E. coli* RNase H (2 units/µl) was added and mixed gently.
9. The tube was again placed in the PCR machine for 37°C for 20 minutes.

10. Forward and reverse primer (obtained from Biosynthesis Inc, TX, USA) were diluted (10 µM/L) and 1 µl of each was mixed with 2 µl of cDNA, 0.2 µl of Taq DNA polymerase (procured from Promega, WI, USA), 2 µl of dNTP (obtained from Promega, WI, USA) and 2 µl of 10X buffer (Promega, WI, USA) and water was added to make up the volume up to 20 µl and mixed thoroughly.
11. Above mentioned mixture was centrifuged and placed in the PCR machine at the following cycles as described in Table 2.2.

Table 2.2. PCR cycles, temperature and time.

Number of cycles	Temperature/time
Cycle 1	96^{o}C for 5 minutes
Cycle 2 (28 cycles)	95^{o}C for 30 seconds
	55^{o}C for 30 seconds
	72^{o}C for 30 seconds
Cycle 3	72^{o}C for 10 minutes

PCR products were discriminated on a 1.2% Agarose gel. Gel was stained in ethidium bromide (Fisher Scientific, NJ, USA), for 20-30 minutes and then washed with water. In the last step, gel was placed in the ultra violet illuminator (UVP, CA, USA) and photographed.

Primers were made in D.S. Gene 1.1 (Accelrys Software, Inc., San Diego, CA) to perform RT-PCR for sterility-associated genes and genes of high similarity (Table 2.3).

Table 2.3. Primers for the previously reported sterility-associated genes and genes of high similarity were constructed to test RNAi efficiency.

Primers for Putative Phenotype Associated Genes	Primers for Known Phenotype Associated Genes
F17A2.3 5'-TCCAAAAATGAGAGGAGAATGC-3'(Forward) 3'-TGCTCCTTTGATTTGATATGGC-5'(Reverse)	C44B9.4 5'-GCAGAAAAAAGAGAAAGAAGGC-3'(Forward) 3'-GATGTGCTGGCTTCTTCAG-5'(Reverse)
B0273.4 5'-GGACGAAATCACAATCACAAC-3'(Forward) 3'-GATTACTATCCGAGAGGAGTCC-5'(Reverse)	F54D12.5 5'-TGCCACACGTGGAAGCAAAC-3'(Forward) 3'-TGTGCATAGTTGAGACGTGGC-5'(Reverse)
ZK185.1 5'-CAAATCACAAGAAACAGGTCG-3'(Forward) 3'-TTACTTAACCATCTCATCGTGC-5'(Reverse)	C32E8.3 5'-CTGCTGCTGGATTCAACTG-3'(Forward) 3'-AATCGGCTTCTTACTTTGTCTC-5'(Reverse)
F58G5.4 5'-GCTGTTGAACAGATGTACCC-3'(Forward) 3'-CGCTGTTGGGAAGAAGATTG-5'(Reverse)	Y71D11A.5 5'-TCTGCAATTCCTATCGCTTG-3'(Forward) 3'-TTTACATAAGAAACCCTCGGC-5'(Reverse)
Y39A1C.2 5'-CGAACTCTTCAAAAGAATGCG-3'(Forward) 3'-GAGCTCAGTGCATCTAGAAAG-5'(Reverse)	D2085.4 5'-CATTTGGAATGGGATTCAATGG-3'(Forward) 3'-TGAAGCCATCAACTGGAGC-5'(Reverse)
C32B5.9 5'-TTTCGGTGGAAGAGTTATTGTC-3'(Forward) 3'-AGGAAAAGAAGTAAGATGCAGC-5'(Reverse)	T10H9.3 5'-CGATCACAACAGTTCAAAGAAC-3'(Forward) 3'-CGATCTCCGTTAAATGCTGC-5'(Reverse)
W01C9.3 5'-CCGCCTTCCTATCGAAATTC-3'(Forward) 3'-TCTGCTGAAGCTGAATAATTCG-5'(Reverse)	C09F5.1 5'-ATGGTGGTGGAGCAAATCG-3'(Forward) 3'-GTTTTTCTCCGGCACTCTTC-5'(Reverse)
M151.1 5'-TGGACCATCCTTTCAGTGG-3'(Forward) 3'-TCTCCAACTTGTAGAGTTCCTG-5'(Reverse)	W09C3.4 5'-TGGCTTCCATCAAAGAAGAAC-3'(Forward) 3'-CACGGATCTTTCCATCAGC-5'(Reverse)
	Control 5'-CTCAACGAAACTACACCCCTGCC-3' (Forward) 3'-GTGTCACTTACCGTATCCCATT (Reverse)

CHAPTER THREE

Results

Several studies have been published to find the RNAi phenotypes in *C. elegans* as described in Table 1.8. Few studies in *C. elegans* have shed light on the homology in other species. In the present study, we found sequences in *C. elegans* whose RNAi phenotypes have not been reported in previous studies and which share high sequence similarity in *C. elegans* itself.

The scope of testing each RNA interference gene target against all possible phenotypes is an unreasonable expedition for most laboratories. By demonstrating a high success rate of narrowing possible phenotypes based on previously known results using bioinformatics techniques allow researchers to focus their efforts on genes or phenotypes of particular interest.

Variations in experiments can lead to different results at the end. For example the mode of introducing dsRNA, deviation in the temperature and time of scoring the phenotypes can significantly affect the outcome of phenotype assays. Since RNAi acts in a sequence specific manner, therefore it is possible that dsRNA knocks out another similar gene. Hence phenotype observed can be a result of multi-gene silencing.

Bioinformatics Analysis

Sequence Similarity Identification

The peptide sequences from 2807 genes with previously reported RNAi-associated phenotype were obtained from the Wormbase database (Wormbase.org 2005, release WS 159). In addition, corresponding peptide sequences from all 19,920 genes in the *C. elegans* genome were retrieved and formatted for standalone use with the BLAST suite of tools (Altschul *et al*. 1990; Altschul *et al*. 1997). High scoring similar sequences in two sets of genes were determined as previously reported (Wolf *et al*. 2001).

A flowchart of whole sequence followed for BLAST experiments and the results obtained at each step are given in Figure 10. Out of 19,920 genes of whole *C. elegans* genome, we observed 1155 genes after first BLAST which share high sequence similarity with genes of known RNAi phenotypes (Step A of Figure 3.1). These 1155 genes were the second best hit whereas first best hit were the same genes. Best results of the initial BLASTp test were compared again to the initial data set (as shown in Step B). The application of reverse BLASTp elucidated 442 genes of high similarity. After removing the duplicates and transcripts of the genes which occur within the same open reading frame (ORF) of the RNAi associated phenotype genes, 141 genes of unknown RNAi-associated phenotype existed with putative similarity to the initial set (Step C). In the last step (Step D of Figure 3.1), this set was further pared by limiting the focus to putative genes with high sequence similarity to RNAi identified genes associated with the sterility and uncoordinated movement phenotypes.

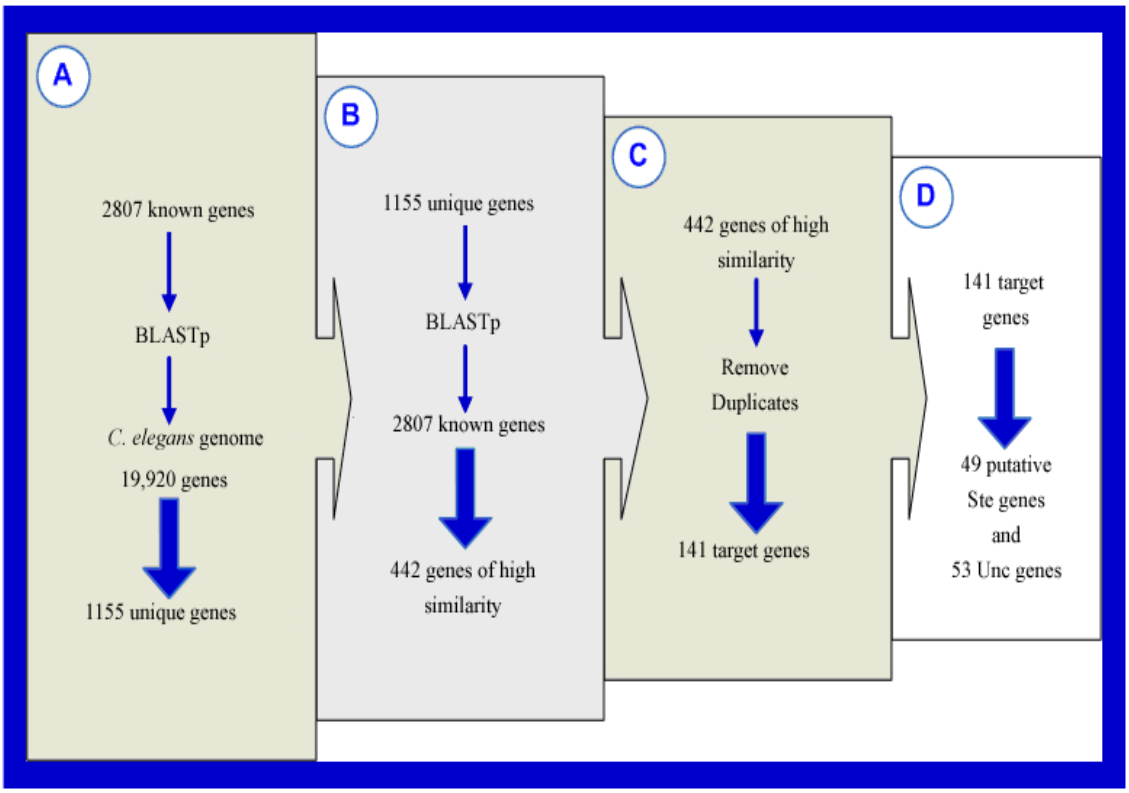

Figure 3.1. Results showing number of genes identified based on sequence similarity at each step of BLAST experiments.

Table 3.1 lists all target 141 genes identified based on sequence similarity and were earlier not associated with any RNAi phenotype. Complete list of 2807 known genes associated with RNAi phenotypes is given in Appendix C.

Distribution of phenotypes associated with the set of 141 unique genes is described in Figure 3.2. Briefly, the results segregated into sixteen phenotype categories, with a maximum category classification occupied by 59 putative genes that aligned well with Gro (retarded growth) genes.

Table 3.1. List of target 141 genes identified by BLAST on basis of sequence similarity.

Genes which are not associated with RNAi phenotype				
R04A9.2	K07E3.7	D2063.2	ZK185.1	ZC21.10
T19D7.1	F18E9.5	R02F11.1	F28E10.4	ZK370.8
F35H12.4	F57C7.1	K11D12.9	F36H12.8	F44E2.8
F55A4.2	F47B10.1	C18C4.5	B0273.4	ZK643.2
C46H3.3	T04F8.4	C04E6.5	T12E12.2	R10E12.1
Y71H10A.1	ZK185.4	R10E12.2	F17A2.3	W06H8.1
F40F4.1	F17E5.1	Y97E10AL.3	H10D12.2	Y39A1C.2
ZK816.1	F22E10.5	C24B5.2	D2063.3	K10B4.1
ZK813.6	C05E7.2	Y22F5A.3	H20E11.1	C32B5.9
C18B2.4	R03G8.6	T04H1.9	R05G6.7	T08E11.6
Y32B12B.2	F09A5.1	R05D8.8	F58G6.4	B0047.2
ZC64.4	F46G10.2	T28H10.1	F56C4.1	F43C11.3
F09F9.2	F55F3.3	R186.2	F38B6.4	F34D6.4
K05B2.2	F16B12.6	F44G3.7	M02B1.1	M151.1
ZC8.6	ZK6.2	F13A7.10	F09E8.8	Y14H12B.2
C24H10.1	ZK488.1	T08G3.7	T23G4.5	R03H10.6
C25F6.3	W07B8.5	Y32B12B.1	BE0003N10.3	C09E7.4
F14D12.6	R12A1.4	R11G1.4	Y39A3B.5	T05C1.3
F13D11.4	R09B5.5	W03G1.8	Y48G9A.11	T05A7.1
C46F4.2	K04F1.7	F58H7.3	H14E04.1	F09C12.1
SSSD1.1	R05D8.5	C45G7.2	F25F2.2	C29F5.7
C25B8.3	T03D3.11	M02B7.1	F26A1.14	F18C5.8
F38B6.2	R08F11.7	Y46C8AL.2	C09E7.3	ZK1290.9
C39D10.7	D1065.1	ZK1025.7	ZK112.5	B0228.7
T08A9.4	F47D2.10	F49F1.11	ZC97.1	W01C9.3
T22C8.5	F56A6.1	F47B3.5	F13G3.11	ZK1025.3
T26C5.1	Y18H1A.11	Y47D9A.3	C36F7.5	F47C12.4
T02G6.7	Y6D1A.2	Y73E7A.8	W02D3.11	Y53H1A.5
Y47H9B.2				

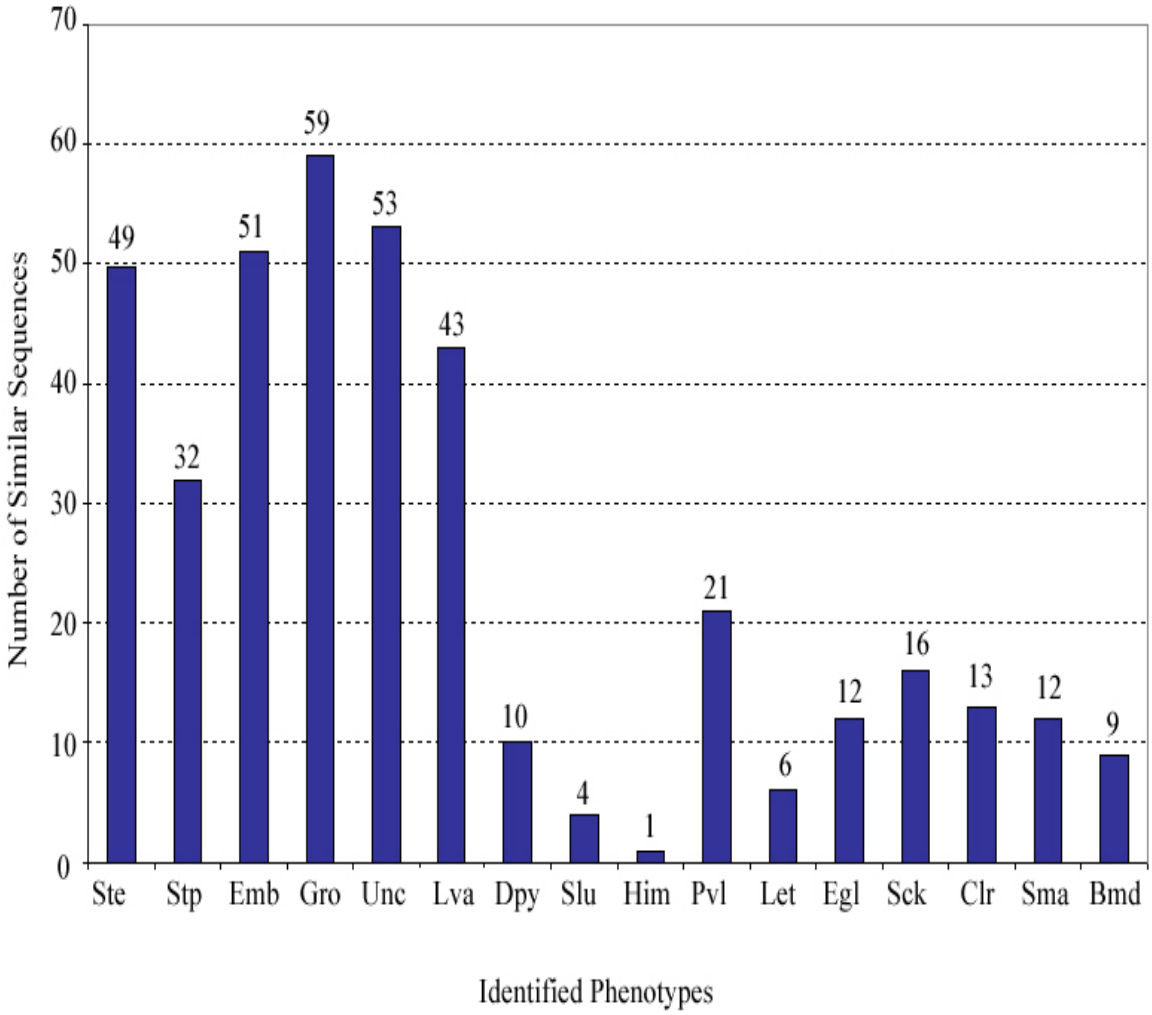

Figure 3.2. Distribution of phenotypes for 141 target genes identified based on the sequence similarity.

Other categories include such genetically defined phenotypes categories such as Emb (embryo abnormal), Unc (uncoordinated movement), Dpy (dumpy), Him (high incidence in male progeny), Pvl (protruding vulva), Let (lethal), Egl (egg laying abnormal), Clr (clear body), Sma (small) and Lon (longer than wild type). Besides there were some phenotypes usually defined in RNAi studies such as Bmd (body morphology), Ste (sterility), Stp (sterile progeny), Lva (larval arrest), Slu (sluggish), and Sck (sick), which can be alternative descriptions of the genetically defined phenotypes. Most of the genes were found to be associated with the multiple phenotypes.

Analysis and Identification of RNAi Phenotypes

RNAi for Sterility Phenotype

Bioinformatics analysis resulted in identification of 49 putative genes associated with Sterility phenotype. Due to limited availability of cDNA library, we were able to perform RNAi by feeding only on 37 putative genes. Previous studies have used a very small sample size of 10 or less worms in RNAi experiments for sterility phenotype. In the present study we have used a large sample size of 25 worms to generate more reliable, accurate and reproducible data. A list of 37 candidate genes for sterility is given in the Table 3.2 with their chromosome number, description and sterility score. Sterility score was assigned to a gene when a brood size of less than 10 was observed among at least 10% of the worms. The *C. elegans* strain *rrf-3*(*pk-1426*) was used in this study to analyze the Ste phenotype.

Table 3.2. List of 37 genes studied by RNAi and their sterility score. This table depicts gene name, chromosome number (Chrom#) and their description. N is number of worms.

Gene	Chrom#	Description	% Sterility (N=25)
C46H3.3	X	Ankyrin repeat	0
F40F4.1	X	F-box domain	0
R11G1.4	X	Protein kinase ATP binding	0
ZC64.4	X	LIM domain	0
F09F9.2	X	Unnamed protein	0
K05B2.2	X	Predicted cell growth/differentiation regulator, contains RA domain	0
ZC8.6	X	Phosphatidylinositol 3- and 4-kinase	0
F14D12.6	V	7 transmembrane receptor	0
K07E3.7	V	Cation transport ATPase	0
F17A2.3	V	Zinc finger, PHD-type	68
F09A5.1	V	Major facilitator superfamily	0
F46G10.2	V	Unnamed protein	0
C24B5.2	V	ATPase family	0
F44G3.7	V	Unnamed protein	0
T08G3.7	V	Uncharacterized protein, contains BRCT, WSN domains and ankyrin repeats	0
W03G1.8	IV	Unnamed protein	0
C45G7.2	IV	Unnamed protein	0
F49F1.11	IV	Galectin, galactose-binding lectin	0
ZK185.1	IV	Zinc finger, C2H2 type	44
ZK354.2	IV	Casein kinase	8

Table 3.2 Continued.

