MOLECULAR IDENTIFICATION AND PHENOTYPIC CHARACTERIZATION OF *LET-65 (MARS-1)* IN *CAENORHABDITIS ELEGANS*

by

Maha Zakariya Al-Riyami
B.Sc., Sultan Qaboos University, 2006

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

In the
Department of Molecular Biology and Biochemistry

SIMON FRASER UNIVERSITY

Summer 2010

APPROVAL

Name: **Maha Zakariya Al-Riyami**

Degree: **Master of Science**

Title of Thesis: **Molecular identification and phenotypic characterization of *let-65 (mars-1)* in *Caenorhabditis elegans*.**

Examining Committee:

Chair: **Dipankar Sen**
Professor, Department of Molecular Biology and Biochemistry

_____

Dr. David L. Baillie
Senior Supervisor
Professor, Department of Molecular biology and Biochemistry

_____

Dr. Barry M. Honda
Supervisor
Professor, Department of Molecular Biology and Biochemistry

_____

Dr. Jack N. Chen
Supervisor
Associate Professor, Department of Molecular Biology and Biochemistry

_____

Dr. Nicholas Harden
Supervisor
Associate Professor, Department of Molecular Biology and Biochemistry

_____

Dr. Frederic Pio
Internal examiner
Associate Professor, Department of Molecular Biology and Biochemistry

Date Defended/Approved: July 26, 2010

Declaration of
Partial Copyright Licence

ABSTRACT

let-65 was first identified in a screen for lethal mutations in *C. elegans*' unc-22 region. I have demonstrated the molecular identity of *let-65* to be *F58B3.5 (mars-1)*, which encodes a methionyl tRNA synthetase (MARS-1).

To develop a deeper insight into MARS-1 activity, I experimentally confirmed, from its sub-cellular localization, that it is a cytoplasmic enzyme. I also attempted to determine the sub-cellular localization of every known *C. elegans* amino-acyl tRNA synthetase using computational methods and, in collaboration with WormBase, renamed the genes appropriately. In addition, I studied *let-65* transcription regulation by analyzing its 5' promoter containing region.

Haplo-insufficiency phenotypes manifest as a consequence of reduction in copy number of genes that encode proteins involved in translation. I have investigated this in *C. elegans*. I used two alleles of each of *let-65* (*mars-1*) and *let-336* (*rps-27*) and examined putative haplo-insufficiency phenotypes for both these genes.

Keywords: *Caenorhabditis elegans*; unc-22 region; Methionyl-tRNA synthetase; *let-65*; Aminoacyl tRNA synthetase; Haploinsufficiency

To my parents

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my senior supervisor, Dr. David Baillie, for his encouragement and support, and for giving me the valuable opportunity to pursue my studies in his lab. My appreciation is also extended to the members of my supervisory committee, Dr. Barry Honda, Dr. Jack Chen, and Dr. Nicholas Harden for their support and advice. Many thanks go to Dr. Johnsen for his help and advice especially during the preparation of this thesis. I would also want to thank Domena Tu for her help and for injecting all my constructs, Dr. Martin Jones for providing me with the oaCGH data, and Rylan Fernandes for providing me with *let-336* strains and helping me with the progeny counts. I also thank Tim Heslip for helping me with the spinning disk confocal microscope. I would also like to acknowledge past and present members of the Baillie's lab: Danyela Lee, ShuYi Chua, Carrie Simms, Jim Huang, Saeideh Shamsi, Heesun Shin, and Allan Mah and the members of the Chen's lab, especially Jeffrey Shih-Chieh Chu. I also thank my family and my friends for their constant encouragement and support.

I also pass my appreciation to Simon Fraser University and Department of Molecular Biology and Biochemistry for providing me with a wonderful learning experience and to Sultan Qaboos University and the government of The Sultanate of Oman for sponsoring my master education.

TABLE OF CONTENTS

Approval ... ii

Abstract ... iii

Dedication ... iv

Acknowledgements ... v

Table of Contents ... vi

List of Figures ... viii

List of Tables ... ix

General Introduction ... 1

Chapter 1: Determination of the Molecular Identity of *let-65* ... 6

1.1 Introduction ... 6

1.2 Materials and Methods ... 11

1.2.1 Maintaining nematode strains ... 11

1.2.2 Homozygous *let-65* isolation ... 11

1.2.3 *C.elegans* genomic DNA extraction ... 12

1.2.4 PCR amplification and sequencing ... 12

1.2.5 Agarose gel electrophoresis ... 13

1.2.6 Sequence alignment and analysis ... 13

1.2.7 Microinjection ... 13

1.2.8 *let-65* rescue using fosmid *WRM0615dH10* ... 14

1.3 Results ... 15

1.3.1 *F58B3.4* and *F58B3.5* sequencing results ... 15

1.3.2 *let-65*(*s1083* and *s1222*) rescue using the fosmid *WRM0615dH10* ... 15

1.4 Discussion ... 16

Chapter 2: AARS Sub-Cellular Localizations ... 26

2.1 Introduction ... 26

2.2 Materials and Methods ... 27

2.2.1 Construction of translational transgenes ... 27

2.2.2 Microinjection ... 29

2.2.3 Microscopy ... 30

2.2.4 Computational prediction of AARS sub-cellular localizations ... 30

2.3 Results ... 32

2.3.1 *F58B3.5* sub-cellular localization based on the expression of *let-65*$_{(p\text{-}2,596)}$:: *let-65::* GFP translational fusion ... 32

2.3.2 Prediction of AARS sub-cellular localizations ... 32

2.3.3 The sub-cellular localization of *srs-1* (*Wo3B1.4,* now named *sars-2*) and *srs-2* (*C47E12.1,* now named *sars-1*) by translational fusion. 34
2.4 Discussion 39

Chapter 3: Mapping of a *5' cis*-Acting Regulatory Motif in *mars-1* Promoter Region.....46
3.1 Introduction 46
3.2 Materials and Methods 47
3.2.1 Construction of *let-65*$_{promoter}$::GFP transgenes 47
3.2.2 Microinjection 48
3.2.3 Microscopy 48
3.2.4 Phylogenetic analysis 48
3.3 Results 49
3.3.1 Sequential *let-65*$_{promoter}$:: GFP deletion constructs 49
3.3.2 Identification of evolutionarily conserved non-coding elements upstream of *mars-1* 52
3.4 Discussion 55

Chapter 4: Investigation of Haplo- Insufficiency Phenotypes Associated with Mutations in Genes Involved in Translation in *C. elegans* 58
4.1 Introduction 58
4.2 Materials and Methods 60
4.2.1 Maintaining nematode strains and culture conditions 60
4.2.2 Brood size assay 60
4.2.3 Developmental delay assay and microscopy 61
4.2.4 Lifespan assay 61
4.3 Results 62
4.3.1 Self-progeny brood size 62
4.3.2 Developmental delay 63
4.3.3 Lifespan assay 70
4.4 Discussion 73

General Discussion 79

Appendices 84
Appendix A. *let-65* alleles and their lethal phenotypes. 84
Appendix B. *F58B3.4* and *mars-1* wild type sequence and primers used for sequencing *let-65* alleles. 85
Appendix C. BlastP (scores and E-values), and WoLF PSORT localization scores for AARS localizations predictions. 89
Appendix D. *gars-1*, *hars-1*, *tars-,1* and *mel-32* gene models. 92
Appendix E. Forward primers used in promoter amplification reactions to generate *let-65*$_{promoter}$::GFP transgene constructs. 93
Appendix F. Genotypes of the generated *mars-1*$_{promoter}$::GFP strains. 94
Appendix G. Upstream regions of *mars-1* in *C. elegans*, *C.briggsae,* and *C. remanei* 95
Appendix H. Punnett square tables of selfed *let-56*, *let-366* and the controls (BC1270 and BC2200). 96

Reference List 100

LIST OF FIGURES

Figure 1. *sDf10* and *sDf84* oligonucleoyide array Comparative Genomic Hybridization (oaCGH) assay data. 9

Figure 2. *let-65* chromosomal location and candidate genes. 10

Figure 3. MARS-1 conservation and location of amino acid mutations in *let-65* alleles. 20

Figure 4. Graphical representation of MARS-1 domains and positions of *let-65* found mutations. 24

Figure 5. AARS genes distribution in *C. elegans'* genome. 25

Figure 6. Gene models of generated translational fusion constructs. 31

Figure 7. Expression of *F58B3.5*, *Wo3B1.4*, and *C47E12.1* translational fusions. 35

Figure 8. Expression differences among translational fusion constructs 45

Figure 9. Gut GFP expression levels for the shortest six *mars-1*$_{promoter}$::GFP constructs. 51

Figure 10. Interspecies comparison of *C. elegans*, *C. briggsae*, and *C. remanei* upstream regions. 53

Figure 11. Sequence alignment of *C. elegans* and *C. briggsae mars-1* gene upstream region. 54

Figure 12. Average number of wild type progeny for *let-65*(*s1222* and *s1777*) and the control BC1270. 65

Figure 13. Average progeny number for *let-336*(*s1413* and *s1495*) and the control BC2200. 66

Figure 14. Body length of *let-65*(*s1222* and *s1777)* at different time points. 67

Figure 15. Body length of *let-336*(*s1413* and *s1495)* at different time points. 68

Figure 16. Average body lengths of *let-65(s1222* and *s1777)* and *let-336(s1413* and *s1495)*. 69

Figure 17. *let-65(s1222* and *s1777)* and *let-336(s1413* and *s1495)* survival curve. 71

Figure 18. Average number of days survived for *let-65(s1222* and *s1777)* and *let-336(s1413* and *s1495)*. 72

LIST OF TABLES

Table 1. EMS mutations found in *F58B3.5* alleles. 16

Table 2. AARS enzymes family, their protein size and observed RNAi phenotype (WS215, 18-jun-2010). 21

Table 3. *C. elegans aars* revised nomenclature, and BlastP and WoLF PSORT sub-cellular localizations....36

Table 4. GFP expression patterns in strains containing truncated *F58B3.5*$_{promoter}$::GFP constructs. 50

Table 5. Genotypes of strains used in *C. elegans* haplo-insufficiency phenotype investigations....62

GENERAL INTRODUCTION

Essential genes, as defined in our laboratory, include not only ones necessary for an organism to grow to fertile adulthood but also genes required for the subsequent generation to reach fertility (Kemphues 2005). Mutations that identify essential genes are categorized as embryonic, larval lethal, sterile, and maternal-effect lethal. Embryonics block development before the eggs hatches while larval lethals block development in one of the four larval stages (L1-L4), which occur before adulthood. Sterile organisms reach adulthood but do not produce fertilized oocytes. Maternal-effect lethals results from mutations in genes whose wild type expression is required in the P_0 for development of its progeny (Kemphues 2005).

The study of essential genes is important for understanding an organism's development, survival, and reproduction. Such studies have been carried out in many organisms besides the nematode *Caenorhabditis elegans* these include mice, Drosophila, and yeast (Jones 1994).

Sydney Brenner (1974) was the first to use *C. elegans* as a model system. The nematode's small size and its short life cycle make it an excellent experimental organism. *C. elegans*' two sexes, self-fertilizing hermaphrodite (5AA: XX) and males (5AA: XO), allows easy strain maintenance in addition to standard genetic analysis (Brenner 1974). In addition *C. elegans* have a transparent body that permits examination of internal development and defects. Also, it allows in *vivo* gene expression using fluorescent reporter genes, such as green fluorescent protein (GFP) (Chalfie *at al.*1994). Moreover,

the genome of *C. elegans* was the first genome of a multicellular organism to be entirely sequenced and it continues to have the most completely annotated genome. In addition, the public availability of *C. elegans* genome sequences and the availability of genome sequences of other closely and more distantly related species provides a powerful tool for comparative genomic analysis (Lee 2009). All these qualities combine to make *C. elegans* an excellent organism for essential gene studies.

let-65 is an essential gene, which was first identified in an ethyl methanesulfunate (EMS) screen and mapped to the unc-22 region in the right half of Chromosome IV (LGIV) (Moerman 1980). In subsequent screens more alleles of *let-65* were isolated (Rogalski *et al.* 1982; Moerman 1980; Rogalski and Baillie 1985; and Clark *et al*. 1988). *let-65* is intriguing because of its high mutability. In the above screens nine *let-65* alleles were isolated and eight of them display a mid larva arrest phenotype while one allele (*s694*) causes sterility (Clark 1990). In the first chapter of this thesis I determine, by sequencing and complementation test, that the molecular identity of *let-65* corresponds to the predicted gene *F58B3.5*.

F58B3.5 (*mars-1*) encodes for a methionyl-tRNA synthetase MARS. MetRS-1 is one of at least 20 cytoplasmic aminoacyl tRNA synthetase enzymes (AARSs) involved in biology's central dogma (information flows from DNA to RNA to proteins). These enzymes act in the translation of mRNAs by catalyzing aminoacylation of their cognate tRNAs (O'Donoghue and Luthey-Schulten 2003).

It is crucial for a protein to be in its correct sub-cellular location. Knowing the protein's sub-cellular location provides insight into protein's function. The sub-cellular location of a protein can be demonstrated experimentally or predicted by computational

methods. One experimental method takes advantage of reporter proteins such as GFP that I employed to locate three *C. elegans* AARSs. In *C. elegans,* beside the cytoplasmic AARSs, the nuclear genome encodes AARSs that function in mitochondrial protein translation. In this thesis, I used computational approaches to predict the sub-cellular locations of all known AARSs enzymes in *C. elegans'* genome.