Gene	Chrom#	Description	% Sterility (N=25)
B0273.4	IV	Netrin receptor for axon guidance (unc-5)	40
H20E11.1	IV	Unnamed protein	0
R05G6.7	IV	Porin/voltage-dependent anion-selective channel protein	0
F58G6.4	IV	Neurotransmitter ion gated channel	52
F56C4.1	IV	Unnamed protein	0
Y39A3B.5	III	Neuropeptide Y receptor	0
ZC97.1	III	Glutathione S-transferase, C-terminal-like	4
ZK370.8	III	Tetratricopeptide-like helical	8
F44E2.8	III	Unnamed protein	0
Y39A1C.2	III	Ubiquitin protein ligase activity	16
C32B5.9	II	F-box domain	12
M151.1	II	Similarity to Halothermothrix orenii H 168 ATPase involved in DNA repair	20
W01C9.3	II	-73 prion like -(Q/N-rich -domain bearing protein)	12
T22C8.5	II	Zinc finger	8
F56A6.1	I	Translation initiation factor 2C (eIF-2C)	0
F47B3.5	I	Uncharacterized protein	0
C36F7.5	I	Uncharacterized protein	0

As a result of RNAi studies on 37 putative genes, we observed 8 new genes which were sterile and not associated with any phenotype in previous studies, indicating a high success rate of 22%. This is the first study when these 8 new genes are associated with Ste phenotype. The new genes and their sterility score are listed in Table 3.3.

Table 3.3. Eight new genes associated with Ste phenotype and their sterility score. N is number of worms.

Gene Name	% Sterility (N=25)
F17A2.3	68
B0273.4	40
ZK185.1	44
F58G6.4	52
M151.1	20
Y39A1C.2	16
W01C9.3	12
C32B5.9	12

Previously reported RNAi phenotypes associated genes which shared sequence similarity with 8 homology identified (new) genes, the chromosome number and description of both new and known genes are summarized in Table 3.4. Sterility criteria of these 8 new genes was divided into three broad categories - High (more than or equal to 40% of RNAi worms tested produced less than 10 viable progeny), Medium (20% to 39% sterile worms), and Low (10% to 19% of sterile worms).

Table 3.4. Summary of known sterility RNAi phenotype genes and homology identified genes, which share high sequence similarity to known RNAi genes. The table summarizes the names of both types of genes, their chromosome number (Chr. #), and description. Also, the sterility score of putative phenotypes is given (High > 40% worms produced < 10 viable progeny, Medium 20%-39%, and Low 10%-19%).

Previously Reported RNAi Phenotype-Associated Genes	Chr. #	Description	Known RNAi Phenotypes	Corresponding Putative RNAi Phenotype-Associated Genes	Chr. #	Description	Sterility
C44B9.4	II	Zinc finger, PHD type	Ste	F17A2.3	X	Zinc finger, PHD type	High
F54D12.5	II	Unknown	Ste	B0273.4	IV	Netrin receptor for axon guidance (unc-5)	High
C32E8.3	I	Putative signal transduction protein p25	Lva, Ste	ZK185.1	IV	Zinc finger, C2H2 type	High
Y71D11A.5	III	Ligand-gated ion channel	Gro, Sck, Lvl, Ste Unc, Emb	F58G6.4	IV	Neurotransmitter ion gated channel	High
W09C3.4	I	RNA polymerase III, subunit C34	Ste	M151.1	II	Similarity to *Halothermothrix orenii* H 168 ATPase involved in DNA repair	Medium
D2085.4	II	E3 ubiquitin protein ligase	Ste	Y39A1C.2	III	Ubiquitin protein ligase activity	Low
T10H9.3	V	SNARE protein Syntaxin 18/UFE1	Clr, Gro, Sck, Lva, Unc, Ste	C32B5.9	II	F-box domain	Low
C09F5.1	III	Unknown	Pvl, Ste	W01C9.3	II	-73 prion like -(Q/N-rich -domain bearing protein)	Low

As evident from the description (Table 3.4) of both new genes and previously known genes, five of the eight genes such as F17A2.3, ZK185.1, F58G6.4, Y39A1C.2, and M151.1 appeared to possess similar protein functions to their previously reported Ste gene pairs. Also the descriptions of 3 genes were largely related to gene expression regulation, including two transcription factors (F17A2.3 and ZK185.1) and an RNA polymerase (M151.1). Two genes, Y39A1C.2 and C32B59.6, are predicted to play a role in protein degradation. F54D12.5 is a gene of unknown function. The gene, F58G6.4, encodes for an acetylcholine-gated ion channel. However, sterility of B0273.4, *unc-5* RNAi was not previously reported (Leung-Hagesteijn *et al*. 1992), suggesting that the UNC-5 might play a role in a novel function involved in fertility. The clone W01C9.3 appeared to be similar to Q/N rich – domain binding protein but the function of W01C9.3 and the reported pair (C09F5.1) is still unknown.

RNAi for Uncoordinated Movement (Unc) Phenotype

As a result of BLAST studies, 53 putative genes associated with Unc phenotype were identified, however only 29 putative genes were studied for RNAi phenotype due to limited availability of cDNA library. Previous studies have used a small sample size of 10 or less in RNAi experiments for Unc phenotype. In the present study we have used a large sample size of 15 worms to generate more reliable, accurate and reproducible data.

Twenty nine genes which shared high sequence similarity with known Unc genes selected in this study are described in Table 3.5 along with their chromosome numbers and description.

Table 3.5. List of 29 genes studied by RNAi for potential Unc phenotype.

Gene	Chromosome	Description
R11G1.4	X	Homolog of Ndr kinase that regulates neuronal cell shape and neurite initiation
C24H10.1	X	Encodes a claudin homolog that may be required for normal cohesion of apical junctions in epithelia
ZC64.4	X	Transcription factor, contains HOX domain
C46F4.2	X	Fatty Acid CoA Synthetase family
F13D11.4	X	Flavonol reductase/cinnamoyl-CoA reductase
C25B8.3	X	Cysteine proteinase Cathepsin L
F38B6.2	X	Unnamed protein
F17E5.1	X	Guanylate Kinase, abnormal cell lineage
F18E9.5	X	DNA-binding protein jumonji/RBP2/SMCY, contains JmjC domain
R03G8.6	X	Puromycin-sensitive aminopeptidase
F55F3.3	X	Na+/K+ ATPase, beta subunit
F16B12.6	X	Unnamed protein
T03D3.11	V	7-transmembrane olfactory receptor
ZK6.2	V	Nuclear hormone receptor
F13A7.10	V	Glutathione S-transferase, N-terminal domain
R186.2	V	Serpentine Receptor, class D (delta)
T28H10.1	V	Alpha-aminoadipic semialdehyde dehydrogenase-phosphopantetheinyl transferase
F58G6.4	IV	Acetylcholine-gated chloride channel
F28E10.4	IV	Unnamed protein
F36H12.8	IV	Casein kinase
W03G1.8	IV	Unnamed protein
F25F2.2	III	CaDHerin family
Y39A3B.5	III	7 transmembrane receptor
ZK370.8	III	Tetratricopeptide-like helical
C32B5.9	II	F-box domain
Y14H12B.2	II	Uncharacterized protein
B0228.7	II	Methylthioadenosine phosphorylase
T22C8.5	II	Zinc finger, C2H2-like
ZK1025.3	I	Predicted secreted small molecules methylase

All these 29 genes were tested for Unc phenotype by thrashing assay as described in "Materials and Methods" section. The thrashing frequency of each mutant was compared to *rrf-3*(*pk-1426*) grown on pPD129 control bacteria. The results were statistically analyzed by Wilcoxon two group test as described in previous study (Lee *et al*. 2005). It compares several distributions by ranking the data, comparing ranks with each other and analyzing the data by JMP Statistical Software. For every gene tested, the *rrf-3* (*pk-1426*) worm strain was used.

Graph in Figure 3.3 represents the number of thrashings presented by box and whisker plots (JMP release 6 software). It depicts the thrashing pattern of 12 genes which were found positive as Unc phenotype in our analysis. In the graph, each box represents an interquartile range (IQR, the central 50% of the data points). Number indicates the average thrashes per 30 seconds and horizontal line in the boxes indicate medians. Average thrashes per 30 seconds for these 12 genes were found to be significantly ($\alpha = 0.05$, N = 15) lower than that in *rrf-3*(*pk-1426*) grown on pPD129 control bacteria, thus indicating uncoordinated movement.

Out of 29 genes on which RNAi was performed, we found 12 new genes which were uncoordinated and not associated with Unc phenotype in previous studies. Thus, we obtained a high success rate of around 41% in the case of Unc phenotype. A detailed description of these newly found 12 Unc genes and genes which share high sequence similarity to these genes is given in Table 3.6.

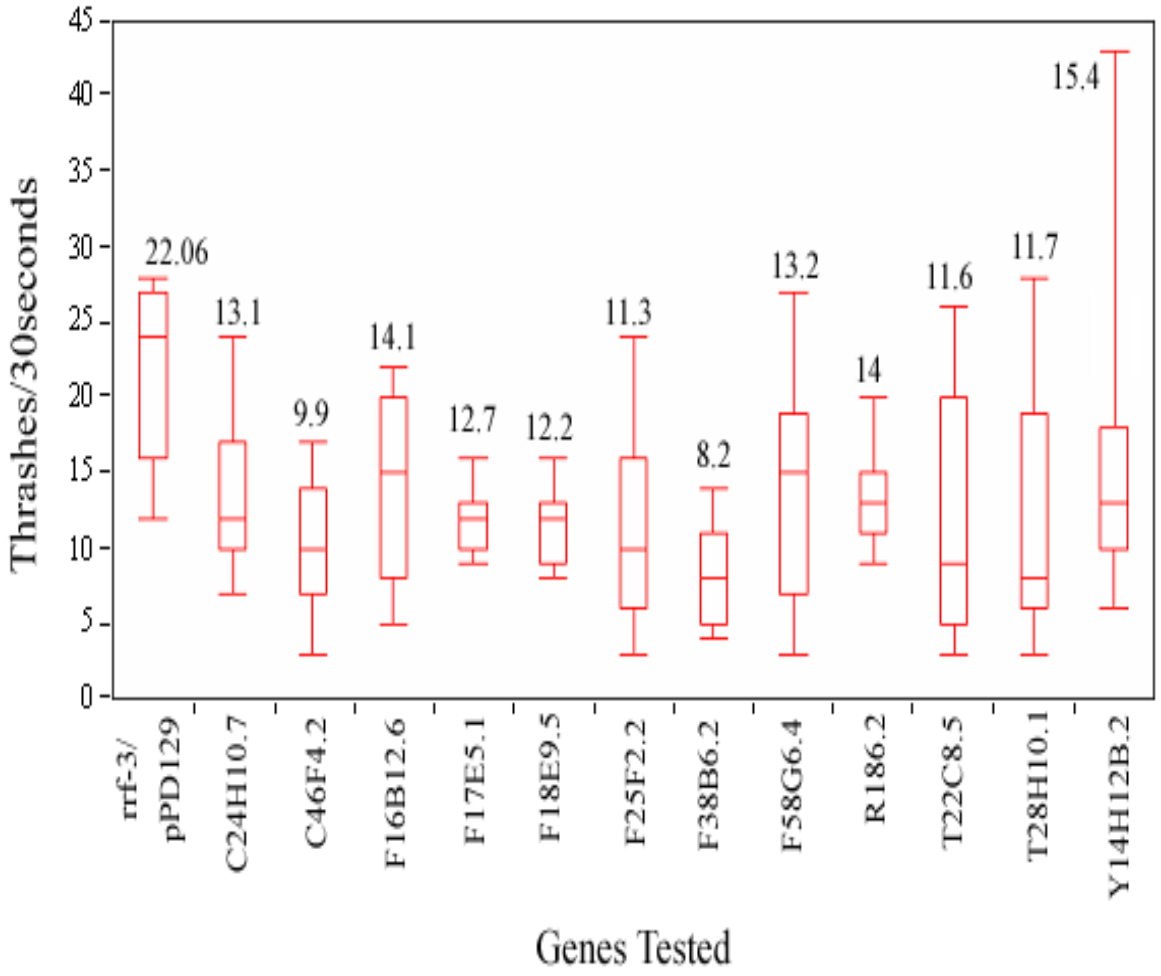

Figure 3.3. Results of thrashing assay of 12 new genes associated with Unc phenotype in comparison to *rrf-3*;pPD129. Number indicates average thrashes/30 seconds, boxes indicate interquartile range of thrashing frequency (N=15. α =0.05) and horizontal line indicates median.

Table 3.6. Summary of known uncoordinated movement RNAi phenotype genes and homology identified genes, which share high sequence similarity to known RNAi genes. Summary includes names of sets of genes, their chromosome number (chr.#) and description.

Previously Reported RNAi Phenotype-Associated Genes	Chr. #	Description	Known RNAi Phenotypes	Corresponding Putative RNAi Phenotype-Associated Genes	Chr. #	Description
Y38F2AL.1	IV	Encodes a claudin homolog that may be required for normal cohesion of apical junctions in epithelia	Dpy, Unc, Egl	C24H10.1	X	Encodes a claudin homolog that may be required for normal cohesion of apical junctions in epithelia
F37C12.7	III	Acyl-CoA synthetase	Stp, Unc, Rup	C46F4.2	X	Fatty Acid CoA Synthetase family
CD4.4	V	Uncharacterized conserved protein	Clr, Gro, Mlt, Unc, Emb, Lva	F38B6.2	X	Unnamed protein
T03F1.8	I	Guanylate kinase	Gro, Pvl, Unc	F17E5.1	X	Guanylate Kinase, abnormal cell lineage
D2021.1	X	TPR repeat ; DNA-binding protein jumonji/RBP2/SMCY, contains JmjC domain	Emb, Gro, Sma, Unc, Pvl	F18E9.5	X	DNA-binding protein jumonji/RBP2/SMCY, contains JmjC domain
Y48G1C.7	I	Unnamed protein	Pvl, Unc , Gro, , Stp, Lva	F16B12.6	X	Unnamed protein
F41C3.4	II	Got1-like protein	Clr, Lvl, Unc, Lva, Emb	R186.2	V	Serpentine Receptor, class D (delta)

Table 3.6 Continued.

Previously Reported RNAi Phenotype-Associated Genes	Chr. #	Description	Known RNAi Phenotypes	Corresponding Putative RNAi Phenotype-Associated Genes	Chr. #	Description
T04G9.4	X	Alpha-aminoadipic semialdehyde dehydrogenase-phosphopantetheinyl transferase	Emb, Sck, Unc, Gro, Prz	T28H10.1	V	Alpha-aminoadipic semialdehyde dehydrogenase-phosphopantetheinyl transferase
Y71D11A.5	III	Ligand-gated ion channel	Gro, Sck, Lvl, Ste, Unc, Emb	F58G6.4	IV	Acetylcholine-gated chloride channel
R10F2.1	III	CaDHerin family	Gro, Lvl, Prz, Unc	F25F2.2	III	CaDHerin family
C16A11.4	II	PHD Zn-finger proteins	Unc, Clr	Y14H12B.2	II	Uncharacterized protein
Y40B1A.4	I	Zn-finger	Bmd, Unc, Dpy, Pvl, Rol, Ste, lva, Rup, Stp,	T22C8.5	II	Zinc finger, C2H2-like

Eight of the twelve genes such as C24H10.1, C46F4.2, F17E5.1, F18E9.5, T28H10.1, F58G6.4, F25F2.2 and T22C8.5 appeared to possess similar protein functions to their previously reported Unc gene pairs. C24H10.1 encodes a caludin homolog that may be required for normal cohesion of apical junctions in epithelia. Apical junction is essential for cell polarization, tissue differentiation and physiology. Caludins are integral membrane proteins with four transmembrane sequences that are found in mammalian tight junctions (Tsukita and Furuse 2000). Nematode four pass transmembrane protein related to claudins, localizes to adherens junctions and contributes to cell adhesion and actin-plasma membrane association (Tepass 2003).

C46F4.2 belongs to fatty acid CoA synthetase family and their function is to hydrolyze the fatty acid. It might be possible that RNAi of this gene disrupts this pathway and more fatty acids are accumulated in the body of *C. elegans*, leading to uncoordinated movement.

F17E5.1 or lin-2 encodes a protein belonging to the membrane associated guanylate kinase (MAGUK) family, with several domains (L27, PDZ, SH3, and guanylate kinase) considered to assemble specific multiprotein complexes in particular regions of the cell; *in vivo*, LIN-2 is required for the correct localization of LET-23 (and, presumably, other membrane proteins) to specific regions of the plasma membrane (Horvitz and Sulston 1980; Kaech *et al*. 1998).

F18E9.5 encodes two isoforms of histone H3 trimethyllysine-27 (H3K27me3) demethylase. H3K27me3 demethylation regulated by UTX/JMJD3 proteins is essential for proper development and contains Jumonji C (Jmjc) domain like its similar pair D2021.1.

The domain organization of F18E9.5 is more similar to human jumonji domain containing 3 histone lysine demethylase (JMJD3) which also has a sole Jmjc domain, than to human ubiquitously transcribed tetratricopeptide repeat, X chromosome and Y chromosome respectively (UTX and UTY) (Agger *et al*. 2007). R186.2 is serpentine receptor class D delta. Serpentine receptor is G-protein coupled receptor and it has 7TM chemoreceptor domain (Robertson and Thomas 2006). Chemosensory neurons make up about 10% of the *C. elegans* nervous system and it is believed that 1000 orphan G protein-coupled receptors act in chemosensation (Bargmann 1998). T28H10.1 is alpha-aminoadipic semialdehyde dehydrogenase-phosphopantetheinyl transferase and its molecular function is magnesium ion binding.

F58G6.4 is classified as ligand-gated ion channels (acetylcholine gated ion channel) which are neurotransmitter receptors that open and desensitize rapidly and this particular capability makes them ideal for short term signaling (Bargmann 1998). At the nematode neuromuscular junction, excitatory motor neurons release acetylcholine, which helps in muscle contraction (Stretton Aow 1985; Jorgensen *et al*. 1995). One more gene F25F2.2 has the same function to its similar pair R10F2.1. Both are related to Cadherin family. F25F2.2 is a homolog of the cadherin superfamily that is involved in cell-cell adhesion (Hill *et al*. 2001). It is expressed in intestinal muscles, Vulval muscle 1 (vm1), eighth pharyngeal muscle cell layer (pm8), head neurons, spermatheca and rectal epithelial cells. It is also homologous to human FAT, a cadherin superfamily, with the characteristic cadherin repeats in the extracellular domain (Wang *et al*. 2006). T22C8.5 is a transcription factor with zinc-finger domain (Okkema and Krause 2005) and expressed in intestine and two nerves in the head (John S Reece-Hoyes 2007).