At the pre-transcriptional level, non-coding DNA sequences play important regulatory roles in gene expression. RNA polymerase II binds to specific sequences of promoter region and functions in transcribing RNA chains from DNA genes. Activator and repressor proteins, known as transcription factors (TFs), bind to upstream *cis*-regulatory elements and manage the expression by alternating the association between RNA polymerase II and promoter sequence (Blackwell and Walker 2006). *C. elegans* provides a powerful tool for studying these elements, since the basal transcriptional machinery is highly similar to that in other eukaryotes (McGhee and Krause 1997). The largest subunit in *C. elegans'* RNA polymerase II has high sequence similarity (80%) to even distant related animals such as vertebrate mammals (McGhee and Krause 1997). Moreover, the TATA-element-binding protein's (TBP) carboxy-terminal is highly conserved with that of yeast, *Drosophila*, and humans with 75–85% sequence identity and can interact directly with basal transcription factors of both humans and *Drosophila* (McGhee and Krause 1997). Reporters have also been used to locate these transcriptionally important regulatory sequences by taking advantage of promoter driven reporter constructs. This methodology is based on sequentially truncating the genes' upstream promoter containing regions (from the 5' end) and then fusing each fragment to a reporter, followed by micro-injecting the constructs into nematodes to generate reporter

transgenics. Such transgenic strains can be screened for loss of expression or expression alternations in order to identify promoter elements necessary for proper gene expression (Lee 2009).

For a variety of organisms, mutations in genes encoding ribosomal proteins (RPs) have been shown to cause cellular and developmental dominant phenotypes. In *Drosophila melanogaster*, disruption of RP genes can result in a haplo-insufficient phenotype (*Minute*). This phenotype includes prolonged development, short and thin bristles, and poor fertility (Lambertsson 1998). In *Arabidopsis* haploinsufficiency of the ribosomal protein *RPS5* results in disruption of cell division (Marygold *et al.* 2007). In humans, haplo-insufficiency of *RPS19* results in Diamond Blackfan anemia (Marygold *et al.* 2007).

An explanation for these defects is that the reduction in the amount of a single RP results in impaired cytoplasmic ribosomal function. This is due to the assembly of incomplete nonfunctional ribosomal subunits that cause a general reduction in protein synthesis (Marygold *et al.* 2007). These mutations and their phenotypes have been extensively studied in *D. melanogaster*. However, in *C. elegans,* there have been no studies of such mutations and their haplo-insufficient phenotypes for components of the cytoplasmic translational mechanisms. The main objectives of this thesis are:

1. To determine the molecular identity of *let-65.*
2. To determine the sub-cellular location of *C. elegans* amino-acryl tRNA synthetase enzymes.

3. To investigate the 5' *cis*- regulatory elements of *let-65(mars-1)* those are required for its transcription.

4. To investigate *C. elegans* haplo-insufficiency phenotypes associated with mutations in genes involved in the translation process.

CHAPTER 1: DETERMINATION OF THE MOLECULAR IDENTITY OF *LET-65*

1.1 Introduction

C. elegans' nuclear genome consists of six chromosome pairs with sizes ranging from 14×10^6 to 22×10^6 base pairs (bp) (Coulson *et al.* 1991). Part of LGIV (17×10^6 bp) is genetically balanced by a reciprocal translocation called *nT1(IV;V)* (Clark *et al.* 1988).

sDf2 is a deficiency which extends approximately two map units on LG IV. The breakpoints of this deficiency define the unc-22 region, which is on the right half of LGIV. This region is also referred to as the *sDf2* region (Clark 1990). The unc-22 region was named after *unc-22,* which encodes twitchin, a large protein expressed in muscle (6,839 amino acids). Mutant *unc-22* animals have an uncoordinated "twitching'' recessive phenotype (Brenner 1974) and a dominant "twitching'' phenotype in a 1% nicotine alkaloid solution (Moerman and Baillie 1979). These "twitching'' worms are also thinner, with slower movement and longer developmental time than wild type (N2) *C. elegans* (Moerman and Baillie 1979). Thus *unc-22* is a useful marker for mutations in genes within the region. The unc-22 region has been sub-divided by 34 deficiencies (Moerman and Baillie 1981; Clark and Baillie 1992; Rogalski and Baillie 1985; Rogalski *et al.* 1982, Clark *et al.* 1988; Schein *et al.* 1993; Marra and Baillie 1994). Two methods were used to determine the extents of these deficiencies: Standard complementation tests and polymerase chain reaction based method (Schein 1994). Recently, Oligonucleotide

array Comparative Genomic Hybridization (oaCGH) has been used for the molecular characterization of several deficiencies and duplication containing strains within the unc-22 region. The results of incorporating the oaCGH data into the genetic map allowed for direct positional cloning of essential genes and rapid identification of some mutations (Jones *et al.* 2007).

Essential genes include not only genes necessary for an organism to grow to fertile adulthood but also genes required for the next generation to reach fertility (Kemphues 2005). Mutations within the unc-22 region were generated by EMS mutagenesis which commonly induces G/C-A/T transitions (Anderson 1995).

let-65 is an essential gene in the unc-22 region. It has nine alleles all of which were generated by EMS treatment (Rogalski *et al,* 1982; Moerman 1980; Rogalski and Baillie 1985; Clark *et al.* 1988, and Clark 1990). Molecular identification of *let-65* is of particular interest because of its large size (as determined by it mutational rate). It is one of four genes in the region with a large number of alleles (*let-56*, *let-65*, *let-59*, and *let-68*). Teresa Rogalski and Jackie Schein conducted preliminary investigations of *let-65* (Rogalski and Baillie 1985; Schein 1994) and I have summarize their studies here:

T. Rogalski conducted complementation tests between mutations and deficiencies in the unc-22 region. Complementation testing of *let-65* with *sDf8* and *sDf10* deficiencies showed that both *sDf8* and *sDf10* fail to complement *let-65* (Rogalski and Baillie 1985). Schein (1994) subsequently demonstrated that *sDf84* complements *let-65*. From the results of these two studies it was concluded that *let-65* resides between *sDf10* left breakpoint and *sDf84* left breakpoint.

Martin Jones identified molecularly the breakpoints of *sDf10* and *sDf84* by oaCGH (Jones, M. pers. comm). This technique showed that *sDf10* left and right breakpoints are at 11,629,654bp and 12,243,134bp respectively. The deficient region of *sDf84* is between 11,639,813bp and 12,749,450bp (Fig. 1). These results limited the candidate genes for *let-65* to *F58B3.4* and *mars-1* which are located between the left breakpoints of *sDf10* and *sDf84* (Fig. 2). I determined the molecular identity of *let-65* by sequencing the genomic regions for each of the candidate genes in a strain (BC962) containing *let-65*(*s254*). I amplified the two candidate genes (*F58B3.4* and *mars-1*) by PCR, and sent the products to Macrogen Korea for sequencing. I found a G to A transition in *mars-1.* Accordingly I sequenced *mars-1* in seven more alleles and found G-C to A-T transitions. Alignment of *mars-1* with its orthologus genes in *Caenorhabditis briggsae*, *Homo sapien,* and *Drosophila melanogaster* showed that most of the mutations I found resulted in amino acid changes in evolutionarily conserved residues. I confirmed the sequencing results by rescuing two *let-65* alleles (*s1083* and *s1222*) with the fosmid *WRM0615dH10* that shows that it spans the genomic region containing *mars-1.*

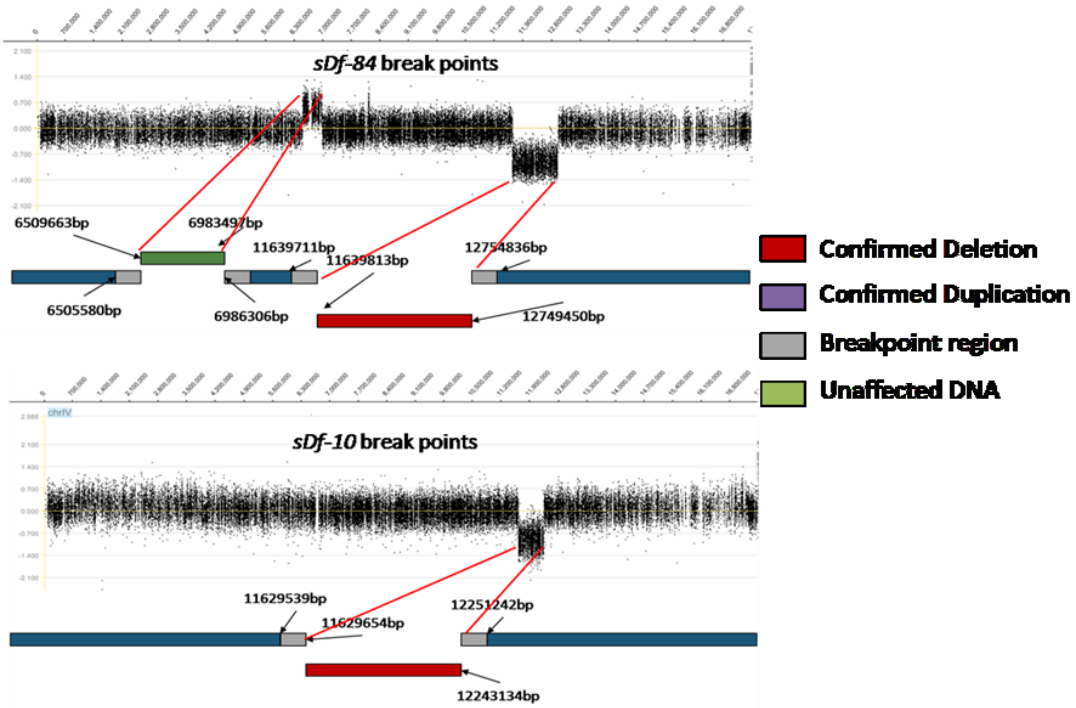


Figure 1. ***sDf10*** **and** ***sDf84*** **oligonucleoyide array Comparative Genomic Hybridization (oaCGH) assay data.**
sDf10 **and** ***sDf84*** **deletions breakpoints found by oaCGH (Jones, M. pers. comm).**

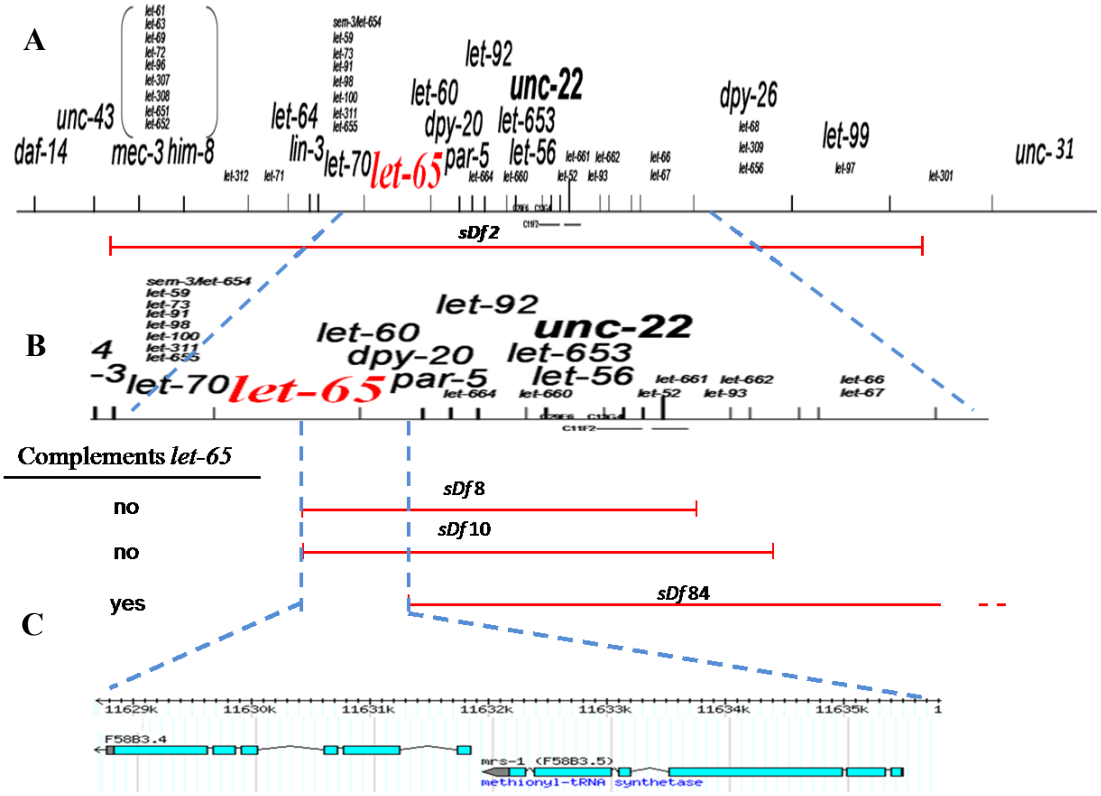


Figure 2. ***let-65*** **chromosomal location and candidate genes.**
Physical map and genetic map correlation in the *let-65* region. A. unc-22 region defined by *sDf2*. B. By complementation test, *sDf-8* and *sDf-10* do not complement *let-65* (Rogalski and Baillie 1985), *sDf-84* complements *let-65* (Schein 1994). *let-65* locates between the left breakpoints of *sDf-10* and *sDf-84*. C. *F58B3.4* and *F58B3.5* (previously named mrs-1), are the two candidate genes for *let-65* locate they are between *sDf-10* and *sDf-84* left breakpoints. *WRM0615dH10* is the fosmid that was used for *let-65* complementation testing (Modified from Clark 1990, C. is modified from WormBase, WS212).