We have performed the RNAi on two categories of phenotypes i.e. sterility and uncoordinated movement. The high protein sequence similarity of putative RNAi-associated genes with known phenotype associated genes necessitates the *in vitro* verification of results. While Western blot techniques provide the most comprehensive means to demonstrate the down regulation of gene expression, the limited availability of antibodies to desired proteins makes this method less palatable than sub-quantitative techniques like RT-PCR.

To verify the specificity and to identify the possible secondary targets, we performed the RT-PCR on all the eight positive sterility genes. RNAi animals were collected and the levels of the target mRNA and possible secondary targets were measured by RT-PCR. Results (Figure 3.4) indicated that in each case, there was a significant and specific decrease in expression only in the genes targeted by RNAi (compare lanes 1 and 4 in Figure 3.4 images a-h) but not in the possible secondary target (compare lanes 2 and 5 in Figure 3.4 images a-h), suggesting that the sterility of eight genes is the target specific phenotypes.

Control Set of Genes

Table 3.7 summarizes the list of 17 control genes (obtained from a script written in Perl program as mentioned in "Materials and Methods" section), their chromosome number, known RNAi phenotypes, corresponding putative RNAi phenotype associated gene and their chromosome number. The table also describes the results of RNAi for Ste and Unc phenotypes for the corresponding putative RNAi phenotype associated genes.

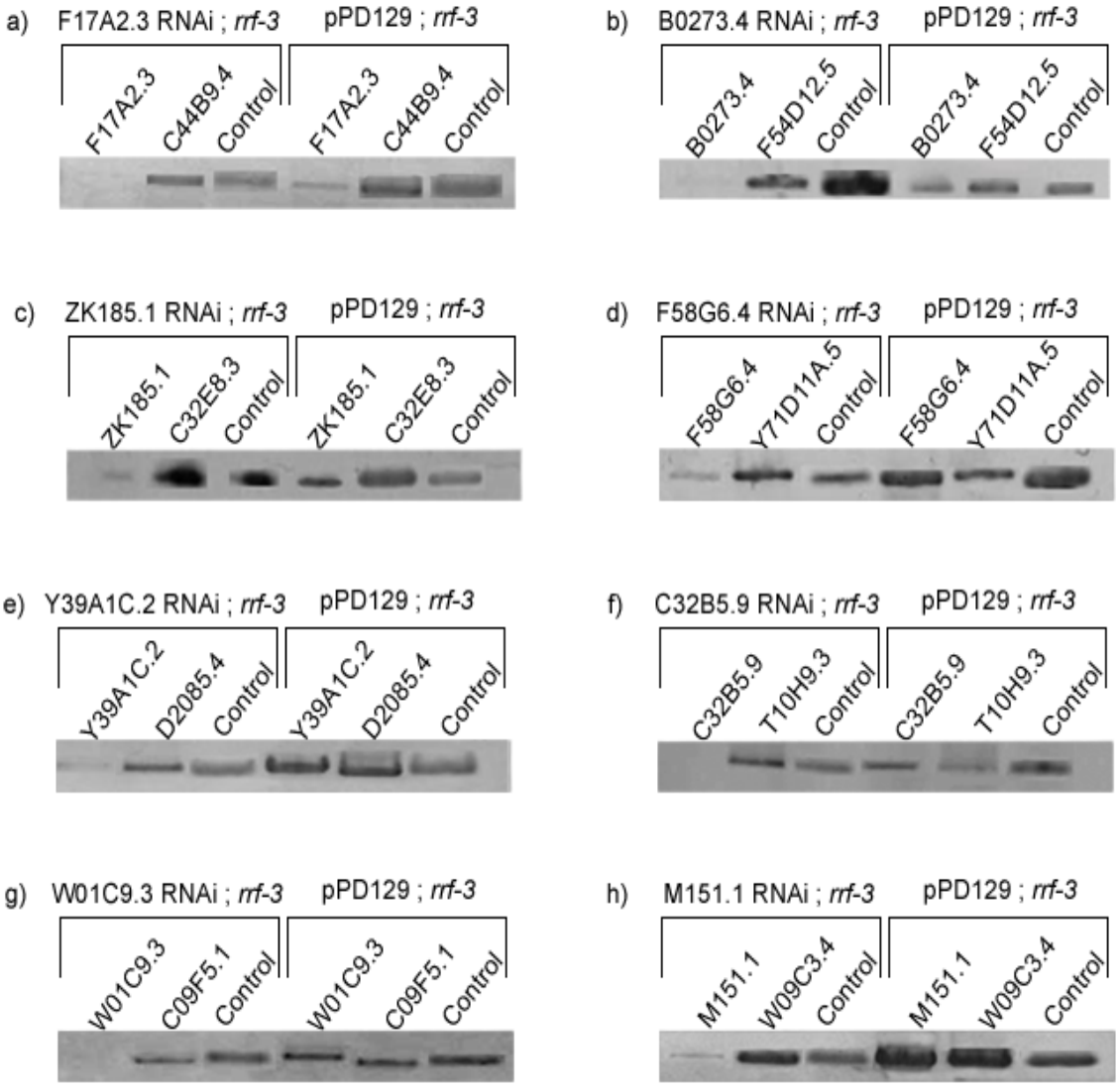

Figure 3.4. RT-PCR of putative Ste phenotype-associated genes.Images a – h, the first and fourth lanes represent the RNAi gene of interest, second and fifth lanes represent high-scoring similar gene of known phenotype, third and sixth lane genes represent control. The cDNA synthesis was carried out for first three lanes from the *rrf-3* worms treated with RNAi gene which is shown in first lane of every picture and for the last three lanes is from *rrf-3;* pPD129 treated worms.

Table 3.7. Summary of control genes phenotypes tested. This table depicts the name of known RNAi phenotype genes, their chromosome numbers (Chr. #), homology identified genes, which share high sequence similarity to known RNAi genes; and Chr. # and their association with Ste and Unc phenotypes.

Previously Reported RNAi Phenotype-Associated Genes	Chr. #	Known RNAi Phenotypes	Corresponding Putative RNAi Phenotype-Associated Genes	Chr. #	Ste	Unc
Y52B11B.1	I	Dpy	ZK816.1	X	No	No
T23B12.3	V	Clr, Gro, Sck,Lva, Emb	F38B6.4	X	No	Yes
C29H12.5	II	Egl	T12E12.2	IV	No	No
F46E10.11	V	Gro, Lvl, Prz	R02F11.1	V	No	No
C15H9.10	X	Lva	T26C5.1	II	No	No
Y102A5C.3	V	Lva	ZK1290.9	II	No	No
T08B2.8	I	Lva, Gro, Emb	ZK488.1	V	No	No
Y53H1C.2	I	Bli, Lva, Pvl, Rup, Sck	R10E12.1	III	No	No
F21D5.5	IV	Gro, Lva	F26A1.14	III	No	No
C08F1.3	II	Emb, Gro , Sma	T08E11.6	II	No	No
Y47D3B.1	III	Egl, Let, Lva	ZK1025.7	I	No	No
B0432.12	II	Gro	F47C12.4	IV	No	No
T28F2.2	I	Gro	T04H1.9	V	No	No
W04C9.3	I	Lvl	T04F8.4	X	No	No
Y71H2AL.2	III	Gro, Lva	F55A4.2	X	No	No
F10E9.7	III	Emb, Gro	ZK643.2	III	No	No
Y49A3A.1	V	Emb, Lon	F22E10.5	X	No	No

RNAi was performed on 17 random set of genes and found none of them was associated with Ste phenotype. Sterility data for the control set is provided in Table 3.8.

Table 3.8. Sterility score for control set of genes (N is number of worms).

Gene name	% Sterility (N =25)
ZK816.1	0
F38B6.4	0
T26C5.1	0
T12E12.2	0
R02F11.1	0
ZK1290.9	0
ZK488.1	0
R10E12.1	0
F26A1.14	0
T08E11.6	0
ZK1025.7	4
F47C12.4	0
T04H1.9	0
T04F8.4	0
F55A4.2	4
ZK643.2	0
F22E10.5	0

All these 17 genes were also tested for Unc phenotype and only one gene F38B6.4 was found positive. Rest of them were not associated with Unc phenotype. To analyze the Unc phenotype, their thrashings were analyzed as compared to *rrf-3*; pPD129 control as described in "Materials and Methods". The graph (Figure 3.5) represents the number of thrashings presented by box and whisker plots (JMP release 6 software). It depicts the thrashing pattern of 17 genes and control *rrf-3*; pPD129. In the graph, each box represents an interquartile range (IQR, the central 50% of the data points). Number indicates the average thrashes per 30 seconds and horizontal line in the boxes indicate medians.

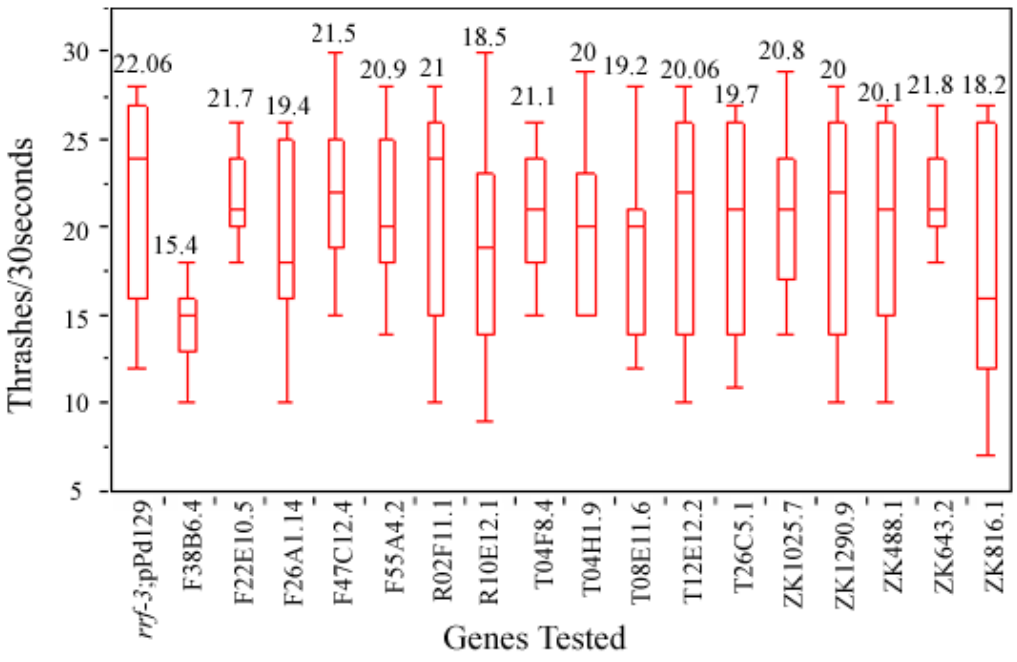

Figure 3.5. Uncoordinated phenotype tested in control set of genes compared to *rrf-3*; pPD129. One gene F38B6.4 was found positive as Unc, rest 16 were not associated with Unc phenotype. Number indicates average thrashes per 30 seconds. Boxes indicate interquartile range of thrashing frequency (N=15, α =0.05).

CHAPTER FOUR

Discussion

With the ultimate goal of developing new and effective methods for diagnosing and treating diseases, biomedical researchers are studying biological processes and pathogenesis of diseases using biotechnology techniques. Several studies have been conducted to identify roles of genes in pathogenesis of diseases. Ethical and practical limitations of doing research on human genes have led to the extensive use of model organisms for understanding biological processes and genes underlying human diseases.

It has been more than 30 years since Sydney Brenner documented the use of *C. elegans* as a model organism (Brenner 1974). Since then, thousands of articles have been published on its behavior, neurobiology, aging, etc. *C. elegans* is established as a model organism not only for general biological processes but also for elucidating the pathogenesis of diseases. Striking similarity with human genome and distinct features of *C. elegans* (short life span, small reproductive cycle, small body plan, easily observable mutant phenotypes and ease of cultivation in lab) make it an ideal choice as an experimental organism.

Last decade is considered to be a decade of genome mining after hundreds of genome sequences of different organisms were published. In the present decade of genome mining, large scales studies are being done to elucidate the functions of genomic content. In the *C. elegans* community also, efforts have been made to elucidate the function, regulation, interaction and expression of the entire complement of genes in the genome (Hillier *et al*. 2005).

To understand the complex network of genes, we need computational tools and ability to apply them to the various levels of data gathered from different genome studies. Bioinformatics tools such as BLAST can help in integrating the data from different RNAi studies and other genome projects. Previous RNAi studies have been able to define phenotypes of just 14% of total *C. elegans* genome using various techniques in laboratory. There is a challenge to define phenotypes to the rest of 86% genome using laboratory techniques. Thus a need to use new computational tools for identifying the phenotypes of these genes was observed.

In this study, we used the computational tool like BLAST to find the similarity in genes and predicted the phenotypes of new genes. In the first step, two sets of *C. elegans* genes were made. One consisted of all the genes in *C. elegans* genome (19,920 genes) and other consisted of all the genes associated with RNAi phenotypes (2807 genes). In the next step, bi-directional pairwise BLAST was performed to find the genes which were similar in sequence. On the basis of sequence similarity, 141 genes of unknown RNAi-associated phenotype were found to possess putative similarity to the initial set. Result was segregated into sixteen different phenotypes. Out of the 16, only two categories of phenotypes, Ste and Unc, were selected to study the association of putative genes with predicted phenotypes in laboratory.

Out of 141 unique genes, 49 genes were identified to be associated with Ste phenotype. However, due to limited cDNA collection, only 37 putative genes available in library were selected to perform RNAi by feeding. As a result of RNAi studies, we assigned sterile phenotype to 8 new genes.

A high success rate of 22% was observed in identifying sterility phenotypes in *C. elegans* based on sequence similarity. Five out of the eight new genes appeared to possess similar protein functions to their previously reported Ste gene pairs.

On the other hand, we identified 53 putative genes associated with Unc phenotype from BLAST studies. In this case also, based on limited availability of cDNA, only 29 putative Unc genes available in library were studied by RNAi feeding. Out of the 29 genes studied, uncoordinated phenotype was assigned to 12 new genes leading to a high success rate of 41%. Eight out of the new twelve genes appeared to possess similar protein function to their previously reported Unc gene pairs.

Functions of all the putative genes were not observed to be similar to their homologues. This could be due to one of the several reasons described herein. RNAi is a convenient method to study the loss of function phenotype in the gene of interest and it has been used in many species. However, results obtained from different studies are not certain because a number of experimental variables can influence the outcome of RNAi assay. Examples of the probable factors influencing the results include temperature, choice of dsRNA, and variation in penetrance level of RNAi feeding method.

Choice of dsRNA can influence the outcome of RNAi assay. The chromosome I and III studies both used amplified genomic DNA and other studies used the cDNA clones from libraries. A recent comparison of predicted genes with the experimentally validated genes showed that about 9% of the genes were erroneously predicted as joined together or split into multiple genes (Reboul *et al*. 2001). Use of cDNA libraries avoids the potential problems of gene misidentification.

However, cDNA libraries have the limitation of being incomplete (Kamath *et al*. 2003). That is why we could not perform the RNAi on all the predicted Ste and Unc genes.

Different methods of dsRNA delivery to produce gene knockdown in *C. elegans* have their pros and cons too. Injection method appears to be useful for studying gene function in the embryos but appears to work less for genes expressed later during development compared to soaking or feeding of dsRNA. Soaking method requires large amount of dsRNA as compared to feeding method. Feeding method was selected in this study since it can be used at any stage of development and requires less amount of dsRNA as compared to soaking method. Also, feeding method is easy to use on a large number of worms and is less expensive. On the other hand, penetrance levels in feeding method seem to be more variable in several kinds of phenotypes although it has similar efficiency in generating phenotypes as comparison to injection and soaking methods.

Temperature could be another important factor leading to inconsistent results in RNAi assays. Both the rate of development and progression of RNAi phenotype can be affected by temperature (Maine 2001). Furthermore, genes have different sensitivities to RNAi. It is not possible to target every single gene by RNAi. Genes that encode proteins with long half lives can be difficult to target (Montgomery *et al*. 1998; Ngo *et al*. 1998).

Different tissues have different sensitivity to RNAi, for example, nervous system. However, this can be overcome by using RNAi hypersensitive strains like *rrf-3* (Simmer *et al*. 2003). That may be one of the possible reasons of getting high percentage of Unc phenotype in our study since *rrf-3* was used.

One of the fundamental paradigms in computational biology is function prediction by homology or sequence similarity. Sequence similarity may be used as an evidence for homology but does not necessarily mean homology (Wei *et al*. 2002). We found positive results on the basis of sequence similarity. There are reports (Zhong-Hui Duan 2006) which conclude that protein pairs of similar biological functions tend to have higher sequence similarity and this can serve as a key measure in protein function prediction. However, sequence similarity must not be the only criteria to predict the function. The resulting observations must be verified through other means like gene expression pattern and functional assay etc. It is always important to assess the biological significance of a sequence alignment. This may lead to find the evidence for similar cellular function, a common overall structure or if possible a similar three dimensional structure.

Protein structure can be defined at four different levels as primary, secondary, tertiary and quaternary. Functionally important areas such as ligand binding sites or enzymatic sites are formed at tertiary and quaternary levels (Pevsner 2003). However, relative to tertiary structure, secondary structure predictions are easier to make. Though it is commonly found that similar sequences have same three dimensional structures but even a single amino acid mutation can change the conformation of protein. It is possible that due to conformation change in structure, the ligand binding site of protein also undergoes conformational change and hence the change in function (Watson *et al*. 2005).

Another possible reason of having similar sequence and different functions is variation in protein folding. Folding depends upon the amino acids in a polypeptide chain.

Amino acids can be hydrophobic, hydrophilic or electrically charged and their interaction with each other and surroundings can produce different folds and different three dimensional structures in spite of having similar sequences (Watson *et al*. 2005).

Further, domain shuffling (Doolittle and Bork 1993; Doolittle 1995) also complicates the matters. Domain is the region of protein that can adopt a particular three dimensional structure. A presence or absence of a domain can have a great impact on the protein molecular function.

Protein families evolve a multiplicity of functions through gene duplication, speciation and other processes (Sjolander 2004). Evolution not only conserves function, it also generates new functions. Gene duplication is the single most contributing factor for diversification of gene functions (Fitch 1970). Paralogus genes, related by duplication events are likely to have more different functions in comparison to orthologus genes (Sjolander 2004). However, it is not impossible in the case of paraloges and it is possible that genes having similar RNAi phenotypes are not under great selective pressure to change. For example, For example, Hemoglobin genes (A, A2, S and F) are paraloges to each other and diverged slightly in function but they have maintained their basic function of oxygen transport (http://en.wikipedia.org/wiki/Homology_(biology). Conversely, recent studies demonstrated that redundant gene pair functions in genes duplicated before the divergence of *C. elegans* and *C. briggsae* suggested that there has been selective pressure to maintain at least some overlap in functions (Tischler *et al*. 2006).

About 60% of *C. elegans* genes have been reported to possess similarity to a human gene (Harris *et al*. 2004). Also, approximately 40% of human disease causing genes have homologues in the *C. elegans* genome (Culetto and Sattelle 2000). These facts indicate that there exists a probability of sequence similarity of 8 new sterile genes and 12 new uncoordinated genes to have homologues in human genome.

Bioinformatics techniques were successfully utilized in this project for identification of phenotypes on the basis of sequence similarity with a relatively high success rate of 22% and 41% for sterility and uncoordinated phenotypes respectively. Only two out of the 16 categories of phenotypes identified by BLAST were studied in this study. We have tried to find few genes' phenotypes in this work and moved a bit closer to understand *C. elegans* genome biology. Same techniques can further be used to study remaining 14 categories of RNAi phenotypes and to assign phenotypes to rest of the genes. Simultaneously, further research can be done on a particular gene of interest such as the expression site and their involvement in different biological pathways. There could be various causes of one phenotype such as in the case of Ste and Unc phenotypes. Further research on particular genes might give important information about genes acting on different pathways to produce a single phenotype.

APPENDICES

APPENDIX A

Table A. *C. elegans* Homologues of Positionally Cloned Genes Mutated in Human Diseases (Culetto and Sattelle 2000).

Human Disease (OMIM Accession no.)	Human Gene (GenBank Accession no.)	*C. elegans* ORF	BLASTp (P value)
Aarskog-Scott syndrome (305400)	FGD1 (U11690)	C33D9.1	3.2×10^{-56}
Achondroplasia (100800)	FGFR3 (M58051)	F58A3.2	1.7×10^{-89}
Adenomatous polyposis coli (175100)	APC (M74088)	K04G2.8A	3.0×10^{-33}
Adrenoleukodystrophy, X-linked (300100)	ALD (Z21876)	T02D1.5	1.5×10^{-196}
Alzheimer disease, type 3 (104311)	PS1 (L76517)	F35H12.3[a]	2.4×10^{-79}
Alzheimer disease, type 4 (600759)	PS2 (L44577)	F35H12.3	1.4×10^{-81}
Amyotrophic lateral sclerosis (105400	SOD1 (K00065)	C15F1.7	3.2×10^{-45}
Aniridia (106210)	PAX6 (M77844)	F14F3.1	4.6×10^{-89}
Ataxia telangiectasia (208900)	ATM (U26455)	B0261.2	7.2×10^{-42}
Autoimmune polyglandular syndrome (240300)	AIRE (O43918)	F26F12.7	2.2×10^{-09}
Barth syndrome (302060)	BTHS (X92762)	ZK809.2	4.4×10^{-54}
Benign familial neonatal convulsions (121200)	KCNQ2 (AF033348)	C25B8.1	2.3×10^{-102}
Bloom syndrome (210900)	BLM (U39817)	T04A11.6	4.4×10^{-116}
Bor syndrome (600257)	EYA1 (Q99502)	C49A1.4	4.2×10^{-17}
Breast and ovarian cancer, early onset (113705)	BRCA1 (U14680)	C36A4.8	7.5×10^{-11}
Bruton agammaglobulinaemia (300300)	BTK (U78027)	M79.1C	3.6×10^{-82}
Carnitine deficiency, primary (212140)	OCTN2 (AB016625)	F52F12.1	4.8×10^{-74}
Ceroid lipofuscinosis, infantile neuronal (256730)	INCL (U44772)	F44C4.5	1.1×10^{-81}
Chediak-Higashi syndrome (214500)	CHS (U67615)	VT23B5.2	3.3×10^{-100}
Chondrodysplasia punctata (302950)	ARSE (X83573)	D1014.1	1.6×10^{-46}

Table A Continued.

Human Disease (OMIM Accession no.)	Human Gene (GenBank Accession no.)	*C. elegans* ORF	BLASTp (P value)
Choroideraemia (303100)	CHM (X78121)	Y57G11C.1	5.2×10^{-42}
Chronic granulomatous disease (306400)	NCF1 (M55067)	Y116A8C.36	3.7×10^{-09}
Citrullinaemia, adult onset type II (603471)	SLC25A13 (AF118838)	K02F3.2	4.8×10^{-177}
Coffin-Lowry syndrome (303600)	RSK2 (P51812)	T01H8.1A	1.1×10^{-223}
Congenital adrenal hyperplasia (201910)	CYP21 (M26856)	F44C8.1	5.9×10^{-37}
Congenital nephrotic syndrome 1 (256300)	NPHS1 (AF035835)	C26G2.1	1.5×10^{-61}
Cyclic haematopoiesis (162800)	ELA2 (P08246)	C07G1.1	3.9×10^{-13}
Cystic fibrosis (219700)	CFTR (M28668)	F21G4.2	1.4×10^{-135}
Darier disease (124200)	ATPLA2 (P16615)	K11D9.2	0.0
Deafness 3, conductive (304400)	POU3F4 (X82324)	K02B12.1	1.1×10^{-60}
Deafness, autosomal recessive 9 (601071)	OTOF (AF107403)	F43G9.6	1.3×10^{-72}
Deafness, neurosensory, autosomal recessive 3 (600316)	MYO15 (AF053130)	T10H10.1	3.1×10^{-200}
Deficiency of coagulation factors V/VIII (227300)	ERGIC-53 (P49257)	K07A1.8	7.5×10^{-85}
Dent disease (300009)	CLCN5 (X91906)	C07H4.2	2.1×10^{-169}
Diastrophic dysplasia (222600)	DTD (U14528)	K12G11.2	1.8×10^{-76}
Duchenne muscular dystrophy (310200)	DMD (M18533)	F38B4.3	1.2×10^{-145}
Duncan disease (308240)	SH2D1A (O60880)	M79.1C	1.3×10^{-09}
Dyskeratosis congenita, X-linked (305000)	DKC1 (O60832)	K01G5.5	1.6×10^{-146}
Epidermolytic palmoplantar keratoderma (144200)	KRT9 (X75015)	W10G6.3	9.1×10^{-23}

Table A Continued.

Human Disease (OMIM Accession no.)	Human Gene (GenBank Accession no.)	*C. elegans* ORF	BLASTp (P value)
Fragile histidine triad (601153)	FHIT (U46922)	Y56A3A.13	7.8×10^{-28}
Fragile site mental retardation, type 2 (309548)	FMR2 (U48436)	F35A5.1	1.6×10^{-05}
Friedreich ataxia (229300)	FRDA (U43747)	F59G1.7	3.5×10^{-23}
Glaucome primary open angle (137750)	GLC1A (Z97171)	C48E7.4	2.9×10^{-24}
Glycerol kinase deficiency (307030)	GK (L13943)	R11F4.1	3.6×10^{-133}
Gonadal dysgenesis (306100)	SRY (L08063)	K08A8.2	2.4×10^{-31}
Groenouw granular dystrophy, type 1 (122200)	BIGH3 (M77349)	F26E4.7	3.0×10^{-09}
Hereditary megaloblastic anaemia (261100)	CUBN (AF034611)	ZC116.3	2.0×10^{-225}
Hereditary multiple exostoses (133700)	EXT1 (U70539)	F12F6.3	5.2×10^{-90}
Hereditary non-polyposis colon cancer (120436)	MLH1 (U07418)	T28A8.7	2.5×10^{-107}
Hereditary non-polyposis colon cancer (120436)	MSH2 (U03911)	H26D21.2	1.1×10^{-99}
Hereditary pancreatitis (276000)	TRYP1 (U70137)	C07G1.1	7.6×10^{-05}
Hermansky-Pudlak syndrome (203300)	HPS (U65676)	F53H8.1	2.7×10^{-133}
Holt-Oram syndrome (142900)	TBX5 (Y09445)	F21H11.3	1.2×10^{-61}
Hyperekplexia (149400)	GLRA2 (X52009)	B0207.12	2.7×10^{-79}
Hypophosphataemic rickets, X-linked (307800)	XLH (U60475)	F18A12.8	2.6×10^{-91}
Kallmann syndrome (308700)	KAL (M97252)	K03D.10.1	7.1×10^{-34}
Lissencephaly (247200)	LIS1 (L13385)	T03F6.5[b]	9.5×10^{-127}
Long QT syndrome, type 1 (192500)	KVLQT1 (U40990)	Y54G9A.3	1.6×10^{-133}
Lowe syndrome (309000)	OCRL (M88162)	C16C2.3	2.4×10^{-57}

Table A Continued.

Human Disease (OMIM Accession no.)	Human Gene (GenBank Accession no.)	*C. elegans* ORF	BLASTp (P value)
Marfan syndrome (154700)	FBN1 (L13923)	ZK783.1	4.6×10^{-121}
Maturity onset diabetes of the young (600496)	TCF1 (X59869)	W03D8.4	4.5×10^{-23}
Menkes syndrome (309400)	MNK (X69208)	Y76A2A.2[c]	1.9×10^{-38}
Miyoshi myopathy (254130)	Dysferlin (AF075575)	F43G9.6[d]	1.4×10^{-82}
Mohr-Tranebjaerg syndrome (304700)	DDP (U66035)	Y39A3CR.E	2.5×10^{-13}
Multiple endocrine neoplasia 2A (171400)	RET (M57464)	F58A3.2	2.4×10^{-68}
Myotonic dystrophy (160900)	DM (L19268)	K08B12.5	4.8×10^{-121}
Myotubular myopathy (310400)	MTM1 (Q13496)	Y110A7A.5[e]	2.6×10^{-130}
Neurofibromatosis, type 1 (162200)	NF1 (M89914)	Z879.8	5.1×10^{-18}
Neurofibromatosis, type 2 (101000)	NF2 (L11353)	C01G8.5A	3.4×10^{-112}
Nevoid basal cell carcinoma syndrome (109400)	PTC (U59464)	ZK675.1	8.2×10^{-169}
Niemann-Pick C1 (257220)	NPC1 (NP000262)	F02E8.6[f]	1.6×10^{-133}
Nigmegen breakage syndrome (251260)	Nibrin (AF051334)	B0041.7	1.3×10^{-05}
Non-syndromic deafness DFNA1 (124900)	DIAPH1 (O60610)	F11H8.2	1.1×10^{-42}
Opitz syndrome (300000)	MID1 (Y13667)	ZK1320.6	1.9×10^{-10}
Pallister-Hall syndrome (146510)	GLI3 (M57609)	Y47D3A.7	1.3×10^{-54}
Pancreatic carcinoma (260350)	DPC4 (U4437)	R12B2.1	7.7×10^{-77}
Parkinson disease juvenile 2 (600116)	Parkin (AB009973)	K08E3.7	6.1×10^{-42}
Pendred symptom (274600)	PDS (O43511)	K12G11.2	2.1×10^{-68}
Polycystic kidney disease, type 1 (173900)	PKD1 (L33243)	ZK945	2.8×10^{-06}

Table A Continued.

Human Disease (OMIM Accession no.)	Human Gene (GenBank Accession no.)	*C. elegans* ORF	BLASTp (P value)
Polycystic kidney disease, type 2 (173910)	PKD2 (U50928)	Y73F8A	2.5×10^{-54}
Retinitis pigmentosa 2, X-linked (312600)	RP2 (AJ007590)	C54G6.2	8.8×10^{-06}
Retinitis pigmentosa 3, X-linked (312610)	RP3 (X97668)	F07C3.4	2.0×10^{-17}
Retinoblastoma (180200)	RB1 (M15400)	C32F10.2	1.3×10^{-10}
Retinoschisis X-linked juvenile	XLRS1 (AF014459)	C25F6.4	1.1×10^{-11}
Rieger syndrome (180500)	RIEG (U69961)	B0564.10	1.3×10^{-25}
Simpson-Golabi-Behmel syndrome (312870)	GPC3 (L47125)	F59D12.4	2.7×10^{-22}
Spinal muscular atrophy (253300)	SMN (U18423)	C41G7.1	3.0×10^{-08}
Spinocerebellar ataxia 1 (164400)	SCA1 (X79204)	K04F10.1	2.6×10^{-20}
Spinocerebellar ataxia 2 (183090)	SCA2 (U70323)	D2045.1	1.8×10^{-13}
Stargardt disease (248200)	ABCR (U88667)	Y39D8C.1	8.3×10^{-168}
Tangier disease (205400)	ABC1 (AF165306)	Y39D8C.1	3.6×10^{-180}
Thiamine-responsive megaloblastic anaemia (249270)	SLC19A2 (AJ238413)	F37B4.7	3.1×10^{-63}
Thomsen disease (160800)	CLC1 (Z25884)	C07H4.2	2.1×10^{-169}
Treacher-Collins syndrome (154500)	TCOF1 (U40847)	K06A9.1	1.6×10^{-07}
Tuberous sclerosis (191090)	TSC2 (X75621)	T27F2.2	3.1×10^{-13}
Waardenburg syndrome (193500)	PAX3 (U02309)	R08B4.2	1.0×10^{-23}
Werner syndrome (277700)	WRN (L76937)	F18C5.2[g]	1.6×10^{-72}
Wilms tumor (194070)	WT1 (X51630)	F53F8.1	2.9×10^{-27}
Wilson disease (277900)	WND (U11700)	Y76A2A.2	1.6×10^{-244}
Wiskott-Aldrich syndrome (301000)	WASP (U12707)	C07G1.4	1.0×10^{-11}

APPENDIX B

Table B. Thirty Three Human Disease Gene Homologues with an RNAi Phenotype (Rubin *et al*. 2000; Wood 2002; Kamath *et al*. 2003).

Predicted Gene	Locus of *C. elegans*	Human Disease	Human Gene	BlastP (E Value)	RNAi Phenotype
B0035.5	-	G6PD deficiency	G6PD	1×10^{-176}	Emb, Clr, Gro
B0350.2A	Unc-44	Hereditary spherocytosis	ANK1	0.00	Slu
C01G6.8	cam-1/kin-8	Insulin-resistant diabetes mellitus	INSR	6×10^{-55}	Unc, Pvl, clear patch
C01G8.5A	-	Neurofibromatosis	NF2	1×10^{-123}	Unc, Lvl, Gro
C06A1.1	-	Zellweger syndrome	PEX1	3×10^{-67}	Emb, Bmd, Sck, Gro
C07H6.7	lin-39	MODY, type IV	IPF1	5×10^{-14}	Egl, Vul, Muv
C17E4.5	-	Oculopharyngeal muscular dystrophy	PABPN 1	3×10^{-41}	Emb, Unc, Lva
C29A12.3	lig-1	DNA ligase I deficiency	DNA ligase1	1×10^{-167}	Emb
C48A7.1	egl-19	Long QT syndrome 3	SCN5A	2×10^{-64}	Egl, Clr
C50H2.1	-	Leydig cell hypoplasia	LHCGR	9×10^{-76}	Gro
D2045.1 2	-	Spinocerebellar ataxia	SCA2	7×10^{-09}	Emb
F01G10.1	-	Wernicke–Korsakoff syndrome	TKT	0.00	Emb, Clr, Gro
F07A5.7	unc-15	Tuberous sclerosis	TSC1	1×10^{-07}	Unc, Prz, Egl
F11C1.6 m	nhr-25	Pseudohyperaldosteronis	NR3C2	7×10^{-24}	Unc, Prz, Clr, Egl
F11H8.4	cyk-1	Nonsyndromic sensorineural deafness	DFNA1	9×10^{-49}	Emb, Adl, Rup, Clr
F20B6.2	vha-12	Renal tubular acidosis	ATP6B1	0.00	Emb, Ste, Adl, Lvl, Prz

Table B Continued.

Predicted Gene	Locus of *C. elegans*	Human Disease	Human Gene	BlastP (E Value)	RNAi Phenotype
F54D8. 1	-	Ehlers–Danlos syndrome, type IV	COL3A1	1×10^{-06}	Dpy
F53G12.3	-	Chronic Granulomatous Disease	X-CGD	3×10^{-34}	Bli, Mlt, Lvl
F58A3.2A	egl-15	Multiple venous malformations	VMCM	1×10^{-62}	Egl
K04G2.8A	apr-1	Adenomatous polyposis of the colon	APC	9×10^{-34}	Unc, Bmd, Lvl
K07A1.12	rba-2	Cockayne syndrome	CKN1	6×10^{-13}	Emb, Pvl, Lvl
M02A10.2	-	Hyperinsulinism	KCNJ11	4×10^{-78}	Unc
R107.8	lin-12	Alagille syndrome	JAG1	2×10^{-90}	Egl
R12B2.1	sma-4	Pancreatic carcinoma	MADH4	2×10^{-39}	Sma, Dpy
T03F6.5	lis-1	Miller–Dieker lissencephaly syndrome	PAF	1×10^{-148}	Emb
W05E10.3	ceh-32	Holoprosencephaly	SIX3	1×10^{-69}	Unc
W10G6.3	ifa-2	Keratoderma	KRT9	7×10^{-26}	Unc, Lvl, Mlt
Y47D3A.6A	tra-1	Grieg cephalopolysyndactyly syndrome	GLI	6×10^{-58}	Rup, clear patch
Y76A2A.2	-	Menkes disease	ATP7A	0.00	Prz, Adl, Unc
ZC506.4	mgl-1	Hypercalcemia	CASR	2×10^{-77}	Gro

APPENDIX C

Table C. List of 2807 Genes with Known RNAi Phenotypes in *C. elegans.*

Genes with Known Phenotypes in *C. elegans*						
AC7.1a	B0261.2a	B0395.2	C01F1.1	C04C3.3	C06E1.10	C08B11.6
AC7.2a	B0261.4	B0395.3	C01F1.2	C04F12.4	C06E4.6	C08B11.7
AC8.6	B0280.11	B0403.4	C01F1.3	C04F6.1	C06E7.1a	C08B6.9
AH6.5	B0280.12a	B0412.4	C01F6.4	C04F6.4a	C06E7.3a	C08C3.1a
B0001.7	B0280.1a	B0414.5	C01F6.8a	C04G2.6	C06G1.1	C08C3.4
B0024.14a	B0280.3	B0416.5a	C01F6.9	C04H5.6	C06G1.4	C08F1.3
B0024.4	B0280.9	B0432.12	C01G12.5	C05B5.4	C06G3.10	C08F8.1
B0025.1a	B0285.1	B0432.3	C01G6.3	C05C10.5a	C06G3.2	C08F8.2
B0025.2	B0285.7	B0464.1	C01G6.8a	C05C10.6a	C06H2.1	C08F8.8
B0035.10	B0286.4a	B0464.2	C01G8.5a	C05C12.1	C07A12.4a	C08H9.2
B0035.11	B0286.5	B0464.5b	C01G8.6	C05C12.3	C07A12.5a	C08H9.3a
B0035.12	B0303.15	B0464.7	C01G8.9a	C05C8.2	C07A9.2	C09D1.1a
B0035.14	B0304.1a	B0464.9	C01H6.5a	C05C8.6	C07A9.3a	C09D4.5
B0035.15	B0334.5	B0491.2	C02B10.5	C05C8.7	C07D10.2a	C09F5.1
B0035.5	B0336.10	B0491.5	C02B8.4	C05D10.2a	C07D8.2	C09F5.2
B0035.7	B0336.2	B0491.8a	C02C6.1a	C05D11.10	C07D8.3	C09G4.4
B0035.8	B0336.3	B0495.4	C02F4.2a	C05D11.12	C07E3.1a	C09G4.5
B0035.9	B0336.6	B0495.6	C02F5.1	C05D11.2	C07E3.2	C09G5.6
B0041.4	B0348.4a	B0495.7	C02F5.9	C05D11.3	C07F11.1	C09G9.6
B0041.7	B0348.5	B0511.10	C03A7.2	C05D11.9	C07F11.2	C09H10.10
B0205.6	B0350.2a	B0511.6	C03B1.12	C05D2.5	C07G1.5	C09H10.2
B0205.7	B0361.10	B0511.8	C03B8.4	C05E11.1	C07G2.3a	C09H10.3
B0207.4	B0361.2a	B0511.9a	C03C10.1	C06A1.1	C07H6.2	C09H10.6
B0207.6	B0361.5a	B0513.3	C03C10.3	C06A1.5	C07H6.3	C09H10.7
B0238.11	B0361.6	B0564.1a	C03D6.1	C06A5.1	C07H6.5	C09H10.8
B0244.8	B0361.8	B0564.1b	C03D6.3a	C06A5.3a	C07H6.6	C09H5.2a
B0250.1	B0365.3	C01A2.3	C03D6.8	C06A8.1a	C07H6.7	C10A4.7
B0250.3	B0365.6	C01A2.5	C03G5.1	C06A8.2	C08B11.1	C10C6.1
B0250.7	B0379.4a	C01B4.6	C03H5.2	C06A8.5	C08B11.3	C10C6.5
B0250.8	B0393.1	C01B7.1a	C04A11.3	C06B8.8	C08B11.4	C10E2.3
B0261.1	B0395.1	C01B9.1	C04A2.3a	C06C3.1a	C08B11.5	C10E2.6

Table C Continued.