1.2 Materials and Methods

1.2.1 Maintaining nematode strains

Maintaining the nematode strains was on Petri plates containing Easiest Worm Plate Agar streaked with *Escherichia coli* OP50 (Brenner 1974). The strains used in this study had the genotype *let-65* mutations that were previously induced on LGIV in strains with the markers *unc-22(s7)* or *unc-22(s7)* and *unc-31 (e169). unc-22* is a conditionally semi-dominant mutation in that *unc-22(s7)/+* individuals twitch in a 1% nicotine solution (Sigma) whereas *unc-22* homozygote's are unconditional twitchers (Moerman and Baillie 1979). To prevent recombination of the lethal alleles away from these markers the lethal bearing chromosomes were balanced over the reciprocal translocation *nT1(IV;V)*. *nT1(IV;V)* was first analyzed by Ferguson and Horvitz (1985). For the list of *let-65* strains analyzed by sequencing, refer to Appendix A. The used wild-type strain was N2 (var. Bristol, Baillie Laboratory strain, BC49).

1.2.2 Homozygous *let-65* isolation

To isolate homozygous individuals, gravid *let-65* hermaphrodite's for each strain were selfed and incubated at 20°C to lay eggs for four-hour. All viable L4 stage progeny were removed leaving developmentally arrested individuals which constantly twitch (*let-65/let-65)* on the plates.

1.2.3 *C.elegans* genomic DNA extraction

Genomic DNA was isolated from worms using single worm lysis. One worm was picked into 2µl of lysis buffer (50mM Kcl, 10mM Tris pH 8.2, 2.5 mM MgCl2, 0.45% NP-40, 0.45% Tween-20, 0.01% DNA free gelatine and 100ug/ml proteinase K) then stored in Liquid Nitrogen. PCR procedure was used on the lysed worms' genomic DNA. The PCR program starts by incubation at 65°C for 60 minutes followed by heating to 95°C for 15 minutes to inactivate proteinase K. Genomic DNA samples were stored at -20°C until needed as a template for a 20µl PCR reaction. Two µl of genomic DNA was then used as template for the 20µl PCR reaction.

1.2.4 PCR amplification and sequencing

Sequences for wild type *F58B3.4* and *F58B3.5* genomic region were obtained from WormBase (WS200, 20-Mar-2009). *F58B3.4* PCR amplification primers: forward: TTCAA TTGTT GACCG TTTGG, nested forward: CCAAA AAGAG ACTGA ACCGA, reverse: GGAAA GATCG CATTG GATTT, nested reverses: TCGGA AATGC TTATC AAACG. *F58B3.5* was dividing into two parts, part 1 and part 2. *F58B3.5* part 1 PCR amplification primers: forward: CGAGC TGAGG AACAT GATGA, nested forward: GTGAG TCGCG TTGTT ATTCG, reverse: ATGCA TTGTG TGCAG CTCTC, nested reverse: CACTG TTCAC GGATT GTTGC. *F58B3.5* part 2 PCR amplification primer: forward: CCTCG TTGCA TCACA AGAGA, nested forward: CAACA GAGCA CTTTC GTTCG, reverse: CGATT CCACC TTCCT TCTCA, nested reverse: TCATC AGCAT CTTCC AATGC. Sequencing primers were designed to be staggered approximately every 500bp on either strand. Primer 3 (Rozen and Skaletsky 2000) was used to design all primers. For the wild

type *F58B3.4* and *F58B3.5* sequences and the list of sequencing primers used, refer to Appendix B. All samples were shipped to Macrogen Korea for sequencing.

1.2.5 Agarose gel electrophoresis

PCR products were resolved by electrophoresis through one-percentage agarose gels in 1X TAE buffer. Ethidium bromide (0.05μg/ml) was added to the agarose gels that were electrophoresed in the appropriate buffer (1X TAE). A 1 kb ladder was used as a marker DNA. DNA band were made visible by exposure to UV light.

1.2.6 Sequence alignment and analysis

Nucleotide mutations and their corresponding amino acid changes were found by aliening sequences using Biological Sequence Alignment Editor (BioEdit) (http://www.mbio.ncsu.edu/Bio Edit/BioEdit.html).

1.2.7 Microinjection

Filament capillary tubes (1.0-mm, 6") (World Precision Instruments) were made into microinjection needles by pulling with a Sutter P-97 horizontal needle puller and mounted into a Leitz Wetzlar micromanipulator. Transgenic organisms were created by microinjecting the fosmid *WRM0615dH10* into the nematode's gonad using these needles. Injected PCR constructs had a finale concentration of 3ng/μl, and were injected along with *pCeh361*(100ng/μl) and *myo-2::GFP* (10ng/μl). *pCeh361* is a *dpy-5* rescue construct and *myo-2* is driven by the *myo-2* promoter and is used as a GFP marker that expresses in the pharynx, including pharyngeal bulbs.

Using mineral oil (Sigma, M-3516), *dpy-5*/*dpy-5* worms were placed atop dry agarose pads covered by 48 x 65mm microscope cover slips (Gold Seal Cover Glass,

reorder number 3335). Olympus BH2-HLSH or Zeiss 47 3016 inverted microscopes were used to conduct all microinjections. The injected worms were plated as (five P_0s per plate). All the injections were done by Domena K. Tu. To isolate the construct carrying worms, from the injected strains, wild-type F1s were individually plated and allowed to self. The only lines with wild-type (non-*dpy*) F_2 lines with the fosmid and expressing GFP were selected to set up a transgenic line; only one wild-type F_2 line was kept per original P_0 plate. Fosmid strain generated: BC8556, seX2251; *dpy-5(e907)/dpy-5(e907)[WRM0615dH10+pCeh361+ myo-2::GFP]*. Fosmid PCR check was conducted using PCR primers designed by ShuYi Chua specific to the fosmid backbone. Fosmid-F: GCGAC CACGT TTTAG TCTAC G, Fosmid-R: TCAAT ACTTG CCCTTG CCCTT GACAG G.

1.2.8 *let-65* rescue using fosmid *WRM0615dH10*

WRM0615dH10 fosmid was used to rescue *let-65 (s1083 and 1222).* This fosmid covers *F58B3.5* (*mars-1*) genomic region but not *unc-22*. The fosmid was microinjected into worms as described in the previous section and the generated worms with *dpy-5(e907)/dpy-5(e907) [WRM0615dH10+myo-2::GFP+pCeh361]* genotype were crossed to +/+ N2 males. This cross obtained males with *dpy-5(e907)/+ [WRM0615dH10+myo-2::GFP+pCeh361]* genotype. Only the GFP- expressing males (carrying the fosmid) were isolated and then crossed to *let-65(s1083* or *s1222):* Subsequently, wild type F_1 hermaphrodites allowed to self in separate plates. F_2 progeny were screened for rescued worms that have a phenotype of unconditional twitchers. These have the genotype *let-65(s1083* or *s1222) unc-22(s7) /let-65*(*s1083* or *s1222*) *unc-22(s7) (IV) [WRM0615dH10+myo-2::GFP+pCeh361]*. To indicate a successful fosmid rescue, F_2

unconditional twitchers individuals were separately plated and incubated at 20ºC for four to five days and then screened for full rescue indicated by the presence of progeny or partial rescue when no progeny were observed.

1.3 Results

1.3.1 *F58B3.4* and *F58B3.5* sequencing results

No mutations were found in *F58B3.4* in the sequenced *let-65(s254).* In the other hand, *F58B3.5* was sequenced in eight out of nine alleles and in all cases, G-C to A-T transitions were found. The mutations found in *F58B3.5* are consistent with EMS mutagenesis (Fig. 3). Transitions found in *s1083*, *s1084*, *s254*, *s1154*, *s1730*, *s694*, *s1777* are in well-conserved residues. The mutation in *s1777* truncates the protein prematurely at 408 amino acids, while the mutation in *s1222* mutates the first base in the first intron thus preventing splicing of the first intron and resulting in a premature truncation of the protein at amino acid 28. Table 1 summarizes the nucleotide changes and their corresponding amino acid mutations for each allele. For domain locations of the found mutations, please refer to figure 4.

1.3.2 *let-65*(*s1083* and *s1222*) rescue using the fosmid *WRM0615dH10*

The fosmid *WRM0615dH10* that spans the genomic region containing *F58B3.5* but not *unc-22* fully rescued *let-65(s1083* or *s1222).* Lines were retained and designated BC8656 (sEX2251) for *s1083* and BC8696 (sEX2251) for *s1222.*

Table 1. EMS mutations found in *F58B3.5* alleles.

let-65 Allele	Nucleotide change	Amino acid change
s254	G1188A	Glu364Lys
s694	G670A	Gly186Glu
s1083	C429T	Gly126Glu
s1084	G922A	Gly275Arg
s1154	C1297T	Pro400Leu
s1222	G96A	-27Arg
s1730	C174T	Ala41Val
s1777	G1324A	Trp408Stop

1.4 Discussion

From previous work conducted in our lab and from some of the published RNAi results targeting *F58B3.5 (mars-1),* that result in larval lethal (Simmer *et al.* 2003) and embryonic lethal (Sonnichsen *et al.* 2005) and for *F58B3.4* resulted in larval lethal (Maeda *et al.* 2001), larval arrest (Simmer *et al.* 2003), I deduced that the most likely candidate genes for *let-65* are *F58B3.4* and *F58B3.5 (mars-1)*. These two candidate genes are located between the left breakpoints of *sDf10* and *sDf84*. By sequencing, I found that eight *let-65* alleles had mutations in *mars-1* thus showing it to be the corresponding gene. I then rescued *let-65*(*s1083)* and *let-65(s1222)* lethality by *WRM0615dH10,* which further proved that *mars-1* is the molecular identity of *let-65.*

Mars-1 encodes a methionyl tRNA synthetase. This enzyme is one of at least 20 aminoacyl tRNA synthetases enzymes in the family that catalyzes aminoacylation of their cognate tRNAs. This is a two-step process that involves activation of the amino

acid by ATP that produce aminoacyl adenylate in the first step (Equation 1), followed by transferring the amino acid onto the 3'-end of the tRNA (Jakubowski 2001).

$$AARS + AA + ATP \Longleftrightarrow AARS.AA\text{-}AMP + PPi \quad (1)$$

$$AARS.AA\text{-}AMP + tRNA^{AA} \Longleftrightarrow AARS + AA\text{-}tRNA^{AA} + AMP \quad (2)$$

AARS enzymes are multi-domain proteins. These proteins are divided into two major classes (Eriani *et al.* 1990) in which the catalytic domain is conserved among all members of each class (O'Donoghue and Luthey-Sculten 2003). Catalytic domain of Class I enzymes exhibit a typical Rossmann fold consisting of three α/β/α topology layer. This fold has an inner core of five parallel beta sheets (O'Donoghue and Luthey-Sculten 2003). Class I enzymes also exhibit two conserved consensus motifs, the HIGH (His-Ile-Gly-His) and KMSKS (Lys-Met-Ser- Lys-Ser) (O'Donoghue and Luthey-Sculten 2003). The class II synthetases exhibit a fold of mixed α+β with a central core of anti-parallel β-sheets flanked by α helices (O'Donoghue and Luthey-Sculten 2003). This fold grants the structure a rigid template for amino acid and ATP binding (Ochsner *et al.* 2007).

Class I AARS enzymes: CysRS, MetRS, ValRS, IIeRS, LeuRS, ArgRS, GlnRS, GluRS, LysRS, TyrRS, and TrpRS.

Class II AARS enzymes: AlaRS, GlyRS, AspRS, AsnRS, HisRS, PheRS, ProRS, SerRS, ThrRS, and LysRS.

Lysyl-tRNA synthetase (LysRS) is the only AARS enzyme that has members in both classes. Class I LysRS has been found in most Archaea and in some Bacteria, while Class II LysRS has been found in all examined eukaryotic genomes, the majority of Bacteria and in some eukaryotes (O'Donoghue and Luthey-Sculten 2003).

MetRS recognizes an initiator tRNA as well as the tRNA delivering methionine for elongation of translated protein chains (Deniziak and Barciszewski 2001). MetRS is evolutionary well conserved with 58% identity and 71% similarity between the human and *C.elegans* proteins (BlastP alignment, (Altschul *et al.* 1997; Altschul 2005)). Seven out of eight of the identified G/C to A/T transitions in *mars-1* resulted in amino acid changes in highly conserved residues. One of these changes resulted in a pre-maturely truncated protein *let-65*(*s1777)*, while another mutation resulted in mis-splicing the first intron thus also giving a truncated protein *let-65(s1222)* (Fig. 3).

MARS-1 in *C. elegans* is a 917 amino acids long protein that consists of multiple functional domains. The catalytic core domain (from amino acid 36 to amino acid 407 (Marchler-Bauer *et al.* 2009)) is the main domain and is distinguished by the Rossmann fold that is a characteristic domain of Class I AARS enzymes. In *let-65* alleles*,* six out of the eight identified mutations are in the core domain (*s1730*, *s1083*, *s694*, *s1084*, *s254* and *s1154*) (Fig. 4). The tRNA binding domain (from 755 aa to 857 aa (Marchler-Bauer *et al.* 2009)) and the anticodon-binding domain (from amino acid 416 to amino acid 545 (Marchler-Bauer *et al.* 2009)). The EMS induced mutation in *s1777* creates a truncated protein that lacks both the tRNA binding domain and the anticodon-binding domain (Fig. 4).

There are 36 known AARS genes in *C. elegans.* So far, *let-65* (*mars-1*) is the only genetically identified one with a lethal phenotype. Nevertheless, all but one of the AARS genes have lethal RNAi phenotypes (WS208, 27-Nov-2009) (Table 2). In *C. elegans,* AARS genes have a non-uniform distribution throughout the genome and are on five of six chromosomes (Fig. 5). LGI contains the largest number (12) of the AARS genes:

aars-2, *cars-1*, *cars-2*, *ears-1*, *fars-1*, *iars-2*, *Iars-2*, *nars-1*, *pars-1*, *vars-2*, *Y105E8A.19*, and *Y105E8A.20*. LGIV contains eight: *ears-2*, *qars-1*, *hars-1*, *iars-1*, *mars-1*, *sars-1*, *sars-2*, and *yars-1*. LGIII contains seven: *dars-1*, *pars-1*, *wars-2*, *gars-1*, *rart-1*, *lars-1*, and *Y66D12A.23*. LGII contain five AARS: *aars-1*, *fars-3*, *kars-1*, *rars-2*, and *tars-1*. LGV contains four genes: *fars-2*, *wars-1*, *dars-2*, and *vars-1*. None of the AARS genes is on LGX. This could be explained by the fact that X-linked genes are transcriptionally silenced (Kamath *et al.* 2003) in which hermaphrodite worms have half the X chromosome expression during mitosis and early meiosis, therefore, necessary genes for essential cellular processes, which are important for the viability of all cells, are expected to be absent from LG-X (Kamath *et al.* 2003).