Genes with Known Phenotypes in *C. elegans*						
C10F3.3	C14B9.4a	C16C10.6	C23G10.8	C26C6.5a	C28C12.8	C31C9.2
C10G11.1	C14B9.7	C16C10.8	C23H3.1	C26C6.5b	C28C12.9a	C31E10.7
C10G11.10	C14C10.2	C16D9.2a	C23H3.4a	C26D10.1	C28H8.11a	C31H2.2
C10G11.5a	C14C10.3a	C17C3.4	C23H3.5	C26D10.2a	C28H8.6a	C32A3.1a
C10G6.1a	C14C10.4	C17C3.5	C24A11.9	C26D10.3	C29E4.1	C32A3.2
C10G8.5a	C14C11.4	C17E4.5	C24D10.4	C26D10.5a	C29E4.2	C32D5.12
C10H11.1	C14F5.5	C17E4.9	C24G6.1	C26E6.11	C29E4.3a	C32D5.2
C10H11.10	C15C6.3	C17F4.5	C24G6.8	C26E6.4	C29E4.7	C32D5.3
C10H11.8	C15C6.4	C17G10.2	C24H11.7	C26E6.6	C29E4.8	C32E12.3
C10H11.9	C15C7.5	C17G1.6a	C24H11.9	C26E6.8	C29E6.1a	C32E12.4
C11E4.6	C15C8.4	C17H12.1	C24H12.5a	C26F1.3	C29E6.2	C32E8.10a
C11H1.3	C15C8.7	C17H12.14	C25A1.11	C26F1.4	C29E6.3	C32E8.2a
C11H1.4	C15F1.3a	C18A3.3	C25A11.2	C26F1.9	C29F3.7	C32E8.3
C12C8.3a	C15F1.4	C18A3.5a	C25A11.4a	C27A12.2	C29F4.1	C32E8.5
C12D8.10a	C15H11.7	C18A3.6a	C25A1.5	C27A2.2a	C29F9.12	C32E8.8
C13B9.3	C15H11.8	C18D1.1	C25A1.6	C27A2.6	C29F9.7	C32F10.2
C13B9.4a	C15H11.9	C18D11.4	C25A1.7	C27C7.5	C29F9.8	C32F10.5
C13D9.2	C15H7.4	C18E3.2	C25A1.9	C27D11.1	C29F9.9	C32F10.6
C13F10.4	C15H9.10	C18E3.3	C25D7.10	C27D6.1	C29H12.1	C33A12.1
C13G3.3a	C15H9.4	C18E3.5	C25D7.3	C27D6.3	C29H12.2	C33B4.3a
C13G5.1	C15H9.6	C18E9.10	C25D7.6	C27D9.1	C29H12.5	C33D3.1
C14A4.1	C15H9.8	C18E9.11a	C25F6.2b	C27F2.10	C30A5.7a	C33D9.1a
C14A4.11	C16A11.4	C18E9.2	C25G4.5	C27F2.4	C30B5.1	C33F10.5a
C14A4.14	C16A3.3	C18E9.3a	C25G4.6	C27F2.8	C30B5.4	C33F10.8
C14A4.2	C16A3.4	C18E9.4	C25H3.11	C27H5.4a	C30C11.1	C33H5.10
C14A4.4a	C16A3.5	C18E9.6	C25H3.6a	C27H6.2	C30C11.2	C33H5.15
C14A4.5	C16A3.6	C18H2.1	C25H3.8	C27H6.3	C30C11.4	C33H5.18a
C14B1.4	C16A3.8	C23F12.1a	C26B9.3	C28A5.3	C30D11.1	C33H5.4a
C14B1.5	C16A3.9	C23F12.2	C26C6.1a	C28A5.4	C30F8.2	C33H5.7
C14B9.1	C16C10.2	C23G10.3	C26C6.2	C28C12.10	C30G7.1	C33H5.9
C14B9.2	C16C10.3	C23G10.4a	C26C6.3	C28C12.2	C31B8.7	C34B2.10

Table C Continued.

Genes with Known Phenotypes in *C. elegans*						
C34B2.4	C36B1.4	C39F7.4	C44C1.1	C47G2.5	C52D10.7	C56C10.13
C34B2.5	C36B1.5	C40D2.3	C44C1.4a	C48A7.1a	C52D10.8	C56C10.3
C34B2.7	C36E6.3	C41C4.6	C44E4.4	C48A7.2	C52D10.9	C56C10.8
C34B2.8	C36E6.5	C41C4.8	C44H4.2	C48B4.9	C52E12.2a	C56E6.1
C34C12.8	C36E8.1	C41D11.1	C45B2.7	C48B6.2	C52E4.3	C56G2.1a
C34C6.4	C36E8.5	C41D11.2	C45G3.1	C48B6.6a	C52E4.4	C56G2.3
C34C6.6a	C36H8.1	C41D11.7	C45G9.11	C48D1.2	C53A5.1	C56G2.6
C34E10.1	C37A2.2	C41G7.1a	C45G9.5	C48E7.2	C53A5.3	CD4.3
C34E10.2	C37A2.4a	C41G7.2	C46A5.1	C48E7.3	C53A5.6	CD4.4
C34E10.4a	C37A2.7	C41G7.3	C46A5.3	C49C3.11	C53B7.4	CD4.6
C34E10.6	C37C3.2a	C41H7.4	C46A5.4	C49F5.1	C53C9.2	D1007.12
C34E7.4	C37C3.3	C42C1.10	C46A5.5	C49H3.11	C53D5.4	D1007.2
C34F11.3a	C37C3.6a	C42C1.14	C46C2.1a	C49H3.5a	C53D5.6	D1007.5a
C34F11.4	C37F5.1	C42C1.3	C46F11.4	C49H3.8	C53H9.1	D1007.6
C34F6.1	C37H5.5	C42C1.5	C46F9.4	C50A2.2	C53H9.2a	D1014.3
C34G6.1	C37H5.6a	C42D4.8	C46G7.1	C50B6.2	C54C6.1	D1037.4
C34G6.6a	C37H5.8	C42D8.5a	C46H11.6	C50C3.6	C54D1.5	D1043.1
C34H3.2	C38C10.4	C42D8.8a	C47B2.3	C50D2.1	C54D1.6	D1046.1
C35A11.4	C38C10.5a	C43E11.4	C47B2.4	C50E10.4	C54G10.2	D1046.2
C35B1.1	C38C3.5a	C43E11.9	C47B2.5	C50E3.5	C54G4.8	D1054.14
C35C5.1	C38C3.5b	C43G2.2	C47C12.3a	C50F2.3	C54H2.5	D1054.15
C35D10.13	C38C6.6	C43H8.1	C47D12.1a	C50F4.11	C55A6.2	D1054.2
C35D10.16	C38D4.3	C43H8.2	C47D12.6a	C50F4.2	C55A6.9	D1054.3
C35D10.1a	C39B10.1	C44B12.1	C47E12.1	C50F4.5	C55B7.5	D1069.3a
C35D10.4	C39E6.1	C44B12.2	C47E12.2	C50F4.7	C55B7.8	D1081.2
C35D10.5	C39E9.10	C44B12.5	C47E12.4a	C50F7.4	C55B7.9	D1081.8
C35D10.6	C39E9.11	C44B7.3	C47E12.5	C50H2.1	C55F2.1a	D2005.1
C35E7.8	C39E9.13	C44B9.2	C47E12.7	C52A11.2	C55F2.2	D2007.4
C36A4.4	C39E9.14a	C44B9.4	C47E8.4	C52B11.4	C56A3.4	D2013.5
C36B1.1a	C39E9.3	C44C10.1	C47E8.5	C52B9.7a	C56A3.8	D2013.6
C36B1.3	C39F7.1	C44C10.8	C47E8.7	C52D10.6	C56C10.12	D2013.7

Table C Continued.

Genes with Known Phenotypes in *C. elegans*						
D2021.1	E02H1.6	F07B7.10	F09E5.15	F10G8.3	F14F4.3a	F18F11.5
D2024.3	E03A3.3	F07B7.11	F09E5.2	F10G8.6	F14F7.1	F18G5.2
D2024.5a	E03H4.2	F07B7.3	F09E5.4	F11A10.2	F15B9.7	F18H3.5a
D2024.6	E03H4.8	F07B7.4	F09E5.5	F11A10.7	F15C11.1	F19B6.1a
D2024.8	E04A4.4a	F07B7.5	F09F3.5	F11A1.3a	F15D3.4	F19B6.2a
D2030.3	E04A4.5	F07B7.6	F09F7.2a	F11A3.2a	F15D3.6	F19C7.1
D2030.4	E04A4.7	F07B7.9	F09F7.3	F11C1.6a	F15D3.7	F19C7.7
D2030.9a	E04A4.8	F07C3.1	F09F7.4a	F11C7.5	F15E6.3	F19H6.1
D2045.1a	E04F6.4	F07D10.1	F09G2.4	F11E6.3	F15E6.6	F20B6.1
D2045.6	E04F6.5a	F07D3.2	F09G2.9	F11E6.5	F15E6.9	F20B6.2
D2045.7	EEED8.5	F07E5.5	F09G8.3	F11G11.10	F15G9.4a	F20B6.3
D2045.8	EGAP7.1	F07E5.9	F10B5.1	F11G11.12	F15H10.3	F20D12.1
D2045.9	F01F1.12a	F08B1.1a	F10B5.3	F11G11.8	F16A11.2	F20D12.2
D2085.1	F01F1.7	F08B12.2	F10B5.6	F11H8.1	F16B4.6	F20D12.4
D2085.3	F01F1.8a	F08B4.1a	F10C1.2a	F11H8.4a	F16B4.8	F20D6.9
D2085.4	F01G10.1	F08B4.5	F10C1.5	F12F6.6	F16D3.4	F20G4.1
D2085.6	F01G10.9	F08B4.6	F10C5.1	F12F6.7	F16F9.2	F20G4.3
D2089.1a	F01G12.2a	F08B4.7	F10D2.9	F13C5.2	F16H11.5	F20H11.2
D2089.2	F01G12.5a	F08B6.4a	F10D7.5a	F13D11.2	F17C11.10	F20H11.3
D2092.8	F01G4.2	F08C6.1a	F10E7.6	F13D12.7	F17C11.7a	F20H11.6
D2096.12	F01G4.6	F08C6.2	F10E7.7	F13D2.2	F17C11.9a	F21A3.7
D2096.8	F02A9.2	F08D12.1	F10E9.4	F13E6.2	F17E9.10	F21C3.5
DY3.2	F02A9.4b	F08F8.2	F10E9.5	F13E6.4	F17E9.12	F21D12.2
E01A2.2a	F02A9.6	F08G12.4	F10E9.6a	F13H10.4	F17E9.9	F21D5.5
E01A2.4	F02D10.1	F08G2.2	F10E9.7	F13H6.1	F18A1.2	F21D5.7
E01B7.1	F02D10.5	F08G2.3	F10E9.8	F13H8.2	F18A1.3a	F21D5.8
E01G4.2	F02E8.1	F09B12.1a	F10G2.2	F13H8.7	F18A1.3b	F21D9.2
E02A10.1	F02E9.2a	F09B9.2a	F10G7.1	F14B4.3	F18C12.2a	F21H11.3
E02D9.1b	F07A11.2a	F09D1.1	F10G7.2	F14D12.2	F18C12.3	F21H12.1
E02H1.1	F07A5.1a	F09E5.1	F10G7.4	F14D7.2	F18E2.2	F21H12.4
E02H1.4	F07A5.7	F09E5.11	F10G7.5	F14F11.2	F18E2.3	F22B3.1

Table C Continued.

Genes with Known Phenotypes in *C. elegans*						
F22B3.2	F23H12.5	F26B1.7	F28B12.3	F29G9.3	F32D1.2	F35C5.7
F22B3.4	F23H12.8	F26C11.3	F28B3.7	F29G9.4	F32D1.6	F35D6.1a
F22B3.8	F25B3.1	F26D10.3	F28B4.3	F29G9.5	F32D8.13	F35G12.10
F22B5.1	F25B3.6	F26D11.11a	F28C6.2	F30A10.1	F32D8.5a	F35G12.11
F22B5.2	F25B4.6	F26D2.2	F28C6.3	F30A10.10	F32D8.6	F35G12.2
F22B5.7	F25B4.9	F26E4.1	F28C6.6	F30A10.2	F32E10.1	F35G12.8
F22B5.9	F25B5.2	F26E4.4	F28C6.7a	F30A10.6	F32E10.4	F35G12.9
F22B7.13	F25B5.4a	F26E4.6	F28C6.8	F30A10.8a	F32E10.6	F35H10.11
F22B7.5a	F25B5.7a	F26E4.8	F28D1.1	F30A10.9	F32H2.1a	F35H10.4
F22D3.5	F25C8.3a	F26E4.9	F28D1.10	F30B5.1	F32H2.3	F35H10.5
F22D6.1	F25D7.1	F26F12.7	F28D1.11	F30B5.4	F32H2.6	F35H10.7
F22D6.10	F25D7.3	F26F4.1	F28D1.7	F30F8.8	F32H5.1	F36A2.3
F22D6.4	F25E5.1	F26F4.10a	F28D1.8	F30H5.1	F33A8.5	F36A2.6
F22D6.5	F25G6.2	F26F4.11	F28D9.1	F31B12.1a	F33C8.1a	F36A2.7
F22E5.12	F25G6.6	F26F4.7	F28F8.5	F31C3.2a	F33D11.10	F36A4.10
F22F4.1	F25H2.10	F26G1.1	F28H1.2	F31C3.5	F33D11.5	F36A4.7
F22F4.2	F25H2.11	F26H11.1	F28H1.3	F31D4.1	F33D4.1a	F36D3.1
F22G12.3	F25H2.2	F26H11.2c	F28H7.9	F31D4.3	F33D4.2a	F36F12.6
F23B12.3	F25H2.4	F26H9.6	F29A7.6	F31E3.1	F33D4.5	F36H1.2
F23B12.5	F25H2.5	F27C1.2a	F29B9.10	F31E3.3	F33D4.7	F36H1.4a
F23B12.7	F25H2.9	F27C1.3	F29B9.11	F31E3.5	F33E2.2a	F36H2.1a
F23C8.6	F25H5.4	F27C1.4	F29B9.5	F32A11.3	F33G12.4	F37A4.6
F23F1.1	F25H5.6	F27C1.6	F29B9.6	F32A5.1a	F33H1.2	F37A4.8
F23F12.6	F25H8.2	F27C1.7a	F29C12.4	F32A5.6	F33H1.3	F37B12.1
F23F1.5	F25H8.3	F27C1.8	F29C4.2	F32A5.7	F33H1.4	F37B12.3
F23F1.8a	F25H8.6	F27C8.6	F29C6.1	F32A7.6	F33H2.2	F37B12.4
F23F1.9	F25H9.6	F27D4.1	F29D11.1	F32B6.3	F33H2.5	F37C12.1
F23H11.2	F26A10.2	F27D4.2	F29D11.2	F32B6.6	F33H2.6	F37C12.11
F23H11.5	F26A3.2	F27D4.5	F29F11.5	F32B6.7	F33H2.8	F37C12.13a
F23H12.2	F26A3.3	F27D9.1a	F29G6.1	F32D1.1	F35A5.5	F37C12.14
F23H12.4	F26B1.3	F27E5.2	F29G6.3a	F32D1.10	F35C5.3	F37C12.2

Table C Continued.

Genes with Known Phenotypes in *C. elegans*						
F37C12.3	F40F8.10	F43C1.3	F45F2.2	F48D6.3	F53A2.4	F54C8.4
F37C12.4	F40F9.6a	F43C1.6	F45F2.3	F48E8.1a	F53A3.3	F54C8.5
F37C12.7	F40G9.1	F43D2.1	F45F2.4	F48E8.5	F53A3.7	F54C9.1
F37C12.9	F41B4.2a	F43D9.1	F45G2.4	F48F7.1	F53A9.10a	F54C9.2
F37C4.4a	F41C3.4	F43D9.3	F45G2.5	F48G7.5	F53B1.4	F54C9.4
F37D6.1	F41C6.1	F43D9.5	F45G2.8	F49C12.11	F53B2.1	F54C9.5
F37E3.1	F41C6.2	F43E2.7a	F45G2.9	F49C12.12	F53B3.1	F54C9.6a
F38A1.8	F41C6.3	F43E2.8	F45H10.2	F49C12.13	F53B8.1	F54C9.9
F38A5.10	F41D3.4	F43G6.9	F45H11.2	F49C12.8	F53E10.6	F54D11.1
F38A5.12	F41E6.13a	F43G9.1	F46A8.3	F49D11.1	F53E4.1	F54D12.4
F38A5.14	F41E6.4a	F43G9.10	F46A9.4	F49D11.5	F53F10.4	F54D12.5
F38A5.5	F41E7.1	F43G9.12	F46A9.5	F49E11.1a	F53F10.5	F54D1.6
F38A5.9	F41F3.4	F43G9.3	F46C8.6	F49E11.7	F53F1.2	F54D5.11
F38A6.1	F41G3.14	F43G9.5	F46E10.11	F49E2.1a	F53F4.10	F54D5.5a
F38B2.1a	F41H10.10	F44A6.2a	F46E10.1a	F49H12.1a	F53F4.11	F54D7.2
F38E11.5	F41H10.7	F44B9.7	F46E10.9	F52B10.1	F53G12.1	F54D8.1
F38E1.7	F41H10.8	F44C4.4a	F46F11.4	F52B11.2	F53G12.10	F54E12.1
F38H4.4	F42A6.7a	F44E5.1	F46F11.5	F52B11.3	F53G12.3	F54E12.3
F38H4.9	F42A8.1	F44F4.11	F46F11.9a	F52B11.4	F53G12.4	F54E12.4
F39B2.11	F42A8.2	F44F4.2	F46F2.2a	F52B5.6	F53G2.7	F54E12.5
F39B2.4a	F42A9.2	F44G4.1	F46G11.3	F52C6.12	F53H1.1a	F54E2.3a
F39B2.6	F42C5.10	F44G4.2	F46H6.1	F52C6.13	F54A3.3	F54E7.2
F39G3.7	F42C5.7	F45E12.2	F47A4.2	F52C6.2	F54B11.2	F54E7.3a
F39H11.2	F42C5.8	F45E12.3	F47B7.2a	F52C6.3	F54B11.3a	F54F2.1
F39H11.3	F42E11.4	F45E12.5a	F47B8.10	F52C6.4	F54B3.3	F54F2.2a
F39H11.5	F42G8.10a	F45E1.6	F47D12.4a	F52C9.7	F54C1.3a	F54F2.7
F39H2.2a	F42G8.12	F45E4.9	F47F6.1a	F52D10.6	F54C1.7	F54F2.8
F40F11.1	F42G8.6	F45E6.1	F47F6.1b	F52E10.5	F54C4.1	F54F7.5
F40F11.2	F42G9.7	F45F2.10	F47F6.5	F52E4.7	F54C4.3	F54G8.3
F40F11.3	F43B10.2	F45F2.12	F48A11.1	F52F12.6	F54C8.2	F54H12.1a
F40F12.7	F43C1.2a	F45F2.13	F48C1.4	F52H3.1	F54C8.3	F54H12.6

Table C Continued.