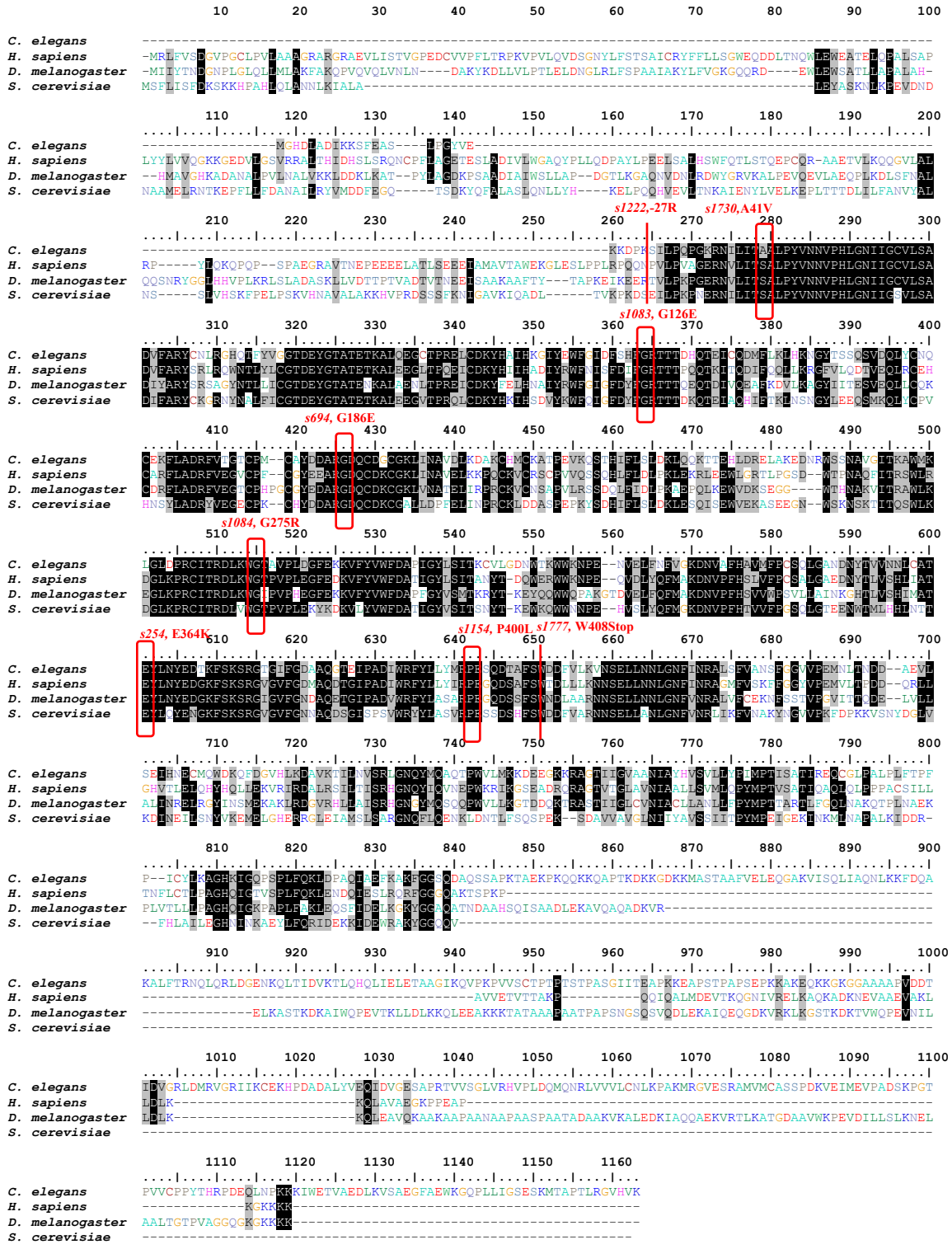


Figure 3. **MARS-1 conservation and location of amino acid mutations in *let-65* alleles. BioEdit alignment of mars-1 with its orthologs in human, *D. melanogaster,* and *S. cerevisiae. let-65* mutations occur in well conserved residues. MARS-1 is evolutionary well conserved with 58% identity and 71% similarity between the human and *C. elegans* orthologs (BlastP alignment, (Altschul *et al.* 1997; Altschul 2005)). For domain boundaries, please refer to Fig. 4).**

Table 2. AARS enzymes family, their protein size and observed RNAi phenotype (WS215, 18-jun-2010).

AARS enzyme	Ch location	Gene and sequence name	Protein size	Observed RNAi phenotype
Alanyl (A) tRNA synthetase	II	*aars-1, W02B12.6*	793 aa	5
Alanyl (A) tRNA synthetase	I	*aars-2, F28H1.3*	968 aa	1, 2, 4
Cysteinyl (C) tRNA synthetase	I	*cars-1, Y23H5A.7*	909 aa	1, 2, 4
Cysteinyl (C) tRNA synthetase	I	*cars-2, Y23H5A.1*	Not attached to a protein	4
Aspartyl (D) tRNA synthetase	III	*dars-1, B0464.1*	531 aa	1, 2, 4
Aspartyl (D) tRNA synthetase	V	*dars-2, F10C2.6*	593 aa	1, 3, 4
Glutamly (E) tRNA synthetase	I	*ears-1, ZC434.5*	1149 aa	1, 2
Glutamly (E) tRNA Synthetase	IV	*ears-2, T07A9.2*	481 aa	1, 2, 3, 5
Phenylalanyl (F) tRNA synthetase	I	*fars-1, T08B2.9*	552 aa	1, 2, 3, 4
Phenylalanyl (F) tRNA synthetase	V	*fars-2, Y60A3A.13*	458 aa	1, 4
Phenylalanyl (F) tRNA synthetase	II	*fars-3, F22B5.9*	591 aa	1, 2, 4, 5
Glycyl (G) tRNA synthetase	III	*gars-1, T10F2.1*	742 aa	1, 2, 4
Histidyl (H) tRNA synthetase	IV	*hars-1, T11G6.1*	520 aa	1, 2
Isoleucyl (I) tRNA synthetase	IV	*iars-1, R11A8.6*	1141 aa	1, 4, 5
Isoleucyl (I) tRNA synthetase	I	*iars-2, C25A1.7*	970 aa	1
Lysyl (K) tRNA synthetase	II	*kars-1, T02G5.9*	572 aa	1, 2, 4, 5
Leucyl (L) tRNA synthetase	III	*lars-1, R74.1*	1186 aa	1, 2, 4

AARS enzyme	Ch location	Gene and sequence name	Protein size	Observed RNAi phenotype
Leucyl (L) tRNA synthetase	I	*lars-2, ZK524.3*	859 aa	1, 4
Methionyl (M) tRNA synthetase	IV	*mars-1, F58B3.5*	917 aa	1, 2, 4
Methionyl (M) tRNA synthetase	I	*Y105E8A.20*	406 aa	1, 5
Asparaginyl (N) tRNA synthetase	I	*nars-1, F22D6.3*	545 aa	1, 2, 4
Asparaginyl (N) tRNA synthetase	III	*Y66D12A.23*	448 aa	1, 4
Prolyl (P) tRNA synthetase	III	*pars-1, T20H4.3*	581 aa	1, 2, 4
Prolyl (P) tRNA synthetase	I	*pars-2, T27F6.5*	454 aa	X
Glutaminyl (Q) tRNA Synthetase	IV	*qars-1, Y41E3.4*	786 aa	1, 2
Arginyl (R) tRNA synthetase	III	*rars-1, F26F4.10*	713 aa	1, 2, 4
Arginyl (R) tRNA synthetase	II	*rars-2, C29H12.1*	512 aa	1, 3
Seryl (S) tRNA synthetase	IV	*sars-1, C47E12.1*	487 aa	1, 2, 4
Seryl (S) tRNA synthetase	IV	*sars-2, W03B1.4*	441 aa	1
Threonyl (T) tRNA synthetase	II	*tars-1, C47D12.6*	725 aa	1, 2, 5
Valyl (V) tRNA synthetase	V	*vars-1, ZC513.4*	918 aa	1, 2, 3
Valyl (V) tRNA synthetase	I	*vars-2, Y87G2A.5*	1050 aa	1, 2, 4, 5
Tryptophanyl (W) tRNA synthetase	V	*wars-1, Y80D3A.1*	417 aa	1, 2, 5
Tryptophanyl (W) tRNA synthetase	III	*wars-2, C34E10.4*	650 aa	1, 2

AARS enzyme	Ch location	Gene and sequence name	Protein size	Observed RNAi phenotype
Tyrosiny (Y) tRNA synthetase	IV	*yars-1, K08F11.4*	447 aa	1, 2, 3, 4
Tyrosiny (Y) tRNA synthetase	I	*Y105E8A.19*	722 aa	1, 2, 4

All AARS genes have lethal RNAi phenotypes (WS215, 18-jun-2010). AARS genes have a non-uniform genomic distribution over the autosomes in *C. elegans*. RNAi phenotypes: 1. Embryonic lethal, 2. Larval lethal, 3. Sterile, 4. Maternal sterile, 5. sterile progeny, X. No lethal phenotypes observed.

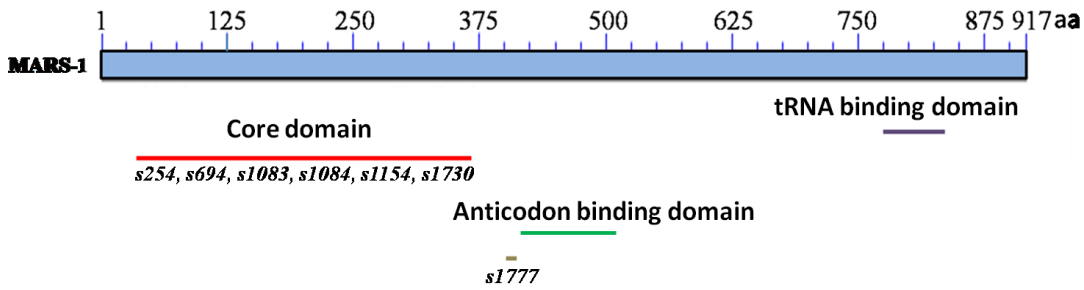


Figure 4. **Graphical representation of MARS-1 domains and positions of *let-65* found mutations. Blue bar represents MARS-1. Numbers indicate amino acid positions. Red, green, and purple bars represent the three major domains: Core domain (the catylatic domain), anticodon binding domain and tRNA binding domain (respectively). Seven out of the eight identified *let-65* mutations are located in the core domain. No mutations were found in the anticodon binding domain or the tRNA binding domain. The brown bar indicates the location of the point mutations in *let-65*(*s1777)* (Modified from Marchler-Bauer et al. 2009).**

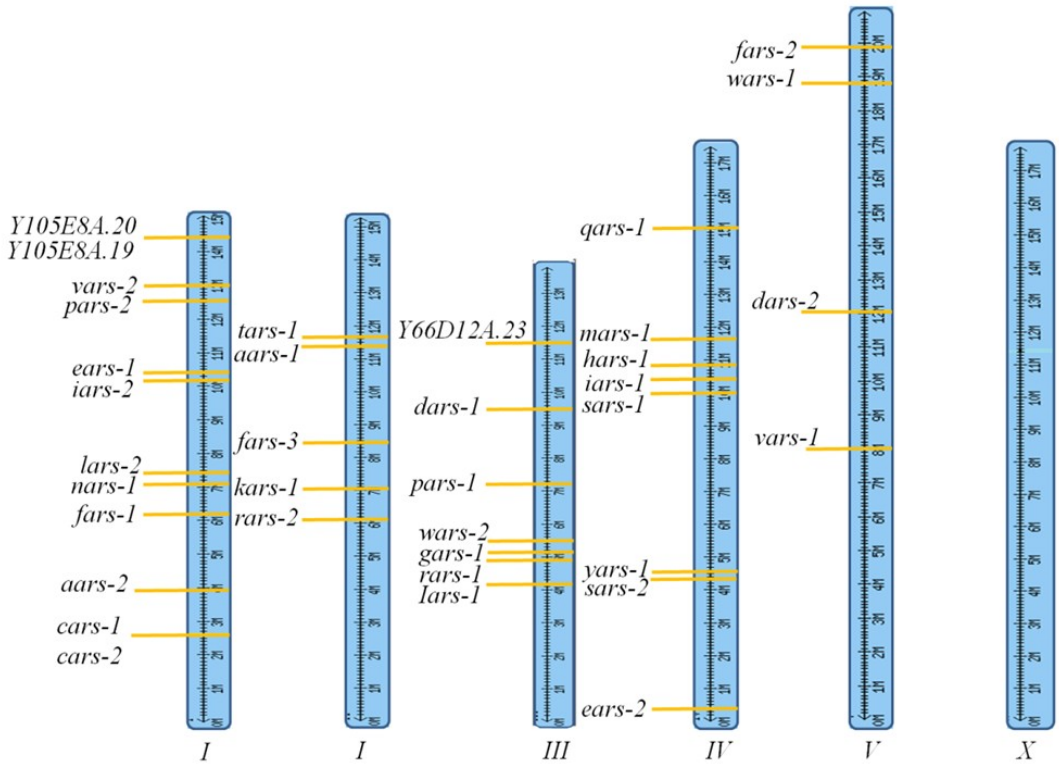


Figure 5. **AARS genes distribution in *C. elegans'* genome.**
LGI contains the largest number (12) of the AARS genes; followed by LGIV with eight genes. LGIII contains seven genes, LGII contains five AARS genes and LGV contains four genes. No AARS genes are on LGX.

CHAPTER 2: AARS SUB-CELLULAR LOCALIZATIONS

2.1 Introduction

Eukaryotic cells have functionally distinct, membrane-enclosed compartments and organelles each of which contains its own characteristic set of enzymes and other specialized molecules (Herrmann and Neupert 2000). In eukaryotic cells, most proteins are synthesized by cytosolic ribosomes, while some are synthesized by mitochondrial ribosomes (Herrmann and Neupert 2000). Proteins that do not reside in the cytosol as their final destination have amino acid sequences (sorting signals) that are specific to direct them to their correct compartment (Paetzel *et al.* 2002).

There are at least two types of sorting signals. One type consists of terminal sequences (Paetzel *et al.* 2002). Once the sorting process has been completed, signal peptidases remove some of these signals from their proteins (Paetzel *et al.* 2002). The second sorting signal type consists of internal amino acids residues that are not adjacent in linear amino acid sequences. Unlike the first type, these signals are not processed by the signal peptidases and persevere in the finished protein (Lingappa *et al.* 1979). Examples of these proteins with such signals include proteins that are translated in endoplasmic reticulum (ER) membrane-bound ribosomes (Palade and Siekevitz, 1956). Each particular protein destination requires specific signal sequences. Mitochondrial proteins have an N-terminal signal sequence (Paetzel *et al.* 2002) which mostly consists of positive and hydrophobic amino acids (Herrmann and Neupert 2000).

To better understand a eukaryotic cell, it is essential to know where proteins locate and function. Protein sub-cellular localization can be determined either by experimental methods or by computational tools. I experimentally characterized *mars-1's* sub-cellular localization in *C. elegans.* To do this I made use of a reporter construct in which GFP encoding sequence was fused, in-frame, with *mars-1* amino acid encoding sequence. I also used computational methods (BlastP (Altschul *et al.* 1997; Altschul *et al.* 2005) and WolF PSORT, (Horton *et al.* 2007, http://wolfpsort.org/) to determine the sub-cellular localization of all known *C. elegans* AARS enzymes. In order to confirm the computational results, I experimentally characterized the sub-cellular localization for two AARS genes *srs-1* and *srs-2*. As a consequence of this analysis of this chapters work, and in consultation with Dr. J. Hodgkin, Oxford University, we renamed the *C. elegans* AARS genes from *xrs-* to *xars-* and I recommended numbering the genes as "xars-1" for genes encoding cytoplasmic proteins and "xars-2" for genes encoding mitochondrial proteins. In cases where there is a second cytoplasmic enzyme xars-3 is to be used. However, in this chapter I will use the previous AARS gene nomenclature as it was before I modified them in this thesis to minimize nomenclature confusion.

2.2 Materials and Methods

2.2.1 Construction of translational transgenes

Fusion PCR was used to generate *let-65*$_{\text{promoter}}$:: *let-65:: GFP* (Fig.6) (Hobert 2002) using DNA template amplified from N2 genomic DNA (Bristol, Baillie Laboratory strain, BC49). The encoding sequences with their promoter containing region were fused upstream to GFP encoding sequence that was amplified off from the *pPD95.75* (GFP) encoding cassette. *pPD95.75* contains artificially introns, and a 3' un-translated region

(UTR) from *C. elegans' unc-54* (Fire Vector Kit, 1995; Boulin, et al, 2006). Amplification of the GFP reporter was carried out using these primers: GFP-C: AGCTT GCATG CCTGC AGGTC G, and GFP-D: AAGGG CCCGT ACGGC CGACT A (Boulin *et al.* 2006).

The forward primer used to amplify *let-65* including 2,596 bases of upstream region of the ATG start codon (putative promoter containing region) was F58B3.5-A: CAAGG TGAAC AAATT AGAAA CGA. The reverse primer had additional sequence complementary to the forward GFP primer to allow the two PCR products to join together. The reverse Primer was *F58B3.5*-B: AGTCG ACCTG CAGGC ATGCA AGCTT TTCAC ATGGA CTCCA CGAA (the additional primer sequence obtained from Boulin *et al.* 2006). The nested primers used to fuse the two PCR products, at *F58B3.5* amino acid 916, were *F58B3.5*-A*: AACGG CAAGC TGAAA TGAAT and GFP-D*: GGAAA CAGTT ATGTT TGGTA TA (Boulin, et al, 2006). *let-65 (F58B3.5)* sequences used for primer design were isolated from WormBase (WS208, 27-Nov-2009). Phusion polymerase (Finnzymes, New England Biolabs Cat: F530) was used for all PCR reactions to ensure the fidelity of the constructs.

The same method was used to generate *srs-1* (*W03B1.4,* now named *sars-2*) and *srs-2* (*C47E12.1,* now named *sars-1*) constructs. To generate the *srs-1* translational fusion construct, $W03B1.4_{(p\text{-}1,870)}$:: *W03B1.4::* GFP (Fig 6), a forward primer used to amplify *srs-1* (*W03B1.4*) with 1,870 bases of the region upstream the ATG start codon and it was *W03B1.4*-A: ACCCC AAGTT CTCAT TGCTG. The gene's reverse primer, which contains additional sequence complementary to the forward GFP cassette primer, to allow joining of the two PCR products was *W03B1.4-B*: AGTCG ACCTG

CAGGC ATGCA AGCTA AGTGA TTTCG CTGGT TGG. The forward nested primer used to fuse the two PCR products (*srs-1* (*W03B1.4*) gene and GFP gene) at *W03B1.4* amino acid (486) is *W03B1.4*-A*: ACAGC TTGGT GAGCT TCAAA. For srs-2 translational fusion construct, *C47E12.1*$_{(p\text{-}2,044)}$:: *C47E12.1::* GFP (Fig 6), the forward primer used to amplify *srs-2* (*C47E12.1*) with 2,044 bases of region upstream of the ATG start codon was *C47E12.1*-A: CCACG GCCTC TACAA TGAAT. The gene's reverse primer, which contains additional sequence complementary to the forward GFP cassette primer, to join the two PCR products together was *C47E12.1*-B: AGTCG ACCTG CAGGC ATGCA AGCTC TTTCC TGTCG CCTTT TTG. The forward nested primer used to fuse the two PCR products (*srs-2* (*W03B1.4*) and the GFP cassette) at *W03B1.4* amino acid (440) was *C47E12.1*-A*: TGAGC ACGTC GTAGA GAACA A. The same primers that were used to amplify GFP for generating the *mrs-1*(now named *mars-1*) construct, was used to generate the *srs-1* (now named *sars-2*) and *srs-2* (now named *sars-1*) constructs.

2.2.2 Microinjection

Refer to section 1.2.7 Materials and Methods for a detailed description of microinjection. The constructs that I generated were co-injected with *pCeh361* at 100ng/µl. All injections were done by Domena K Tu. Generated lines were retained and designated BC8933 (sEx2893) and BC8934 (sEx2894) for *F58B3.5*, BC9056 (sEx2999) and BC9057 (sEx2100) for *srs-2 W03B1.4,* and BC9056 (sEx2999) and BC9066 (sEx2100) for *srs-1 (C47E12.1)*.

2.2.3 Microscopy

Screening of transgenic strains was done using a spinning disk confocal microscope. Worms were immobilized on moist agarose (2%) pads using 1mM levamisole (in M9) (Signa, L9756-5G). Images were processed using Velocity Software. Screening of *srs-1* and *srs-2* transgenic strains was done using a Zeiss Axioscope equipped with a QImaging camera. Images were process using Adobe Photoshop CS. All figures were taking in 1s exposure time.

2.2.4 Computational prediction of AARS sub-cellular localizations

Prediction of the sub-cellular localization of *C. elegans* AARS enzymes was conducted by using two programs: WoLF PSORT, which predicts protein localization sites from the proteins amino acids sequences (Horton *et al.* 2007), http://wolfpsort.org/) and BlastP which searches the query protein in protein databases (Altschul *et al.* 1997; Altschul 2005). For BlastP scores and E values please refer to Appendix C.

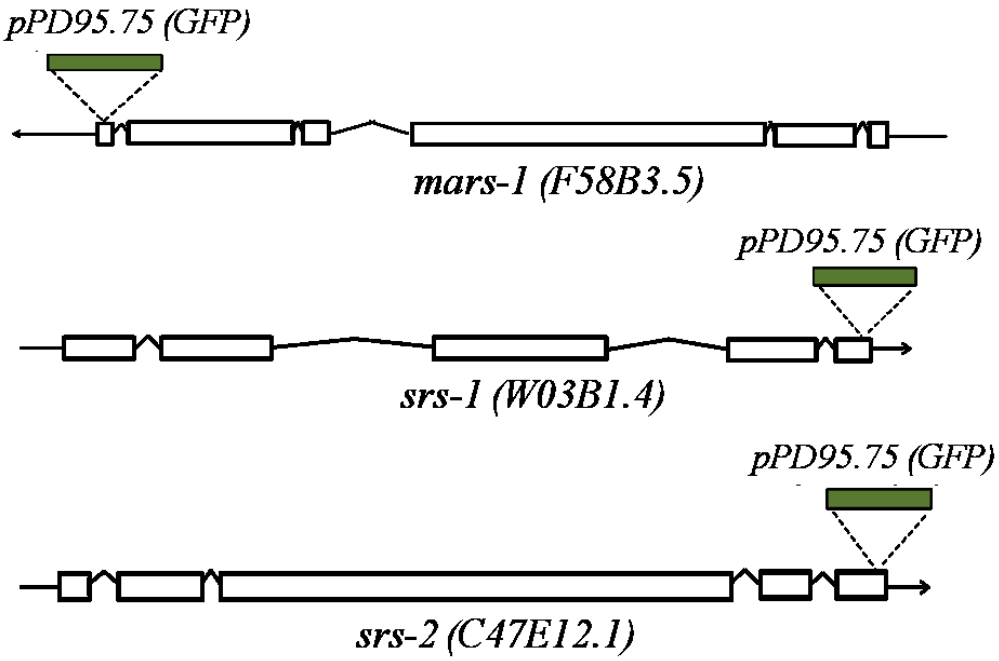


Figure 6. **Gene models of generated translational fusion constructs.**
GFP cassette was fused into the last exon of the translation constructs. A. GFP was fused at amino acid 916 of *F58B3.5*, B. amino acid 486 of *srs-1* (*W03B1.4,* now named *sars-2*) and C. amino acid 440 of *srs-2* (*C47E12.1,* now named *sars-1*). The shown models are obtained from WormBase, WS215 and do not represent the proportional sizes of the genes.

2.3 Results

2.3.1 *F58B3.5* sub-cellular localization based on the expression of *let-65*$_{(p-2,596)}$:: *let-65::* GFP translational fusion

To determine the sub-cellular localization of *F58B3.5,* I used a translation fusion construct of GFP encoding sequence fused to the gene at amino acid 916. GFP expression was visible in the pharyngeal bulbs, pharynx, and anterior and posterior neurons in adult. I looked for expression at the sub-cellular level and the expression of the construct was in the cytoplasm (Fig.7). No expression was observed in the cell's nucleus or mitochondria.

2.3.2 Prediction of AARS sub-cellular localizations

Because AARS enzymes are involved in protein translation in both cytosol and mitochondria, the algorithm WoLF PSORT, which predicts proteins sub-cellular localization sites by their amino acid sequences (Horton *et al.* 2007), was used to detect the sub-cellular localization of every AARS enzyme in *C. elegans*. Using WoLF PSORT I predicated 20 AARS enzymes localized to the cytoplasm (*ars-1*, *ars-2*, *drs-1*, *ers-1, ers-3*, *frs-1*, *frs-2*, *hrs-1*, *irs-1*, *krs-1*, *mrs-1*, *Y10.5E8A.20*, *nrs-1, nrs-2, prs-1*, *rrt-1*, *rrt-2*, *srs-2*, *wrs-1,* and *Y105E8A.19*). I predicted using WoLF PSORT that eight enzymes target the mitochondria (*drs-2*, *ers-2*, *frs-3*, *grs-1*, *irs-2*, *prs-2*, *trs-1,* and *yrs-1*); six are nuclear proteins (*lrs-1*, *lrs-2, Y66D12.A23*, *srs-1*, *vrs-2,* and *wrs-2*), one enzyme was predicted to be a plasma membrane protein (*crs-1*), and one to functions in the cytoskeleton (*vrs-1*) (Table. 3). *lrs-2* was previously confirmed to be a mitochondrial protein by translational fusion (Lee et al. 2003). A mutation in *lrs-2* (*mg312)* was found to extend lifespan (Lee et al. 2003).

For a further confirmation of the WoLF PSORT results, AARS enzymes proteins sequences obtained from WormBase (WS209) were used as query sequences to search swissport database using BlastP. Out of the BlastP matches, *Saccharomyces cerevisiae* genes were used to make the final judgment of the protein localization in *C. elegans. S.cerevisiae* was used mainly because it's one of the simplest in regards of identifying open reading frames (ORFs), and it is the most soundly characterized eukaryotic genome (Fisk et al. 2006).