Genes with Known Phenotypes in *C. elegans*						
F54H5.4a	F56A11.1	F56H6.5	F58E2.9	F59G1.5	H27M09.1	K02A4.1
F55A11.2	F56A12.1	F56H6.8	F58E6.10	F59G1.7	H27M09.2	K02B12.1
F55A12.2a	F56A3.2	F57B10.1	F58F12.1	F59H6.12	H27M09.4	K02B12.3
F55A12.3	F56A3.3a	F57B10.10	F58F6.4	F59H6.3	H28O16.1a	K02B12.8
F55A12.7	F56A3.4	F57B10.3a	F58F6.5	H02I12.1	H31G24.4	K02B2.5
F55A12.8	F56A8.6	F57B1.2	F58G1.2	H02I12.6	H32C10.1	K02D10.5
F55A3.2	F56B3.1	F57B9.10a	F58G1.4	H02I12.7	H35B03.2a	K02D7.1
F55A8.1	F56B3.12	F57B9.2	F58G1.5	H04M03.1	H37A05.1	K02D7.3
F55B11.1	F56B3.2a	F57B9.3	F59A2.1a	H04M03.4	H37N21.1	K02D7.4
F55B11.2	F56B3.4a	F57B9.5	F59A2.3	H06A10.2	H38K22.1	K02E10.2a
F55B12.1	F56B3.8	F57B9.6a	F59A2.4	H06H21.3	H38K22.2a	K02F2.2
F55B12.4	F56B6.4a	F57C9.5	F59A3.1	H06I04.3a	H39E23.1a	K02F2.3
F55C10.2	F56C11.1	F57F4.3	F59A3.3	H06I04.4a	H41C03.1	K02F2.6
F55C10.3	F56C11.6a	F57F4.4	F59A6.4	H06O01.1	H43I07.2	K02F3.2
F55C5.4	F56C9.1	F57F5.1	F59A7.8	H12I13.4	JC8.10a	K02H11.2
F55C5.5	F56D1.1	F57G8.8	F59B10.1	H13N06.3a	JC8.2	K03A1.6
F55C5.8	F56D12.1a	F57H12.1	F59B2.3	H13N06.4a	JC8.5	K03B4.1
F55C7.7a	F56D1.3	F58A3.1a	F59C6.4	H14A12.2a	JC8.6a	K03B4.3a
F55D10.2	F56D1.4a	F58A3.2a	F59C6.5	H14N18.1a	K01A6.4	K03B8.4
F55D10.3	F56D1.7	F58A4.11	F59D8.2	H15N14.1c	K01B6.2	K03E5.3a
F55D12.4	F56D2.1	F58A4.3	F59E10.1	H15N14.2	K01C8.10	K03E6.7
F55F10.1	F56D2.6a	F58A4.4	F59E10.3	H17B01.4a	K01C8.6	K03H1.10
F55F8.2a	F56E10.4	F58A4.8	F59E12.11	H19M22.2a	K01C8.7	K03H1.2
F55F8.3	F56F3.1	F58A4.9	F59E12.12	H19M22.3a	K01C8.9	K03H4.2
F55F8.4	F56F3.2a	F58A6.1	F59E12.3	H19N07.1	K01G5.1	K04A8.6
F55F8.5	F56F3.5	F58B3.4	F59E12.4a	H19N07.2a	K01G5.10	K04C2.2
F55G1.10	F56F3.6	F58B3.5	F59E12.5a	H20J04.2	K01G5.4	K04C2.5
F55G1.2	F56F4.5	F58B6.2	F59F3.2	H20J04.5	K01G5.5	K04D7.1
F55G1.3	F56G4.4	F58B6.3a	F59F3.5	H24G06.1a	K01G5.7	K04D7.2a
F55G1.4	F56H1.3	F58B6.3b	F59F5.1	H24O09.2	K01G5.8a	K04D7.5
F55H2.2	F56H1.4	F58D5.1	F59G1.3	H25P06.2a	K02A11.1a	K04E7.2

Table C Continued.

Genes with Known Phenotypes in *C. elegans*						
K04F10.4a	K07A1.11	K08D12.3a	K10G6.1	M01F1.2	M195.2	R06A4.4a
K04G2.1	K07A1.12	K08E3.5a	K10G9.2	M01F1.3	M28.5	R06A4.7
K04G2.3	K07A1.2	K08E3.6	K11B4.1	M01F1.4a	M57.2	R06A4.9
K04G2.8a	K07A12.3	K08E4.1	K11C4.3a	M01F1.6	M6.1a	R06C1.2
K04G7.1	K07A12.5	K08E5.3a	K11C4.5	M01F1.7	M6.3	R06C1.3
K04G7.10	K07B1.6a	K08E7.3	K11D2.3	M01G5.5	M7.1	R06C7.10
K04G7.11	K07B1.7a	K08F11.4a	K11D9.1a	M02G9.1	M88.2	R06C7.5a
K04G7.4a	K07C11.2	K08F4.2	K11D9.2a	M03A1.1a	M88.5a	R06C7.8
K04H4.1a	K07C11.4	K08F8.5a	K11G12.1a	M03A8.1	M88.6a	R06F6.1
K04H4.2a	K07C5.1	K08F9.4	K11G9.4	M03C11.7	PAR2.4a	R06F6.2
K05C4.1	K07C5.4	K09A9.1	K11H12.2	M03D4.6	R01B10.1a	R07B1.1
K05C4.2	K07C5.6	K09A9.3	K11H3.2	M03F4.2a	R01B10.3	R07B5.9
K05C4.6	K07C5.8	K09A9.5	K11H3.6	M03F4.6	R01E6.3a	R07E4.6a
K06A4.6	K07D4.3	K09B11.2	K12C11.2	M03F4.7a	R02D3.3	R07E5.1
K06A5.4	K07D4.7a	K09E2.4a	K12D12.1	M03F8.3	R02D3.5	R07E5.10
K06A5.7	K07D8.1	K09E3.1	K12D12.2	M04B2.1	R02E12.8	R07E5.12
K06B4.1	K07E12.1a	K09E4.1	K12F2.1	M04B2.4	R02F2.7	R07E5.14
K06B4.2	K07E8.3	K09F6.6	K12H4.1	M04F3.1	R03E1.1	R07E5.3
K06B4.8	K07F5.14	K09H11.3	K12H4.3	M05B5.2	R03E1.2	R07E5.7
K06B9.2	K07F5.8	K09H9.3	K12H4.4	M106.1	R03G5.1a	R07G3.1
K06B9.4	K07F5.9	K09H9.6	K12H4.5	M106.5	R04B5.6	R07G3.3a
K06C4.10	K07H8.10	K10B2.1	K12H4.8	M110.4a	R04F11.2	R07H5.1
K06C4.11	K07H8.3	K10B2.5	LLC1.3	M110.5a	R05C11.3	R07H5.8
K06C4.12	K08A8.2a	K10B3.10	M01A10.2a	M117.2	R05D11.3	R08C7.10a
K06C4.13	K08A8.3	K10C2.4	M01A10.3	M142.4	R05D11.7	R08C7.2a
K06C4.15	K08B12.5	K10C3.6a	M01B12.3	M142.6	R05D3.1	R08C7.3
K06C4.2	K08B4.1a	K10D2.2	M01B12.5a	M163.4	R05D3.11	R08D7.1
K06C4.3	K08C7.3a	K10D2.3	M01D7.7a	M176.2	R05D3.4a	R08D7.2
K06C4.4	K08C9.2	K10D2.6	M01E10.2	M176.3	R05F9.1a	R08D7.3
K06C4.5	K08D10.3	K10D3.2	M01E11.6	M176.6a	R05H10.2	R09B3.1a
K06H7.6	K08D12.1	K10F12.3a	M01E5.5a	M18.5	R06A10.2	R09B3.4

Table C Continued.

Genes with Known Phenotypes in *C. elegans*						
R09B3.5	R12E2.3	T01B7.7	T04A8.5	T05H10.6a	T08A11.2	T10B9.4
R09E10.6	R12E2.5	T01C3.1	T04A8.6	T05H4.10	T08A9.9a	T10C6.11
R09F10.3	R13.1	T01C3.11	T04A8.7a	T05H4.12	T08B1.1	T10C6.12
R107.6	R13A5.12	T01C3.6	T04C10.2a	T05H4.4	T08B1.2a	T10C6.13
R107.8	R13A5.5	T01C3.7	T04C12.4	T05H4.5	T08B2.10	T10C6.14
R10D12.10	R13A5.8	T01C8.5	T04C12.5	T05H4.6	T08B2.5a	T10E10.1
R10E11.1a	R13F6.1	T01D1.2a	T04C12.6	T06A10.1	T08B2.7a	T10E10.2
R10E11.2	R13F6.10	T01E8.2	T04D3.2	T06A10.2	T08B2.8	T10E10.5
R10E11.8	R13F6.9	T01E8.3	T04D3.5	T06A1.4	T08B2.9a	T10E10.6
R10E4.2a	R13G10.1	T01E8.6	T04G9.3	T06D8.5	T08B6.3	T10E9.7a
R10F2.1	R144.2a	T01G9.4	T04G9.4	T06D8.6	T08D2.1	T10F2.1a
R10H10.1	R144.3	T01G9.6a	T05A10.3	T06D8.8	T08G11.4	T10F2.3
R10H10.2	R144.7	T01H3.1	T05A1.2	T06E6.1	T08G5.5	T10F2.4
R119.4	R144.9	T01H3.4	T05A6.1	T06E6.2a	T09A5.10	T10G3.6
R119.6	R148.7	T01H8.5	T05C12.10	T06G6.11	T09A5.11	T10H9.3
R11A5.2	R151.3	T02C12.2	T05C12.6a	T06G6.8	T09A5.5	T10H9.4
R11A5.7	R151.9	T02E1.3a	T05C12.7	T06G6.9	T09A5.6	T11B7.4d
R11A8.2	R160.1a	T02E1.5	T05D4.3	T07A9.11	T09A5.9	T11F8.3
R11A8.3	R166.4	T02E9.3	T05D4.4	T07A9.2	T09B4.10	T11F9.13
R11A8.6	R193.2	T02G5.7	T05E11.1	T07A9.6	T09B4.9	T11F9.9
R11D1.8	R31.1	T02G5.9a	T05E11.3	T07A9.8	T09E11.4	T11G6.1a
R11D1.9	R53.1a	T02H6.11	T05E11.5	T07A9.9a	T09E11.5	T12A2.2
R11E3.6	R53.3a	T02H6.3	T05E8.3	T07C4.1	T09E11.7	T12A2.7
R12B2.1	R53.4	T03E6.7	T05F1.1	T07C4.6	T09E8.1a	T12C9.7
R12B2.4	R53.6	T03F1.8	T05F1.3	T07C4.7	T09F3.3	T12D8.1
R12B2.5a	R53.7a	T03F1.9	T05G5.10	T07F8.3a	T10B11.9	T12D8.7
R12C12.2	R74.1	T03F6.2	T05G5.2	T07G12.11	T10B5.3	T12E12.4a
R12C12.8a	R90.1	T03F6.5	T05G5.3	T07G12.13	T10B5.5a	T12F5.4
R12E2.10	T01B11.3	T03G11.8	T05G5.4	T07G12.6	T10B5.6	T12G3.5
R12E2.12	T01B7.5a	T04A8.11	T05G5.6	T07G12.8	T10B9.1	T13A10.11a
R12E2.2	T01B7.6	T04A8.14	T05G5.8	T07H6.3a	T10B9.2	T13F2.3a

Table C Continued.

Genes with Known Phenotypes in *C. elegans*						
T13F2.7	T19D12.2a	T21H3.3	T23G7.1	T27E9.1a	W02A2.6	W04D2.4
T13H5.4	T19E10.1a	T22A3.5	T23G7.5	T27E9.2	W02A2.7	W04D2.5
T13H5.5	T19E7.3	T22A3.8	T23H2.1	T27F2.1	W02B12.6	W04D2.6a
T13H5.6	T19H12.10	T22B11.5	T23H2.5	T27F2.3	W02B12.9	W04G3.2
T14A8.1	T19H12.11	T22B3.1	T24A11.1a	T27F6.5	W02B9.1a	W04G3.3
T14B4.2	T20B12.1	T22B7.1a	T24B8.1	T27F7.1	W02D3.4	W04G3.8
T14B4.6	T20B12.2	T22C1.7	T24B8.7a	T27F7.3a	W02D3.7	W05B10.1
T14B4.7a	T20B12.3	T22D1.10	T24C4.1	T28C6.4	W02D3.9	W05B10.5
T14D7.2	T20B12.7	T22D1.4	T24C4.5	T28C6.6	W02D7.7	W05B2.1
T14F9.1	T20B12.8	T22D1.9	T24H10.1	T28D9.10	W02D9.1	W05B2.5
T14F9.4a	T20B3.2	T22D2.1	T24H10.3	T28F12.2a	W02D9.3	W05B2.6
T14G10.5	T20B5.1	T22E5.5	T24H7.1	T28F2.2	W02F12.5	W05E10.3
T15B7.16	T20F10.1	T22F3.4	T25B9.9	T28F2.5	W03B1.4	W05F2.6
T15B7.2	T20F5.2	T22G5.5	T25C12.1a	T28H11.5	W03C9.3	W05H12.2
T16G12.5	T20G5.1	T22H6.2a	T25C8.2	VC5.4	W03C9.4	W06A7.2
T16H12.4	T20G5.2	T22H9.1	T25D3.2	VF13D12L.1	W03C9.7	W06B11.2
T16H5.1a	T20G5.6	T23B12.2	T25G3.2	VF36H2L.1	W03D2.4	W06E11.1
T17E9.1a	T20H4.3a	T23B12.3	T25G3.3	VW02B12L.1	W03F11.1	W06E11.2
T17E9.2a	T20H4.5	T23B12.7	T26A5.3	VZK822L.1	W03F8.1	W06F12.1a
T17H7.4a	T21B10.1	T23D8.1	T26A5.7a	W01A8.4	W03F8.10	W06H3.3
T17H7.4d	T21B10.2c	T23D8.3	T26A5.9	W01A8.5	W03F9.1	W07A12.6
T19A5.1	T21B10.3	T23D8.4	T26A8.4	W01B11.3	W03F9.10	W07A12.7
T19A5.2a	T21B10.7	T23D8.5	T26E3.3	W01B6.9	W03F9.2a	W07A8.1
T19A5.3a	T21C9.12	T23D8.6	T26E3.4	W01C8.5	W03H9.4	W07A8.3
T19A6.2a	T21C9.5	T23D8.9a	T26G10.1	W01D2.1	W04A4.5	W07A8.5
T19B10.2	T21D12.2	T23E1.2	T26G10.5	W01D2.2a	W04A8.1	W07B3.2a
T19B10.6	T21D12.4	T23F2.1	T26H2.6	W01F3.3	W04A8.7	W07E11.1
T19B10.9	T21D12.9a	T23F6.4	T26H5.5	W01G7.3	W04B5.4	W07E11.3a
T19B4.4	T21E3.1	T23G11.2	T27A3.2	W02A11.1	W04C9.1	W07E6.1
T19B4.5	T21G5.4	T23G11.3	T27B1.2	W02A11.4	W04C9.3	W07E6.2
T19B4.7	T21G5.5a	T23G5.1	T27C4.4a	W02A11.8	W04C9.5	W07E6.4

Table C Continued.