S.cerevisiae BlastP matches for *C. elegans* AARS enzymes, predict that 18 of the AARS genes encode enzymes with the cytoplasm as their final locations (*crs-1*, *drs-1*, *ers-1*, *ers-2*, *frs-1, frs-2*, *irs-1*, *krs-1*, *lrs-1*, *mrs-1*, *nrs-1, nars-2*, *prs-1*, *rrt-1*, *srs-2*, *trs-1*, *wrs-1,* and *Y105E8A.19*. It also predicted that 12 AARS genes encodes enzymes that are mitochondrial proteins (*drs-2*, *ers-3*, *frs-3*, *irs-2*, *lrs-2*, *Y105E8A.20*, *Y66D12A.23*, *prs-2*, *rrt-2*, *srs-1*, *wrs-2,* and *yrs-1)*. ARS-1 and ARS-2 were matched with ALA1 (YOR335C) that encodes Alanyl tRNA synthetase, a protein that functions in the cytoplasm and mitochondria. Moreover, VRS-1 and VRS-2 were matched by BlastP to the same gene (VAS1) which encodes both cytoplasmic and mitochondrial Valyl tRNA synthetase. GRS-1 matched to GRS1 in *S. cerevisiae* that have both cytoplasmic and mitochondrial localizations. Similarly, HRS-1 matched to HTS-1 that has both cytoplasmic and mitochondrial localizations. Four enzymes (LysRS, GlyRS, HisRS, and TyrRS) have only one gene each that encodes a cytoplasmic enzyme with no identified genes in *C. elegans* for the mitochondrial enzyme (Table. 3).

BlastP matches showed that NRS-2 has ANA2 (YGR124W) as its *S. cerevisiae* BlastP match, and this gene encodes asparagine synthetase. *H. sapiens* and *S. pombe*

matched genes also encode Asparagine synthetase and not asparaginyl tRNA synthetase. Moreover, three genes were found to encode Phenylalanyl tRNA synthetase: *frs-1* and *frs-2* (both cytoplasmic) and *frs-3* (mitochondrial).

2.3.3 The sub-cellular localization of *srs-1* (*Wo3B1.4,* now named *sars-2*) and *srs-2* (*C47E12.1,* now named *sars-1*) by translational fusion.

To confirm BlastP and WoLF PSORT predictions for AARS sub-cellular localizations, translational fusions were made for two AARS genes *srs-1* (*Wo3B1.4,* now named *sars-2*) and *srs-2* (*C47E12.1,* now named *sars-1*). To generate a *srs-1* (*Wo3B1.4*) construct, GFP encoding sequence was fused to the gene at amino acid 486. In the sub-cellular level GFP expression was visible in the mitochondria. To generate a *srs-2* (*C47E12.1*) construct, GFP encoding sequence was fused to the gene at amino acid 440. GFP expression was visible in the cytoplasm (Fig.7).

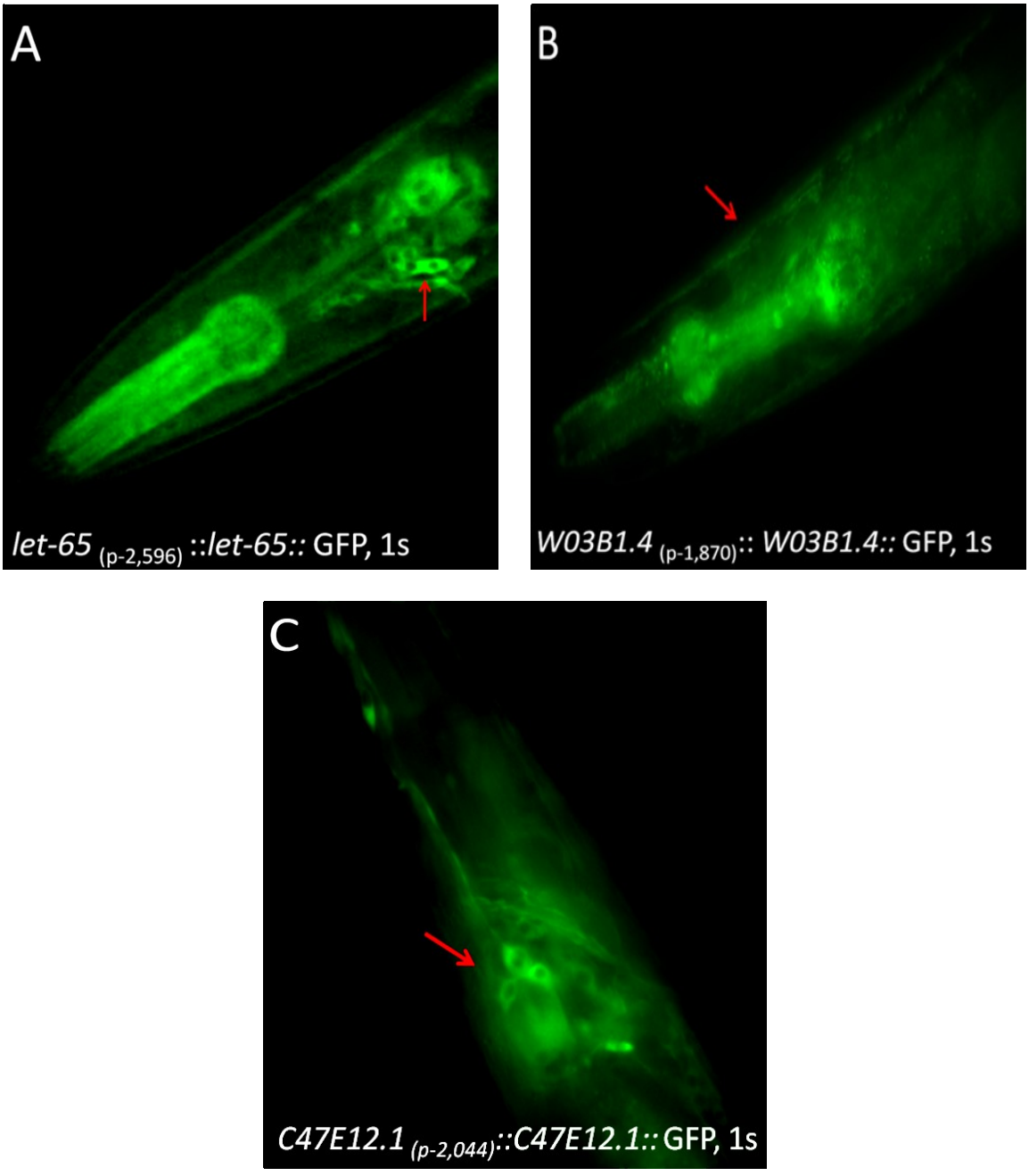


Figure 7. **Expression of *F58B3.5*, *Wo3B1.4*, and *C47E12.1* translational fusions.**
(A) Expression of the *F58B3.5* transgene is visible in the pharyngeal bulbs, pharynx, and anterior neurons. (B) *Wo3B1.4* expression is visible in all cells and is localized to the mitochondria. (C) *C47E12.1* expression was also visible in all cells. However the expression is localized to the cytoplasm.

Table 3. ***C. elegans aars*** **revised nomenclature, and BlastP and WoLF PSORT sub-cellular localizations.**

AARS enzyme	Previous genes and sequence Nomenclature	Revised genes Nomenclature	WoLF PSORT localization	BlastP localization based on *S. cerevisiae* matches	*S. cerevisiae* BlastP matches
Alanyl (A) ARS	*ars-1, W02B12.6*	*aars-1, W02B12.6*	Cyto	Cyto/ Mito	ALA1, YOR335C
Alanyl (A) ARS	*ars-2, F28H1.3*	*aars-2, F28H1.3*	Cyto	Cyto/ Mito	ALA1, YOR335C
Cysteinyl (C) ARS	*crs-1, Y23H5A.7*	*cars1, Y23H5A.7*	plas	Cyto	YNL247W
Cysteinyl (C) ARS	*crs-2, Y23H5A.1*	*cars-2, Y23H5A.1*	-	-	-
Aspartyl (D) ARS	*drs-1, B0464.1*	*dars-1, B0464.1*	Cyto	Cyto	DPS1, YLL018C
Aspartyl (D) ARS	*drs-2, F10C2.6*	*dars-2, F10C2.6*	Mito	Mito	MSD1, YPL104W
Glutaminyl (Q) ARS	*ers-1, Y41E3.4*	*qars-1, Y41E3.4*	Cyto	Cyto	GLN4, YOR168W
Glutamyl (E) ARS	*ers-3, T07A9.2*	*ears-2, T07A9.2*	Cyto	Mito	MSE1, YOL033W
Glutamyl (E) ARS	*ers-2, ZC434.5*	*ears-1, ZC434.5*	Mito	Cyto	GUS1, YGL245W
Phenylalanyl (F) ARS	*frs-1, T08B2.9*	*fars-1, T08B2.9*	Cyto	Cyto	FRS2, YFL022C
Phenylalanyl (F) ARS	*frs-3, Y60A3A.13*	*fars-2, Y60A3A.13*	Mito	Mito	MSF1, YPR047W
Phenylalanyl (F) ARS	*frs-2, F22B5.9*	*fars-3, F22B5.9*	Cyto	Cyto	FRS1, YLR060W
Glycyl (G) ARS	*grs-1, T10F2.1*	*gars-1, T10F2.1*	Mito	Cyto/ Mito	GRS1, YBR121C
Histidyl (H) ARS	*hrs-1, T11G6.1*	*hars-1, T11G6.1*	Cyto	Cyto/ Mito	HTS1, YPR033C
Isoleucyl (I) ARS	*irs-1, R11A8.6*	*iars-1, R11A8.6*	Cyto	Cyto	ILS1, YBL076C
Isoleucyl (I) ARS	*irs-2, C25A1.7*	*iars-2, C25A1.7*	Mito	Mito	ISM1, YPL040C

AARS enzyme	**Previous genes and sequence Nomenclature**	**Revised genes Nomenclature**	**WoLF PSORT localization**	**BlastP localization based on *S. cerevisiae* matches**	***S. cerevisiae* BlastP matches**
Lysyl (K) ARS	*krs-1, T02G5.9*	*kars-1, T02G5.9*	Cyto	Cyto	KRS1, YDR037W
Leucyl (L) ARS	*lrs-1, R74.1*	*lars-1, R74.1*	Nul	Cyto	CDC60, YPL160W
Leucyl (L) ARS	*lrs-2, ZK524.3*	*lars-2, ZK524.3*	Nul	Mito	NAM2, YLR382C
Methionyl (M) ARS	*mrs-1, F58B3.5*	*mars-1, F58B3.5*	Cyto	Cyto	MES1, YGR264C
Methionyl (M) ARS	*Y105E8A.20*	*mars-2 Y105E8A.*20	Cyto	Mito	MSM1, YGR171C
Asparaginyl (N) ARS	*nrs-1, F22D6.3*	*nars-1, F22D6.3*	Cyto	Cyto	DED81, YHR019C
Asparaginyl (N) ARS	*nrs-2, F25G6.6*	-	Cyto	Cyto	ANA2, YGR124W
Asparaginyl (N) ARS	*Y66D12A.23*	*nars-2*	Nul	Mito	SLM5, YCR024C
Prolyl (P) ARS	*prs-1, T20H4.3*	*pars-1, T20H4.3*	Cyto	Cyto	YHR020W
Prolyl (P) ARS	*prs-2, T27F6.5*	*pars-2, T27F6.5*	Mito	Mito	AIM10, YER087W
Arginyl (R) ARS	*rrt-1, F26F4.10*	*rars-1, F26F4.10*	Cyto	Cyto	YDR341C
Arginyl (R) ARS	*rrt-2, C29H12.1*	*rars-2, C29H12.1*	Cyto	Mito	MSR1, YHR091C
Seryl (S) ARS	*srs-1, W03B1.4*	*sars-2, W03B1.4*	Nul	Mito	DIA4, YHR011W
Seryl (S) ARS	*srs-2, C47E12.1*	*sars-1, C47E12.1*	Cyto	Cyto	SES1, YDR023W
Threonyl (T) ARS	*trs-1, C47D12.6*	*tars-1, C47D12.6*	Mito	Cyto	THS1, YIL078W
Valyl (V) ARS	*vrs-1, ZC513.4*	*vars-1, ZC513.4*	Cysk	Cyto/Mito	VAS1, YGR094W
Valyl (V) ARS	*vrs-2, Y87G2A.5*	*vars-2, Y87G2A.5*	Nul	Cyto/Mito	VAS1, YGR094W
Tryptophanyl (W) ARS	*wrs-1, Y80D3A.1*	*wars-1, Y80D3A.1*	Cyto	Cyto	WRS1, YOL097C

AARS enzyme	Previous genes and sequence Nomenclature	Revised genes Nomenclature	WoLF PSORT localization	BlastP localization based on *S. cerevisiae* matches	*S. cerevisiae* BlastP matches
Tryptophanyl (W) ARS	*wrs-2, C34E10.4*	*wars-2 C34E10.4*	Nul	Mito	MSW1, YDR268W
Tyrosinyl (Y) ARS	*yrs-1, K08F11.4*	*yars-2, K08F11.4*	Mito	Mito	MSY1, YPL097W
Tyrosinyl (Y) ARS	*Y105E8A.19*	*yars-1, Y105E8A.19*	Cyto	Cyto	TYS1, YGR185C

(-) Pseudogene. (Cyto) Cytoplasmic, (Mito) mitochondrial, (plas) plasma membrane, (nul) Nuclear, and (cysk) cytoskeleton. In the WormBase (WS214), *Y66D12A.23*, is not named as *nars-2* because I wasn't aware of this gene in the initial analysis. *Y105E8A.*20 is not named as *mars-2* and *K08F11.4* and *Y105E8A.19* are not named correctly as *Yars-1* and *Yars-2* respectively due to a mistake in my initial analysis.