Genes with Known Phenotypes in *C. elegans*						
W08D2.1	Y102A11A.3	Y110A7A.19	Y18D10A.16	Y37D8A.18	Y39E4B.1	Y41E3.4
W08D2.4	Y102A5C.3	Y110A7A.4	Y18D10A.17	Y37D8A.19	Y39E4B.10	Y42G9A.1
W08D2.5	Y105C5A.1	Y110A7A.8	Y18D10A.20	Y37D8A.21	Y39E4B.3a	Y42G9A.4a
W08D2.7	Y105C5A.14	Y111B2A.11	Y18D10A.5	Y37D8A.9	Y39E4B.5	Y43E12A.1
W08E3.1	Y105C5B.12a	Y111B2A.12	Y18D10A.9	Y37E11AL.8	Y39F10B.1a	Y43F4B.4
W08F4.6	Y105E8A.16	Y111B2A.14	Y19D10A.4	Y37E11AM.1	Y39G10AR.10	Y43F4B.5a
W08F4.8	Y105E8A.17	Y111B2A.15	Y19D2B.1	Y37E3.10	Y39G10AR.12a	Y43F4B.6
W09B6.1a	Y105E8A.19	Y111B2A.17	Y22D7AL.5	Y38A10A.2	Y39G10AR.13	Y43F8C.8
W09C2.1	Y105E8A.20	Y111B2A.18	Y23H5A.1a	Y38A10A.5	Y39G10AR.14	Y43H11AL.2
W09C2.3a	Y105E8A.23	Y113G7A.3	Y23H5A.3	Y38A8.2	Y39G10AR.7	Y43H11AL.3
W09C3.4	Y105E8A.24a	Y113G7B.17	Y23H5A.7a	Y38A8.3	Y39G10AR.8	Y44F5A.1
W09C3.6	Y105E8A.25	Y113G7B.18	Y23H5B.5	Y38C1AA.7	Y39G8C.1	Y45F10A.5
W09C5.1	Y105E8A.29	Y113G7B.21	Y23H5B.6	Y38E10A.24	Y39H10A.6	Y45F10C.3
W09C5.2	Y105E8A.6	Y113G7B.23	Y23H5B.7a	Y38F1A.5	Y39H10A.7a	Y45F10D.11
W09C5.4	Y105E8A.9	Y116A8A.9	Y24D9A.1a	Y38F2AL.1	Y40B10A.4	Y45F10D.12
W09C5.6a	Y105E8B.1a	Y116A8C.32	Y24D9A.4a	Y38F2AL.3a	Y40B1A.4	Y45F10D.4
W09C5.8	Y105E8B.2a	Y116A8C.35	Y25C1A.13	Y38F2AL.4	Y40B1B.5	Y45F10D.7
W09D10.1	Y105E8B.8a	Y116A8C.42	Y25C1A.5	Y38H8A.1	Y40B1B.7	Y45F10D.9
W09D10.3	Y106G6A.2a	Y119C1B.8a	Y2H9A.1	Y39A1A.13	Y40D12A.2	Y45F3A.1
W09G12.5	Y106G6E.6	Y119D3B.11	Y32F6A.2	Y39A1A.14	Y41C4A.10	Y45G12B.1a
W09G12.7	Y106G6H.14	Y119D3B.15	Y32F6A.3	Y39A1A.18	Y41C4A.9	Y45G5AM.1a
W09G12.8	Y106G6H.2a	Y119D3B.21	Y32G9B.1	Y39A1A.19	Y41D4B.11	Y46G5A.1a
W09G3.7a	Y106G6H.3	Y11D7A.12	Y34D9A.1	Y39A1A.22	Y41D4B.19a	Y46G5A.31
W10C6.1	Y106G6H.7	Y11D7A.9	Y34D9A.10	Y39A1B.3	Y41D4B.4	Y46G5A.4
W10C8.2	Y108F1.2	Y16B4A.1	Y34D9A.4	Y39B6A.12a	Y41D4B.5	Y46G5A.6
W10D5.2	Y110A2AL.4a	Y17G7B.15a	Y37A1B.1a	Y39B6A.14	Y41D4B.6	Y46H3C.4
W10D5.3a	Y110A2AL.8a	Y17G7B.18a	Y37D8A.1	Y39B6A.3	Y41E3.1	Y47D3A.16
W10D9.5	Y110A7A.11	Y17G7B.20	Y37D8A.10	Y39B6A.33	Y41E3.11	Y47D3A.26
W10G11.19	Y110A7A.13	Y17G7B.5a	Y37D8A.13	Y39B6A.36	Y41E3.13	Y47D3A.6a
W10G6.2	Y110A7A.14	Y18D10A.1	Y37D8A.14	Y39B6A.39	Y41E3.16	Y47D3B.1
W10G6.3	Y110A7A.17a	Y18D10A.13	Y37D8A.16	Y39C12A.1	Y41E3.2	Y47D3B.10

Table C Continued.

Genes with Known Phenotypes in *C. elegans*						
Y47D3B.5a	Y48G1BL.1	Y51H7C.6a	Y54E10BR.8	Y56A3A.6	Y65B4A.9	Y71F9AL.4
Y47D3B.7	Y48G1C.4	Y51H7C.9	Y54E2A.1	Y57A10A.19	Y65B4BL.2	Y71F9AM.4a
Y47D7A.1	Y48G1C.7	Y52B11A.10	Y54E2A.11a	Y57A10A.27	Y65B4BL.3	Y71F9AM.5
Y47G6A.10	Y48G1C.8	Y52B11A.9	Y54E2A.3	Y57E12AL.1a	Y65B4BR.5a	Y71F9B.4
Y47G6A.12	Y48G8AL.1	Y52B11B.1	Y54E5A.4	Y57E12AL.6	Y65B4BR.8	Y71G12B.11a
Y47G6A.15a	Y48G8AL.14	Y52E8A.1	Y54E5B.3a	Y57G11A.1a	Y66A7A.5	Y71G12B.14
Y47G6A.18	Y48G8AL.5	Y53C10A.12	Y54F10BM.14	Y57G11C.12	Y66A7A.8	Y71G12B.9a
Y47G6A.20a	Y48G8AL.8a	Y53C10A.3	Y54F10BM.2	Y57G11C.15	Y66H1A.3	Y71H10B.1a
Y47G6A.23	Y48G8AR.1	Y53C12A.1	Y54G11A.10	Y57G11C.16	Y66H1A.4	Y71H2AL.1
Y47G6A.29	Y49A3A.1	Y53C12A.4	Y54G11A.8a	Y57G11C.17	Y66H1B.2a	Y71H2AL.2
Y47G6A.8	Y49A3A.2	Y53C12B.1	Y54G2A.15	Y57G11C.24a	Y66H1B.3	Y71H2AM.10
Y47G6A.9	Y49E10.1	Y53C12B.2	Y54G2A.2a	Y57G11C.31	Y66H1B.4	Y71H2AM.17
Y47H9C.7	Y49E10.15	Y53F4B.13	Y54G9A.5	Y57G7A.10a	Y67D2.2	71H2AM.20a
Y48A6B.11a	Y49E10.19	Y53F4B.22	Y54H5A.2	Y57G7A.5	Y67D8C.10a	Y71H2AM.23
Y48A6B.3	Y49E10.2	Y53F4B.6	Y54H5A.3	Y59A8A.1	Y67D8C.3a	Y71H2AM.4
Y48A6B.5	Y49E10.20	Y53G8AL.2	Y55B1BM.1a	Y59A8B.2	Y67H2A.1	Y71H2AM.5
Y48A6C.1	Y49E10.21	Y53G8AR.3	Y55D5A.1a	Y59A8B.20	Y67H2A.5	Y71H2B.10a
Y48A6C.4	Y49E10.3a	Y53G8AR.9	Y55F3AM.15	Y59A8B.6	Y67H2A.6	Y71H2B.3
Y48B6A.1	Y49F6B.1	Y53H1C.2	Y55F3AM.3a	Y60A3A.9	Y69A2AR.30a	Y73B3A.12
Y48B6A.13a	Y49F6B.2	Y54E10A.1	Y55F3AR.1	Y61A9LA.10	Y69A2AR.32a	Y73B3A.18a
Y48B6A.2	Y4C6B.2a	Y54E10A.10	Y55F3AR.3	Y62E10A.1	Y69H2.6	Y73B3A.5
Y48B6A.3	Y50D4C.1a	Y54E10A.15a	Y55F3BL.1	Y62E10A.17	Y6B3A.1a	Y73B3B.4
Y48B6A.4	Y50D7A.11	Y54E10A.16a	Y55F3BR.1	Y62E10A.2	Y6B3B.9	Y73B3B.5
Y48C3A.7	Y50D7A.4	Y54E10A.2	Y55H10A.1	Y62F5A.1a	Y6D11A.1	Y73B6A.5a
Y48E1A.1a	Y50D7A.7	Y54E10A.4a	Y56A3A.1	Y63D3A.5	Y70C5C.6a	Y73B6BL.22
Y48E1B.14a	Y50E8A.4a	Y54E10A.7	Y56A3A.17a	Y63D3A.7	Y71A12B.1	Y73B6BL.3
Y48E1B.5	Y51A2D.15	Y54E10BL.1	Y56A3A.18	Y63D3A.8	Y71A12B.13a	Y73B6BL.33
Y48E1B.6	Y51A2D.7a	Y54E10BL.6	Y56A3A.19	Y64G10A.6	Y71D11A.5	Y73B6BL.38
Y48G10A.4	Y51H4A.15	Y54E10BR.4	Y56A3A.20	Y65B4A.1	Y71F9AL.12	Y73B6BL.6
Y48G1A.4	Y51H4A.3	Y54E10BR.5	Y56A3A.21	Y65B4A.3	Y71F9AL.13a	Y73F8A.24
Y48G1A.5	Y51H7C.11	Y54E10BR.6	Y56A3A.32	Y65B4A.6	Y71F9AL.17	Y73F8A.27

Table C Continued.

Genes with Known Phenotypes in *C. elegans*						
Y74C10AR.1	Y92C3B.2a	ZC434.5	ZK1236.3	ZK270.1	ZK632.2	ZK809.4
Y74C9A.2	Y92H12BR.8	ZC477.9a	ZK1236.5a	ZK287.2	ZK637.3	ZK809.7
Y75B7AL.4	Y94H6A.5a	ZC504.4a	ZK1236.7	ZK287.5	ZK637.7a	ZK829.4
Y75B8A.24	Y95D11A.1	ZC506.3	ZK1240.1	ZK328.2	ZK637.8a	ZK836.1
Y75B8A.27	ZC101.2a	ZC506.4	ZK1248.14	ZK328.5b	ZK6.4	ZK856.10
Y75B8A.2a	ZC123.3	ZC513.4	ZK1251.9	ZK377.2a	ZK6.5	ZK856.11
Y75B8A.7	ZC13.3	ZC518.2	ZK1307.6	ZK381.1	ZK652.1	ZK856.13
Y76A2A.2	ZC168.3	ZC581.1	ZK131.1	ZK430.1	ZK652.4	ZK856.7
Y76A2B.1	ZC168.4	ZK1010.1	ZK131.10	ZK430.7	ZK662.4	ZK856.8
Y76B12C.7	ZC196.8	ZK1010.3	ZK131.2	ZK430.8	ZK669.1a	ZK856.9
Y77E11A.13a	ZC239.3	ZK1010.7	ZK131.3	ZK484.4a	ZK673.7	ZK858.1
Y77E11A.7a	ZC247.1	ZK1037.5	ZK131.4	ZK507.6	ZK675.1	ZK858.4
Y79H2A.11	ZC247.3	ZK1058.2	ZK131.5	ZK512.2	ZK675.2	ZK858.7
Y79H2A.3a	ZC250.3	ZK1098.1	ZK131.6	ZK512.6	ZK6.7a	ZK863.6
Y79H2A.6	ZC308.1a	ZK1098.2	ZK131.7	ZK512.7	ZK682.5	ZK863.7
Y80D3A.1	ZC328.1	ZK1098.5	ZK131.8	ZK520.1	ZK686.1	ZK867.1a
Y80D3A.11	ZC373.1	ZK1098.7	ZK131.9	ZK524.3a	ZK686.2	ZK899.2
Y80D3A.2	ZC373.5	ZK1098.8	ZK154.7	ZK546.13	ZK686.3	ZK899.8a
Y80D3A.5	ZC376.6	ZK1127.4	ZK177.6	ZK546.14a	ZK688.9	ZK909.2a
Y82E9BR.13	ZC376.8	ZK1127.5	ZK177.8a	ZK546.1a	ZK742.1a	ZK930.3a
Y82E9BR.15	ZC395.10	ZK1127.6	ZK180.3a	ZK550.3	ZK770.3	ZK945.2
Y82E9BR.16a	ZC395.3	ZK1127.7	ZK180.4	ZK550.4	ZK783.1	ZK945.3
Y82E9BR.2	ZC395.4	ZK1127.9a	ZK20.3	ZK593.5	ZK792.2	ZK945.9
Y82E9BR.3	ZC395.6	ZK1128.3	ZK20.6	ZK593.7	ZK792.3	ZK970.2
Y87G2A.1	ZC395.8	ZK1128.4	ZK250.3	ZK616.6	ZK792.5	ZK970.3
Y87G2A.10	ZC404.7	ZK1128.5	ZK250.8	ZK617.1a	ZK792.6	ZK970.4
Y87G2A.4	ZC410.7a	ZK1151.1a	ZK262.8	ZK622.3a	ZK795.3	ZK973.10
Y87G2A.5	ZC434.2	ZK1151.1b	ZK265.5	ZK632.13	ZK809.1	ZK973.5
Y8G1A.2	ZC434.4	ZK1193.5a	ZK265.6	ZK632.1a	ZK809.3	ZK973.6
Y74C10AR.1	Y92C3B.2a	ZC434.5	ZK1236.3	ZK270.1	ZK632.2	ZK809.4
Y74C9A.2	Y92H12BR.8	ZC477.9a	ZK1236.5a	ZK287.2	ZK637.3	ZK809.7

BIBLIOGRAPHY

Ackley BD, Kang SH, Crew JR, Suh C, Jin Y *et al*. (2003) The basement membrane components nidogen and type XVIII collagen regulate organization of neuromuscular junctions in *Caenorhabditis elegans*. J Neurosci 23(9): 3577-3587.

Adekunle L (2002) Infertility: a sociological analysis of problems of infertility among women in a rural community in Nigeria. Afr J Med Med Sci 31(3): 263-266.

Agger K, Cloos PA, Christensen J, Pasini D, Rose S *et al*. (2007) UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature 449(7163): 731-734.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3): 403-410.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z *et al*. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17): 3389-3402.

Anderson P (1995) Mutagenesis. Methods Cell Biol 48: 31-58.

Arenz C (2003) RNA interference: from an ancient mechanism to a state of the art therapeutic application? Naturwissenschaften 90(8): 345-359.

Arpagaus M, Fedon Y, Cousin X, Chatonnet A, Berge JB *et al*. (1994) cDNA sequence, gene structure, and in vitro expression of ace-1, the gene encoding acetylcholinesterase of class A in the nematode *Caenorhabditis elegans*. J Biol Chem 269(13): 9957-9965.

Bailey CH, Bartsch D, Kandel ER (1996) Toward a molecular definition of long-term memory storage. Proc Natl Acad Sci U S A 93(24): 13445-13452.

Bargmann CI (1998) Neurobiology of the *Caenorhabditis elegans* genome. Science 282(5396): 2028-2033.

Barnes TM, Hekimi S (1997) The *Caenorhabditis elegans* avermectin resistance and anesthetic response gene unc-9 encodes a member of a protein family implicated in electrical coupling of excitable cells. J Neurochem 69(6): 2251-2260.

Barnes TM JY, Horvitz HR, Ruvkun G, Hekimi S. (1996) The *Caenorhabditis elegans* behavioral gene unc-24 encodes a novel bipartite protein similar to both erythrocyte band 7.2 (stomatin) and nonspecific lipid transfer protein. J Neurochem 67(1): 46-57.

Barr MM, Sternberg PW (1999) A polycystic kidney-disease gene homologue required for male mating behaviour in *C. elegans*. Nature 401(6751): 386-389.

Brenner S (1974) The genetics of *Caenorhabditis elegans*. Genetics 77(1): 71-94.

Brundage L, Avery L, Katz A, Kim UJ, Mendel JE *et al*. (1996) Mutations in a *C. elegans* Gqalpha gene disrupt movement, egg laying, and viability. Neuron 16(5): 999-1009.

Byerly L, Cassada RC, Russell RL (1976) The life cycle of the nematode *Caenorhabditis elegans*. I. Wild-type growth and reproduction. Dev Biol 51(1): 23-33.

Cassada RC, Russell RL (1975) The dauer larva, a post-embryonic developmental variant of the nematode *Caenorhabditis elegans*. Dev Biol 46(2): 326-342.

Caterina MJ, Schumacher MA, Tominaga M, Rosen TA, Levine JD *et al*. (1997) The capsaicin receptor: a heat-activated ion channel in the pain pathway. Nature 389(6653): 816-824.

Chalfie M, Sulston JE, White JG, Southgate E, Thomson JN *et al*. (1985) The neural circuit for touch sensitivity in *Caenorhabditis elegans*. J Neurosci 5(4): 956-964.

Coburn CM, Bargmann CI (1996) A putative cyclic nucleotide-gated channel is required for sensory development and function in *C. elegans*. Neuron 17(4): 695-706.

Cogoni C, Macino G (1997) Isolation of quelling-defective (qde) mutants impaired in posttranscriptional transgene-induced gene silencing in Neurospora crassa. Proc Natl Acad Sci U S A 94(19): 10233-10238.

Cogoni C, Macino G (1999) Gene silencing in Neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase. Nature 399(6732): 166-169.

Colbert HA, Smith TL, Bargmann CI (1997) OSM-9, a novel protein with structural similarity to channels, is required for olfaction, mechanosensation, and olfactory adaptation in *Caenorhabditis elegans*. J Neurosci 17(21): 8259-8269.

Corsi AK (2006) A biochemist's guide to *Caenorhabditis elegans*. Anal Biochem 359(1): 1-17.

Cowden C, Sithigorngul P, Brackley P, Guastella J, Stretton AO (1993) Localization and differential expression of FMRFamide-like immunoreactivity in the nematode Ascaris suum. J Comp Neurol 333(3): 455-468.

Culetto E, Sattelle DB (2000) A role for *Caenorhabditis elegans* in understanding the function and interactions of human disease genes. Hum Mol Genet 9(6): 869-877.

Cully DF, Vassilatis DK, Liu KK, Paress PS, Van der Ploeg LH *et al*. (1994) Cloning of an avermectin-sensitive glutamate-gated chloride channel from *Caenorhabditis elegans*. Nature 371(6499): 707-711.

Dehio C, Schell J (1994) Identification of plant genetic loci involved in a posttranscriptional mechanism for meiotically reversible transgene silencing. Proc Natl Acad Sci U S A 91(12): 5538-5542.

Doolittle RF (1995) The multiplicity of domains in proteins. Annu Rev Biochem 64: 287-314.

Doolittle RF, Bork P (1993) Evolutionarily mobile modules in proteins. Sci Am 269(4): 50-56.

Driscoll M CM (1991) The mec-4 gene is a member of a family of *Caenorhabditis elegans* genes that can mutate to induce neuronal degeneration. Nature 49(6310): 588-593.

Dye J (2005) Fertility of American Women:. Current Population Reports: 20-555.

Elmayan T, Balzergue S, Beon F, Bourdon V, Daubremet J *et al*. (1998) Arabidopsis mutants impaired in cosuppression. Plant Cell 10(10): 1747-1758.

Fire A, Xu SQ, Montgomery MK, Kostas SA, Driver SE *et al*. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391(6669): 806-811.

Fitch WM (1970) Distinguishing homologous from analogous proteins. Syst Zool 19(2): 99-113.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF *et al*. (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 269(5223): 496-512.

Fraser AG, Kamath RS, Zipperlen P, Martinez-Campos M, Sohrmann M *et al*. (2000) Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. Nature 408(6810): 325-330.

Furner IJ, Sheikh MA, Collett CE (1998) Gene silencing and homology-dependent gene silencing in Arabidopsis: genetic modifiers and DNA methylation. Genetics 149(2): 651-662.

Gengyo-Ando K, Kamiya Y, Yamakawa A, Kodaira K, Nishiwaki K *et al*. (1993) The *C. elegans* unc-18 gene encodes a protein expressed in motor neurons. Neuron 11(4): 703-711.

Gogarten JP, Olendzenski L (1999) Orthologs, paralogs and genome comparisons. Curr Opin Genet Dev 9(6): 630-636.

Gonczy P, Echeverri C, Oegema K, Coulson A, Jones SJ *et al*. (2000) Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. Nature 408(6810): 331-336.

Goodman MB, Hall DH, Avery L, Lockery SR (1998) Active currents regulate sensitivity and dynamic range in *C. elegans* neurons. Neuron 20(4): 763-772.