2.4 Discussion

Methionyl tRNA synthetase plays a major role in protein translation in both the cytoplasm and the mitochondria. Both the experimental approach, using translational fusion of *mrs-1* (*mars-1*) and GFP reporters and the computational methods using the WoLF PSORT algorithm and BlastP searches, predicted cytoplasmic localization for *mrs-1* (*mars-1*). The cytoplasmic expression of *mrs-1* (*mars-1*), seen using the translational fusion construct, indicates that it functions in protein translation in cytoplasmic ribosomes.

WoLF PSORT (Horton *et al.* 2007, http://wolfpsort.org/) gives predicted protein sub-cellular localization sites based on amino acid sequences of these proteins. BlastP was used to search the Swiss-Port database with AARS amino acid sequences as protein queries (Altschul et al. 1997; Altschul 2005). BASTP searches confirmed the WoLF PSORT results for most of the AARS enzymes.

Using WoLF PSORT, I got predictions for eight proteins with localizations other than in the cytoplasm or the mitochondria: six were predicted to encode proteins with nuclear localization (*lrs-1*, *lrs-2, Y66D12.A23, srs-1*, *vrs-2* and *wrs-2*); one was predicted to encode a plasma membrane protein (*crs-1*) and one to function in the cytoskeleton (*vrs-1*). On the other hand, I got BlastP predictions of either cytoplasmic or mitochondrial localizations for all of them, which is consistent with their roles in translation (Table. 3). For these eight genes, the BlastP predictions gave cytoplasmic localization for *crs-1*, *lrs-1* and mitochondrial localization for *lrs-2*, *Y66D12A.23*, *srs-1*, and *wrs-2*.

In cases in which there were differences in the predicted localizations between WoLF PSORT and BlastP, *S. cerevisiae* BlastP matches were used to make the final determination of the query protein. WoLF PSORT predictions of non-cytoplasmic or non-mitochondrial localizations for some of the AARS enzymes might be an indication of the involvement of these enzymes in non-translational cellular processes. It has been previously reported that some AARS enzymes are multi-functional and have regulatory functions in different cell processes (Lee *et al.* 2004). For example, ThrRS in *Escherichia coli* was reported to block its synthesis by binding to its own mRNA leader sequence which prevents the 30S ribosomal subunit from binding to the ribosome-binding sequence of the mRNA (Lee *et al.* 2004). Moreover, *Neurospora crassa* mitochondrial TyrRS is involved in the splicing process of Group one introns (Akins and Lambowitz 1987). Human TyrRS was found to split into two fragments each exhibit cytokine activities (Lee *et al.* 2004). These various other potential functions of AARS could explain the non-cytoplasmic, non-mitochondrial localization predicted by WoLF PSORT algorithm for some of the enzymes.

BlastP searches using ARS-1 and ARS-2 gave matches with *S. cerevisiae* genes AlA1 (YOR335C). AIA1 encodes for Alanyl tRNA synthetase which functions in both cytoplasm and mitochondria. WoLF PSORT gave a score of 12.0 for ARS-2 to being a cytoplasmic protein and a score 4.0 to be a mitochondrial protein. Moreover, the *H. sapiens* BlastP match for ARS-2 is AARS (AlaRS) (score: 1139 and E value: 0) which encodes cytoplasmic Alanyl tRNA synthetase. This indicates that ARS-2 encodes a second cytoplasmic AlaRS protein or a protein that functions in both the mitochondria and the cytoplasm.

All VRS-1 and VRS-2 matched by BlastP to VAS1 which encodes cytoplasmic and mitochondrial Valyl tRNA synthetase. The *H. sapiens* BlastP match for VRS-1 is VARS2 (score: 533, E value: 3e-150) which encodes mitochondrial Valyl tRNA synthetase. However, VRS-2's best human match is VARS (score: 1109, E value: 3e-1050) that localizes to the cytosol. This indicates that VRS-1 might be a mitochondrial enzyme and VRS-2 encodes a cytoplasmic enzyme. The localization of these four genes (*ars-1, ars-2, vars-1,* and *vars-2*) can be confirmed experimentally by translational fusions.

GlnRS is absent in all known Archaea and most Bacteria. In these organisms, an indirect pathway is utilized for the formation of aminoacyl-tRNA. In this pathway, $tRNA^{Gln}$ is first misacylated with Glu by GluRS. Glu-$tRNA^{Gln}$ amidotransferase enzyme converts the misacylated Glu to Gln, which results in a correctly charged tRNA. But in eukaryotes, charging of tRNA with Gln is completed by the direct pathway previously described in Chapter one (O'Donoghue and Luthey-Sculten 2003). In WormBase (WS211) *Y41E3.4* is reported to encode a glutamyl (E) tRNA Synthetase (*ers-1*) and there was no gene that encodes glutaminyl (Q) tRNA Synthetase. However, the best result of a BlastP search for *Y41E3.4* encoded amino acids sequence based on *S. cerevisiae, D. melanogaster* and Human scores, suggests that *Y41E3.4* encodes GlnRS. GluRs was not one of the best BlastP results. Based on this, I propose that *Y41E3.4* encodes GlnRS.

In WormBase (WS211), *nrs*-2 (F25G6.6) is reported to encode a second cytoplasmic asparaginyl tRNA synthetase. A search using an amino acid sequence of this gene resulted in a match with asparagine synthetase not asparaginyl tRNA synthetase as

it is annotated. Asparagine synthetase synthesizes asparagine from aspartate (Kiriyama et al. 1989). A BlastP search showed that *Y66D12A.23* encodes the mitochondria asparaginyl tRNA synthetase.

In *C. elegans,* three genes encode phenylalanyl tRNA synthetase: *frs-1* (*T08B2.9*), *frs-2* (*F22B5.9*), and *frs-3* (*Y60A3A.13*). *frs-1* (now named *fars-1*) and *frs-2* (now named *fars-3*) both encode cytoplasmic enzymes while *frs-3* (now named *fars-2)* encodes a mitochondrial enzyme. The best *S. cerevisiae* BlastP match for *frs-1* is FRS2 (YFL022C) that encodes FARS alpha subunit. Moreover, the *S. cerevisiae* BlastP match for *frs-2* (now named as *fars-3*) is FRS1 (YLR060W) that encodes the beta subunit of the *FARS* enzyme. Although the majority of aminoacyl-tRNA synthetases are formed from monomers or dimers made up of identical subunits, FheRS is a tetramer with a $\alpha_2\beta_2$ structure. With a few exceptions, this quaternary structure is exhibited in both prokaryotes and eukaryotes. This tetramer structure occurs in some prokaryotes such as *E. coli* and has been found in some eukaryotes like yeast and mammals (Sanni *et al.* 1988). Therefore, *frs-1* (now named *fars-1*) and *frs-2* (now named *fars-3*) do not encode two separate FheRS enzymes but these two genes encode the two subunits: α (*frs-1*, now named *fars-1*) and β (*frs-2*, now names *fars-3*) that forms the FRAS tetramer.

In *C. elegans,* most of AARS enzymes have two genes; one encodes the cytoplasmic enzyme while the other encodes the mitochondrial enzymes. GARS, KARS, HARS and ThrRS have one gene that encodes for both the cytoplasmic as well as the mitochondrial enzymes. The GRS, HRS, and TRS loci encode two different transcripts producing two different AARS enzymes, one of which functions in the cytoplasm while the other functions in the mitochondria. This phenomenon of forming the cytoplasmic

and mitochondrial proteins from the same loci was previously observed for *mel-32* which encodes serine hydroxymethyltransferase, a gene that was previously identified in Dr. Baillie's laboratory (Vatcher *et al.* 1988). *Krs-1* is a more complex case in which the two proteins (cytoplasmic and mitochondrial) are derived from mRNA alternative splicing. For *grs-1*(*gars-1)*, *hrs-1* (*hars-1)*, *trs-1*(*tars-1)* and *mel-32* gene models please refer to appendix D.

To confirm WolF PSORT and BlastP predictions, I used an experimental method (reporter translational fusions) to determine the localization of two *C.elegans* AARSs: *srs-1* (*Wo3B1.4,* now named *sars-2*) and *srs-2* (*C47E12.1,* now named *sars-1*). WoLF PSORT predicted nuclear localization for *srs-1* (*Wo3B1.4*), however a BlastP prediction showed that this gene encodes a mitochondrial enzyme. For *srs-2* (*C47E12.1*), both WoLF PSORT and BlastP predicted that it encodes a cytoplasmic protein. GFP expression of the *srs-1* (*Wo3B1.4,* now named *sars-2*) construct was localized to the mitochondria. The expression of the *srs-2* (*C47E12.1,* now named *sars-1*) construct was in the cytoplasm. These results confirm the BlastP computational based predictions.

There are expression differences between the translational fusion constructs of *mars-1* (*F58B3.5*), *srs-1* (*Wo3B1.4*) and *srs-2* (*C47E12.1*) (Fig. 8). Using an exposure time of 1s, the expressions of *srs-1* (*Wo3B1.4,* now named *sars-2*) and *srs-2* (*C47E12.1,* now named *sars-1*) are both stronger than *let-65* (*F58B3.5*) expression. Expression variability could reflect differences in the enzyme's specific activities. MARS activity is lower than that for SARS Gabius *et al.* (1983). This difference in the reported enzymes' activities could have resulted from differences at the transcriptional level. Another

possibility for the observed lower GFP expression in *mars-1* (*F58B3.5*) construct is to be because of copy number variations in the GFP extra-chromosomal array.

Since that *C. elegans* nomenclature now allows four letter gene names, with consultation with Dr. Jonathan Hodgkin (University of Oxford; Genetics Unit, Department of Biochemistry, UK), (who is in charge of *C. elegans* nomenclature and he accepted my recommendation), we rename all the worm AARS genes from *xrs-* to *xars-*, to in order bring them in line with most of their homolog's in other organisms. Moreover, to standardize the AARS genes nomenclature, I recommended numbering the genes as "*xars-1*" for genes encoding cytoplasmic proteins and "*xars-2*" for genes encoding mitochondrial proteins. In cases where there is a second cytoplasmic enzyme *xars-3* is to be used. The recommendations were accepted and the nomenclature for the *C. elegans* AARS genes has been modified in WormBase, WS213 release, updated 30-Mar-2010 (Table 3).

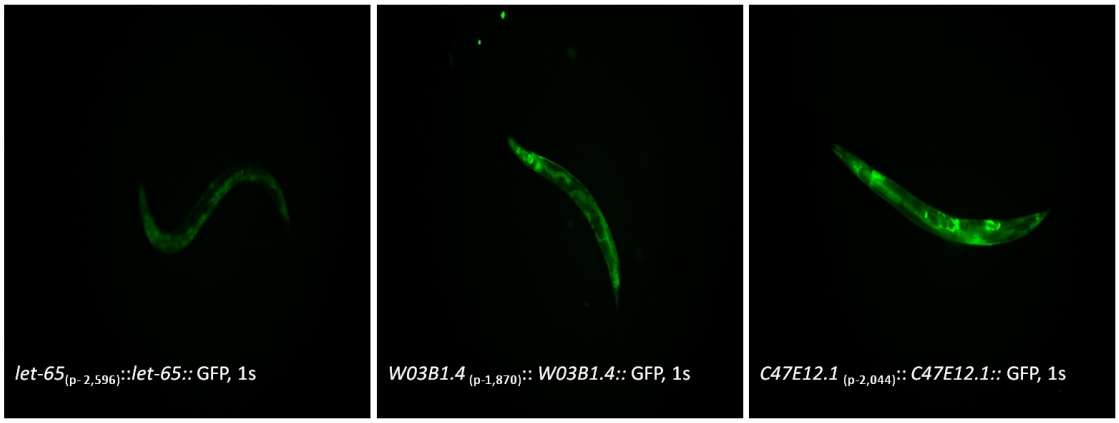


Figure 8. **Expression differences among translational fusion constructs. Using a 1s exposure time, the observed expressions of *srs-1* (*Wo3B1.4,* now named *sars-2*) and *srs-2* (*C47E12.1,* now named *sars-1*) are stronger than the observed expression of *let-65* (*mars-1*).**

CHAPTER 3: MAPPING OF A *5' CIS*-ACTING REGULATORY MOTIF IN *MARS-1* PROMOTER REGION

3.1 Introduction

Non-coding DNA upstream genes may contain control sequences, which are generically defined as *cis*-regulatory elements. These elements function at the transcriptional level by the binding of general transcription factors that control the assembly of a pre-initiation complex (PIC). PIC consists of RNA polymerase II (Pol II), general transcription factors (GTFs) and a mediator (Blackwell and Walker 2006). These elements control genes expression by alternating the association strength between Pol II and the TATA Box which located upstream of the ATG start codon (Blackwell and Walker 2006).

The study of regulatory elements that control gene expression patterns leads to a deeper understanding of mechanisms for gene expression and provides important insights into the function of those genes. Studying these elements offers a glimpse into gene expression in different tissues and developmental stages. In *C. elegans*, *cis*-acting regulatory elements are generally located close to the coding region in which the minimal promoter containing region required for appropriate expression of most Pol II transcripts lies within a couple of kilobases upstream of the start codon (Blackwell and Walker 2006; Boulin *et al,* 2006).

Fluorescent reporters are common tools used for studying transcriptional regulation in *C. elegans*. Having the minimal promoter region usually locating within few

kilobases upstream of the gene makes is easy to amplify the region by PCR and fused by PCR to a reporter gene. To identify *cis*-regulatory elements for *mars-1*(*let-65*), I used GFP, which is one of the commonly used reporters in transcriptional fusions, to make promoter::GFP reporter constructs. In my study I observed both complete loss of GFP expression and changes in expression intensity.