Gottschalk A, Almedom RB, Schedletzky T, Anderson SD, Yates JR, 3rd *et al*. (2005) Identification and characterization of novel nicotinic receptor-associated proteins in *Caenorhabditis elegans*. Embo J 24(14): 2566-2578.

Hall DH, Winfrey VP, Blaeuer G, Hoffman LH, Furuta T *et al*. (1999) Ultrastructural features of the adult hermaphrodite gonad of *Caenorhabditis elegans*: relations between the germ line and soma. Dev Biol 212(1): 101-123.

Hansen M, Hsu AL, Dillin A, Kenyon C (2005) New genes tied to endocrine, metabolic, and dietary regulation of lifespan from a *Caenorhabditis elegans* genomic RNAi screen. PLoS Genet 1(1): 119-128.

Harris TW, Lee R, Schwarz E, Bradnam K, Lawson D *et al*. (2003) WormBase: a cross-species database for comparative genomics. Nucleic Acids Res 31(1): 133-137.

Harris TW, Chen N, Cunningham F, Tello-Ruiz M, Antoshechkin I *et al*. (2004) WormBase: a multi-species resource for nematode biology and genomics. Nucleic Acids Res 32(Database issue): D411-417.

He L, Hannon J, (2004) MicroRNAs: Small RNAs with a big role in gene regulation. Nat Rev Gen 5(7): 522-531.

Hill E, Broadbent ID, Chothia C, Pettitt J (2001) Cadherin superfamily proteins in *Caenorhabditis elegans* and Drosophila melanogaster. J Mol Biol 305(5): 1011-1024.

Hillier LW, Coulson A, Murray JI, Bao Z, Sulston JE *et al*. (2005) Genomics in *C. elegans*: so many genes, such a little worm. Genome Res 15(12): 1651-1660.

Hodgkin J (1988) The nematode *C. elegans* (W.B. Wood ed.). pp243-280.

Hope CEAPABIA (2000) *C. elegans* A practical approach.

Horvitz HR, Sulston JE (1980) Isolation and genetic characterization of cell-lineage mutants of the nematode *Caenorhabditis elegans*. Genetics 96(2): 435-454.

Huang M, Gu G, Ferguson EL, Chalfie M (1995). A stomatin-like protein necessary for mechanosensation in *C. elegans*. Nature 378(6554): 292-295.

Hubbard EJ, Greenstein D (2000) The *Caenorhabditis elegans* gonad: a test tube for cell and developmental biology. Dev Dyn 218(1): 2-22.

Iwasaki K, McCarter J, Francis R, Schedl T (1996) emo-1, a *Caenorhabditis elegans* Sec61p gamma homologue, is required for oocyte development and ovulation. J Cell Biol 134(3): 699-714.

Iwasaki K, Staunton J, Saifee O, Nonet M, Thomas JH (1997) aex-3 encodes a novel regulator of presynaptic activity in *C. elegans*. Neuron 18(4): 613-622.

JE Mendel HK, KS Liu, YM Hajdu-Cronin, MI Simon, RH Plasterk, and PW Sternberg (1995) Participation of the protein Go in multiple aspects of behavior in *C. elegans*. Science 267: 1652 - 1655.

John S Reece-Hoyes JS, Denis Dupuy, Christian A Grove, Albertha JM Walhout, Marc Vidal and Ian A Hope (2007) Insight into transcription factor gene duplication from *Caenorhabditis elegans* Promoterome-driven expression patterns. BMC Genomics 8:27 (doi:10.1186/1471-2164-8-27).

Johnson CD, Duckett JG, Culotti JG, Herman RK, Meneely PM *et al*. (1981) An acetylcholinesterase-deficient mutant of the nematode *Caenorhabditis elegans*. Genetics 97(2): 261-279.

Jorgensen EM, Mango SE (2002) The art and design of genetic screens: *Caenorhabditis elegans*. Nat Rev Genet 3(5): 356-369.

Jorgensen EM, Hartwieg E, Schuske K, Nonet ML, Jin Y *et al*. (1995) Defective recycling of synaptic vesicles in synaptotagmin mutants of *Caenorhabditis elegans*. Nature 378(6553): 196-199.

Kaech SM, Whitfield CW, Kim SK (1998) The LIN-2/LIN-7/LIN-10 complex mediates basolateral membrane localization of the *C. elegans* EGF receptor LET-23 in vulval epithelial cells. Cell 94(6): 761-771.

Kaletta T, Hengartner MO (2006) Finding function in novel targets: *C. elegans* as a model organism. Nat Rev Drug Discov 5(5): 387-398.

Kamath RS, Ahringer J (2003) Genome-wide RNAi screening in *Caenorhabditis elegans*. Methods 30(4): 313-321.

Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R *et al*. (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. Nature 421(6920): 231-237.

Kennedy S, Wang D, Ruvkun G (2004) A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. Nature 427(6975): 645-649.

Kennerdell JR, Carthew RW (1998) Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. Cell 95(7): 1017-1026.

Komatsu H, Mori I, Rhee JS, Akaike N, Ohshima Y (1996) Mutations in a cyclic nucleotide-gated channel lead to abnormal thermosensation and chemosensation in *C. elegans*. Neuron 17(4): 707-718.

L'Hernault. SW (1997) *C. elegans II*, Chapter 11.

Lander (2001) Initial sequencing and analysis of the human genome. Nature(409): 860-921.

Lee M, Shen B, Schwarzbauer JE, Ahn J, Kwon J (2005) Connections between integrins and Rac GTPase pathways control gonad formation and function in *C. elegans*. Biochim Biophys Acta 1723(1-3): 248-255.

Lee RY LL, Hengartner M, Horvitz HR, Avery L. (1997) Mutations in the alpha1 subunit of an L-type voltage-activated Ca2+ channel cause myotonia in *Caenorhabditis elegans*. 16(20):6066-76.

Leung-Hagesteijn C, Spence AM, Stern BD, Zhou Y, Su MW *et al*. (1992) UNC-5, a transmembrane protein with immunoglobulin and thrombospondin type 1 domains, guides cell and pioneer axon migrations in *C. elegans*. Cell 71(2): 289-299.

Levitan D, Greenwald I (1995) Facilitation of lin-12-mediated signaling by sel-12, a *Caenorhabditis elegans* S182 Alzheimer's disease gene. Nature 377(6547): 351-354.

Levitan D, Doyle TG, Brousseau D, Lee MK, Thinakaran G *et al*. (1996) Assessment of normal and mutant human presenilin function in *Caenorhabditis elegans*. Proc Natl Acad Sci U S A 93(25): 14940-14944.

Lipardi C, Wei Q, Paterson BM (2001) RNAi as random degradative PCR: siRNA primers convert mRNA into dsRNAs that are degraded to generate new siRNAs. Cell 107(3): 297-307.

Lohmann JU, Endl I, Bosch TC (1999) Silencing of developmental genes in Hydra. Dev Biol 214(1): 211-214.

Maeda I, Kohara Y, Yamamoto M, Sugimoto A (2001) Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. Curr Biol 11(3): 171-176.

Maine EM (2001) RNAi As a tool for understanding germline development in *Caenorhabditis elegans*: uses and cautions. Dev Biol 239(2): 177-189.

Maruyama IN, Brenner S (1991) A phorbol ester/diacylglycerol-binding protein encoded by the unc-13 gene of *Caenorhabditis elegans*. Proc Natl Acad Sci U S A 88(13): 5729-5733.

McIntire SL, Reimer RJ, Schuske K, Edwards RH, Jorgensen EM (1997) Identification and characterization of the vesicular GABA transporter. Nature 389(6653): 870-876.

Miguel-Aliaga I, Culetto E, Walker DS, Baylis HA, Sattelle DB *et al*. (1999) The *Caenorhabditis elegans* orthologue of the human gene responsible for spinal muscular atrophy is a maternal product critical for germline maturation and embryonic viability. Hum Mol Genet 8(12): 2133-2143.

Misquitta L, Paterson BM (1999) Targeted disruption of gene function in Drosophila by RNA interference (RNA-i): a role for nautilus in embryonic somatic muscle formation. Proc Natl Acad Sci U S A 96(4): 1451-1456.

Mittelsten Scheid O, Afsar K, Paszkowski J (1998) Release of epigenetic gene silencing by trans-acting mutations in Arabidopsis. Proc Natl Acad Sci USA 95(2): 632-637.

Montgomery MK, Xu SQ, Fire A (1998) RNA as a target of double-stranded RNA-mediated genetic interference in *Caenorhabditis elegans*. Proc Natl Acad Sci USA 95(26): 15502-15507.

Nakae J, Biggs WH, 3rd, Kitamura T, Cavenee WK, Wright CV *et al*. (2002) Regulation of insulin action and pancreatic beta-cell function by mutated alleles of the gene encoding forkhead transcription factor Foxo1. Nat Genet 32(2): 245-253.

Ngo H, Tschudi C, Gull K, Ullu E (1998) Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. Proc Natl Acad Sci U S A 95(25): 14687-14692.

Ogg S, Paradis S, Gottlieb S, Patterson GI, Lee L *et al*. (1997) The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. Nature 389(6654): 994-999.

Okkema PG, Krause M (2005) Transcriptional regulation. WormBook: 1-40.

Perez-Reyes E, Cribbs LL, Daud A, Lacerda AE, Barclay J *et al*. (1998) Molecular characterization of a neuronal low-voltage-activated T-type calcium channel. Nature 391(6670): 896-900.

Pevsner J (2003) Bioinformatics and functional genomics.

Piano F, Schetter AJ, Mangone M, Stein L, Kemphues KJ (2000) RNAi analysis of genes expressed in the ovary of *Caenorhabditis elegans*. Curr Biol 10(24): 1619-1622.

Piano F, Schetter AJ, Morton DG, Gunsalus KC, Reinke V *et al*. (2002) Gene clustering based on RNAi phenotypes of ovary-enriched genes in *C. elegans*. Curr Biol 12(22): 1959-1964.

Podbilewicz B (2006) WormBook, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.52.1.

Ponting CP, Hofmann K, Bork P (1999) A latrophilin/CL-1-like GPS domain in polycystin-1. Curr Biol 9(16): R585-588.

Poulin G, Nandakumar R, Ahringer J (2004) Genome-wide RNAi screens in *Caenorhabditis elegans*: impact on cancer research. Oncogene 23(51): 8340-8345.

Reboul J, Vaglio P, Tzellas N, Thierry-Mieg N, Moore T *et al*. (2001) Open-reading-frame sequence tags (OSTs) support the existence of at least 17,300 genes in *C. elegans*. Nat Genet 27(3): 332-336.

Reece-Hoyes JS, Shingles J, Dupuy D, Grove CA, Walhout AJ *et al*. (2007) Insight into transcription factor gene duplication from *Caenorhabditis elegans* Promoterome-driven expression patterns. BMC Genomics 8: 27.

Riddle (1988) In "The nematode *C. elegans*" (W. B. Wood ed.). pp393-412.

Robertson HM, Thomas JH (2006) The putative chemoreceptor families of *C. elegans*. WormBook: 1-12.

Rongo C, Whitfield CW, Rodal A, Kim SK, Kaplan JM (1998) LIN-10 is a shared component of the polarized protein localization pathways in neurons and epithelia. Cell 94(6): 751-759.

Rose KL, Winfrey VP, Hoffman LH, Hall DH, Furuta T *et al*. (1997) The POU gene ceh-18 promotes gonadal sheath cell differentiation and function required for meiotic maturation and ovulation in *Caenorhabditis elegans*. Dev Biol 192(1): 59-77.

Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR *et al*. (2000) Comparative genomics of the eukaryotes. Science 287(5461): 2204-2215.

Ruiz F, Vayssie L, Klotz C, Sperling L, Madeddu L (1998) Homology-dependent gene silencing in Paramecium. Mol Biol Cell 9(4): 931-943.

Sanchez Alvarado A, Newmark PA (1999) Double-stranded RNA specifically disrupts gene expression during planarian regeneration. Proc Natl Acad Sci U S A 96(9): 5049-5054.

SchedlT (1997) *C. elegans* II, Chapter 10.

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270(5235): 467-470.

Segalat L, Elkes DA, Kaplan JM (1995) Modulation of serotonin-controlled behaviors by Go in *Caenorhabditis elegans*. Science 267(5204): 1648-1651.

Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, Levesque G *et al*. (1995) Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature 375(6534): 754-760.

Sijen T, Fleenor J, Simmer F, Thijssen KL, Parrish S *et al*. (2001) On the role of RNA amplification in dsRNA-triggered gene silencing. Cell 107(4): 465-476.

Simmer F, Tijsterman M, Parrish S, Koushika SP, Nonet ML *et al*. (2002) Loss of the putative RNA-directed RNA polymerase RRF-3 makes *C. elegans* hypersensitive to RNAi. Curr Biol 12(15): 1317-1319.

Simmer F, Moorman C, van der Linden AM, Kuijk E, van den Berghe PV *et al*. (2003) Genome-wide RNAi of *C. elegans* using the hypersensitive rrf-3 strain reveals novel gene functions. PLoS Biol 1(1): E12.

Sjolander K (2004) Phylogenomic inference of protein molecular function: advances and challenges. Bioinformatics 20(2): 170-179.

Sonnhammer EL, Durbin R (1997) Analysis of protein domain families in *Caenorhabditis elegans*. Genomics 46(2): 200-216.

Starich TA, Lee RY, Panzarella C, Avery L, Shaw JE (1996) eat-5 and unc-7 represent a multigene family in *Caenorhabditis elegans* involved in cell-cell coupling. J Cell Biol 134(2): 537-548.

Stretton AO, Cowden C, Sithigorngul P, Davis RE (1991) Neuropeptides in the nematode Ascaris suum. Parasitology 102 Suppl: S107-116.

Stretton Aow DR, Angstadt Jd, Donmoyer Je, Johnson Cd (1985) Neural control of behavior in ascaris. Trends in Neurosciences 8(6): 294-300.

Strome S (1986) Fluorescence visualization of the distribution of microfilaments in gonads and early embryos of the nematode *Caenorhabditis elegans*. J Cell Biol 103(6 Pt 1): 2241-2252.

Sulston JE, Schierenberg E, White JG, Thomson JN (1983) The embryonic cell lineage of the nematode *Caenorhabditis elegans*. Dev Biol 100(1): 64-119.

Sundaram M, Greenwald I (1993) Suppressors of a lin-12 hypomorph define genes that interact with both lin-12 and glp-1 in *Caenorhabditis elegans*. Genetics 135(3): 765-783.

Tabara H, Grishok A, Mello CC (1998) RNAi in C-elegans: Soaking in the genome sequence. Science 282(5388): 430-431.

Talbot K, Ponting CP, Theodosiou AM, Rodrigues NR, Surtees R *et al*. (1997) Missense mutation clustering in the survival motor neuron gene: a role for a conserved tyrosine and glycine rich region of the protein in RNA metabolism? Hum Mol Genet 6(3): 497-500.

Tepass U (2003) Claudin complexities at the apical junctional complex. Nat Cell Biol 5(7): 595-597.

Timmons L, Fire A (1998) Specific interference by ingested dsRNA. Nature 395(6705): 854.

Timmons L, Court DL, Fire A (2001) Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. Gene 263(1-2): 103-112.

Tischler J, Lehner B, Chen N, Fraser AG (2006) Combinatorial RNA interference in *Caenorhabditis elegans* reveals that redundancy between gene duplicates can be maintained for more than 80 million years of evolution. Genome Biol 7(8): R69.

Treinin M, Chalfie M (1995) A mutated acetylcholine receptor subunit causes neuronal degeneration in *C. elegans*. Neuron 14(4): 871-877.

Troemel ER, Chou JH, Dwyer ND, Colbert HA, Bargmann CI (1995) Divergent seven transmembrane receptors are candidate chemosensory receptors in *C. elegans*. Cell 83(2): 207-218.

Tsukita S, Furuse M (2000) The structure and function of claudins, cell adhesion molecules at tight junctions. Ann N Y Acad Sci 915: 129-135.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ *et al*. (2001) The sequence of the human genome. Science 291(5507): 1304-1351.

Walrond JP, Stretton AO (1985) Reciprocal inhibition in the motor nervous system of the nematode Ascaris: direct control of ventral inhibitory motoneurons by dorsal excitatory motoneurons. J Neurosci 5(1): 9-15.

Walrond JP, Kass IS, Stretton AO, Donmoyer JE (1985) Identification of excitatory and inhibitory motoneurons in the nematode Ascaris by electrophysiological techniques. J Neurosci 5(1): 1-8.

Wang D, Kennedy S, Conte D, Jr., Kim JK, Gabel HW *et al*. (2005) Somatic misexpression of germline P granules and enhanced RNA interference in retinoblastoma pathway mutants. Nature 436(7050): 593-597.

Wang P, Zhao J, Corsi AK (2006) Identification of novel target genes of CeTwist and CeE/DA. Dev Biol 293(2): 486-498.

Wargelius A, Ellingsen S, Fjose A (1999) Double-stranded RNA induces specific developmental defects in zebrafish embryos. Biochem Biophys Res Commun 263(1): 156-161.

Watson JD, Laskowski RA, Thornton JM (2005) Predicting protein function from sequence and structural data. Curr Opin Struct Biol 15(3): 275-284.

Wei A, Jegla T, Salkoff L (1996) Eight potassium channel families revealed by the *C. elegans* genome project. Neuropharmacology 35(7): 805-829.

Wei L, Liu Y, Dubchak I, Shon J, Park J (2002) Comparative genomics approaches to study organism similarities and differences. J Biomed Inform 35(2): 142-150.

Wen JY KN, Morrison G, Rambaldini G, Runciman S, Rousseau J, van der Kooy D. (1997) Mutations that prevent associative learning in *C. elegans*. Behav Neurosci 111(2): 354-368.

Wianny F, Zernicka-Goetz M (2000) Specific interference with gene function by double-stranded RNA in early mouse development. Nat Cell Biol 2(2): 70-75.

Wittenburg N, Eimer S, Lakowski B, Rohrig S, Rudolph C *et al*. (2000) Presenilin is required for proper morphology and function of neurons in *C. elegans*. Nature 406(6793): 306-309.

Wolf YI, Rogozin IB, Kondrashov AS, Koonin EV (2001) Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. Genome Res 11(3): 356-372.

Wood V (2002) The genome sequence of *Schizosaccharomyces pombe*. Nature 415: 871–880.

Wood WB (1988a) The nematode *C. elegans*. pp1-16.

Wormbase.org (2005) (Release WS159).

Xu X DL, Shih HY, Seo S, Ahn J and Lee M (2005) Linking integrin to IP(3) signaling is important for ovulation in *Caenorhabditis elegans*. FEBS Lett 579 549-553.

Yu S, Avery L, Baude E, Garbers DL (1997) Guanylyl cyclase expression in specific sensory neurons: a new family of chemosensory receptors. Proc Natl Acad Sci U S A 94(7): 3384-3387.

Zhong-Hui Duan BH, Lothar Reichel, Dianne M Perez and Ting Shi (2006) The relationship between protein sequences and their gene ontology functions. BMC Bioinformatics 7(Suppl 4): S11.