3.2 Materials and Methods

3.2.1 Construction of *let-65*$_{promoter}$::GFP transgenes

PCR fusion was used to generate *let-65*$_{promoter}$::GFP transgenes (Hobert 2002). Sequences putatively containing promoter elements were amplified from N2 genomic DNA (Bristol, Baillie Laboratory strain, BC49) and then fused upstream of the GFP encoding sequence which had been amplified from the *pPD95.67* (GFP) coding cassette. *pPD95.67* consist of a 5' nuclear localization sequence (NLS) from SV40 virus, five artificially derived introns, and *C. elegans unc-54* 3' un-translated region (UTR) (Lee 2009). Amplification of the GFP reporter was carried out using the following primers: GFP-C: AGCTT GCATG CCTGC AGGTC GACT; and GFP-D*: GGAAA CAGTT ATGTT TGGTA TATTG GG (designed by Allan K. Mah). The reverse Primer used to amplify the putatively entire *let-65* promoter containing region was *F58B3.5*: AGTCG ACCTG CAGGC ATGCA AGCTG GCAAG GCGTC ACTGA ATA. This reverse primer used for the promoter region had additional sequence complementary to the forward GFP primer to allow joining of the two PCR products. In all PCR reactions, phusion polymerase (Finnzymes, New England Biolabs Cat: F530) was used to ensure the constructs fidelity. Forward primers used to amplify the *let-65* putative promoter contacting regions and their distances upstream of the ATG

start site are shown in Appendix E. *let-65* sequences used for primer design were obtained from WormBase (WS209).

3.2.2 Microinjection

let-65$_{promoter}$::GFP (1µl) constructs were co-injected with *pCeh361* at 100 ng/µl. All injections were done by Domena K. Tu. In cases where there was no observed GFP expression, PCRs were conducted to ensure that lack of fluorescence was not due to the lack of transgene uptake, but the loss of expression was due to inadequate 5' *cis*-regulation. The primers used here are GFP-specific primers designed by Carrie L. Simms: GFP-F: CCATG CCCGA AGGTT ATGTA, and GFP-R: AAAGG GCAGA TTGTG TGGAC. For a list of strains generated in this section, refer to Appendix F. Refer to section 1.2.7 Materials and Methods for a detailed description of microinjection.

3.2.3 Microscopy

Screening transgenic strains was done using a Zeiss Axioscope set with a QImaging camera. 1mM levamisole (in M9) (Sigma, L9756-5G) was added immediately prior to imaging to immobilize worms on a moist agarose (2%) pads. All images were captured using the same equipments including: GFP filter, lens, camera, and QCapture software. Images were processed using Adobe Photoshop CS. Exposure times are indicated in the figure legends.

3.2.4 Phylogenetic analysis

The -683bp to +3bp sequence for *C. elegans MARS-1* and -850bp to +3bp, the orthologous genes in both *C. briggsae* (*CBG06108*) and *C. remanei* (*CRE12616*) were obtained from WormBase (WS213). BioEdit, set to its default parameters, was used to

make Pair-wise alignments which were visualized using GeneDoc (http:://www.psc.edu./biomed/genedoc). Triple alignment of the non-coding region upstream of *mars-1* and its orthologs from *C. briggsae* and *C. remanei* was done using the FamilyRelations II program (Brown *et al.* 2005, http://family.caltech.edu/), using the same regions as above. Sequences used are presented in Appendix G.

3.3 Results

3.3.1 Sequential *let-65*$_{promoter}$:: GFP deletion constructs

Generated constructs containing putative promoter elements used to examine GFP expression were from positions -2521bp, -2053bp, -1393bp, -684bp, and -603 to position +3bp relative to the *mars-1* ATG codon. For -2521bp, -2053bp, -1393bp, and -684bp fragments, the expression was variable (data not shown). For -603bp fragment, GFP expression was seen in the gut, anterior neurons, pharynx procorpus, pharyngeal bulbs and muscles. Consequently, the region from -603bp to the translation start site (ATG) was chosen for finer deletion analysis in order to target reduction or complete loss of GFP expression.

The new fragments were from positions -477bp, -344bp, -283bp, -124bp and -58bp to the translation start site (ATG). Fragment -477bp had the same GFP expression intensity as fragment -603bp for all observed cells. For the -344bp fragment, lower GFP expression was seen in the gut, and pharyngeal bulbs. Reduction in the GFP expression was observed in -283bp and -124bp constructs in the anterior neurons, pharynx procorpus and isthmus, and pharyngeal bulbs compared to that observed in -603bp, -477bp, and -344bp constructs. Brighter GFP expression was observed in the gut for the -124bp

fragment compared to that observed in the -283bp construct. However, less expression was seen in the anterior neurons, pharynx procorpus and isthmus, and pharyngeal bulbs. Mostly no GFP expression was detected for the 58bp fragment; however, faint GFP expression was detected in the pharynx procorpus and isthmus, and pharyngeal bulbs of a few nematodes (Table 4). Examples of the transgenic worms' GFP gut expression patterns for the shortest six constructs are shown in Figure 9.

Table 4. GFP expression patterns in strains containing truncated *F58B3.5*$_{promoter}$::GFP constructs.

Tissue	-603	-477	-344	-283	-124	-58
Gut	+++++	+++++	++++	++++	+++++	-
Anterior neuron	+++	+++	+++	++	+	-
Pharynx procorpus and isthmus	+++++	+++++	+++++	++	++	+/-
Pharyngeal bulbs	++++	++++	+	+	+	+/-
Muscles	+++	+++	++++	+++	+++	-

Negative numbers in the heading row indicates the start position of the used promoter fragment relative to the translation start site (ATG) in base pairs. In the -58bp construct the +/- indicates a very low level of GFP expression.

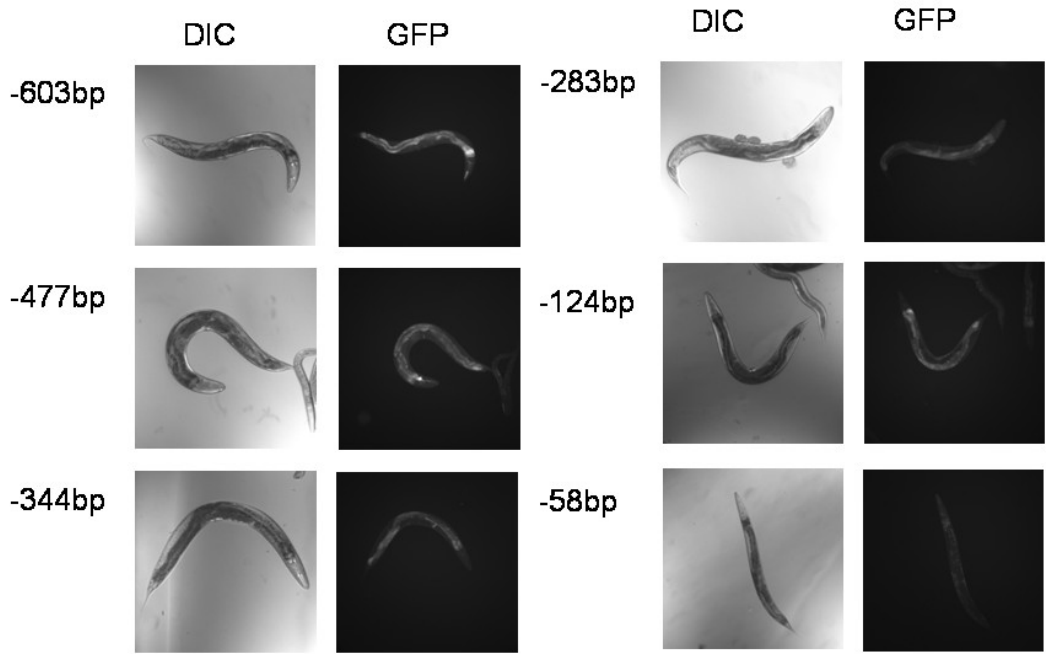


Figure 9. **Gut GFP expression levels for the shortest six *mars-1*$_{promoter}$::GFP constructs. Gut GFP expression levels varied among the strains containing different promoter::GFP constructs. To ensure comparable levels of transgenic GFP expression, all images were taken of young adults worms at 1s exposures. The sizes of the shown constructs are indicated in the left of each construct image.**

3.3.2 Identification of evolutionarily conserved non-coding elements upstream of *mars-1*

Identification of evolutionary conserved non-coding elements upstream of *mars-1* in *C. elegans* was conducted by comparing the upstream region in *C. elegans mars-1* to the upstream regions of the *mars-1* orthologous in two other *Caenorhabditis* species: *C. briggsae* and *C. remanei.* For *C. elegans* I chosen the region -683bp upstream the translations start site (ATG). Since *C. briggsae* and *C. remanei* genomic sequence, in the publicly available data base, are missing sequences in the upstream region (-629 to -319 and -347 to -337 respectively), larger regions were used (-850bp for each) (All sequences were obtained from WormBase release WS213).

C. elegans, C. briggsae, and *C. remanei* regions were analyzed using the FamilyRelations II program (Brown *et al.* 2005) that revealed the existence of conserved inverted sequences (Fig. 10A). Highly conserved areas were revealed by increasing the stringency of the comparison (Fig. 10B). The alignment shows two conserved sites besides the splice junction site (*C. elegans* sequence: TATTCAGT) (fig 10 C, D, E and F). These conserved sequence blocks are: AAGTGtGCtCtAcgGAGAATT and TTTTTGATTGACAgTT (lower case letters are mismatches between *C. elegans* and *C. briggsae*). The second conserved sequence is in all three species but in *C. remanei* it is in the reverse complement orientation.

BioEdit was used to locate the conserved sequences Pair-wise sequence alignment was conducted using *C. elegans* and *C. briggsae* sequences with default parameters. The alignment was visualized using GeneDoc program (http:://www.psc.edu./biomed/ genedoc) (Fig. 11).

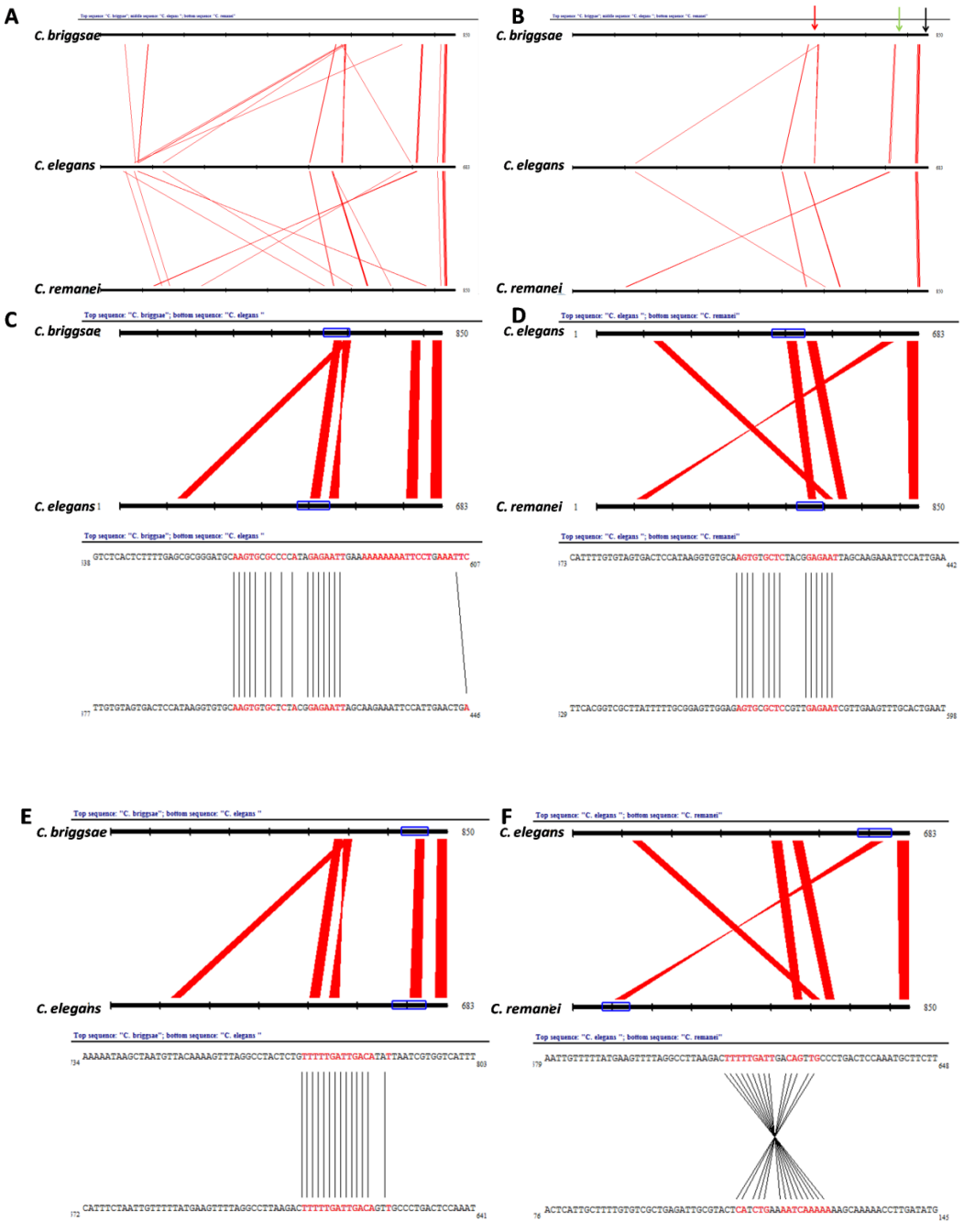


Figure 10. **Interspecies comparison of *C. elegans*, *C. briggsae*, and *C. remanei* upstream regions. A and B show comparison of the upstream regions relative to the splice junction sequence (Black arrow in B) of *C. elegans, C. briggsae* and *C. remanei* was conducted using FamilyRelations II (Brown *et al.* 2005). B shows the almost completely conserved regions. C and D show identified conserved sequence (Red arrow in B), E and F, also show conserved sequence (green arrow in B).**

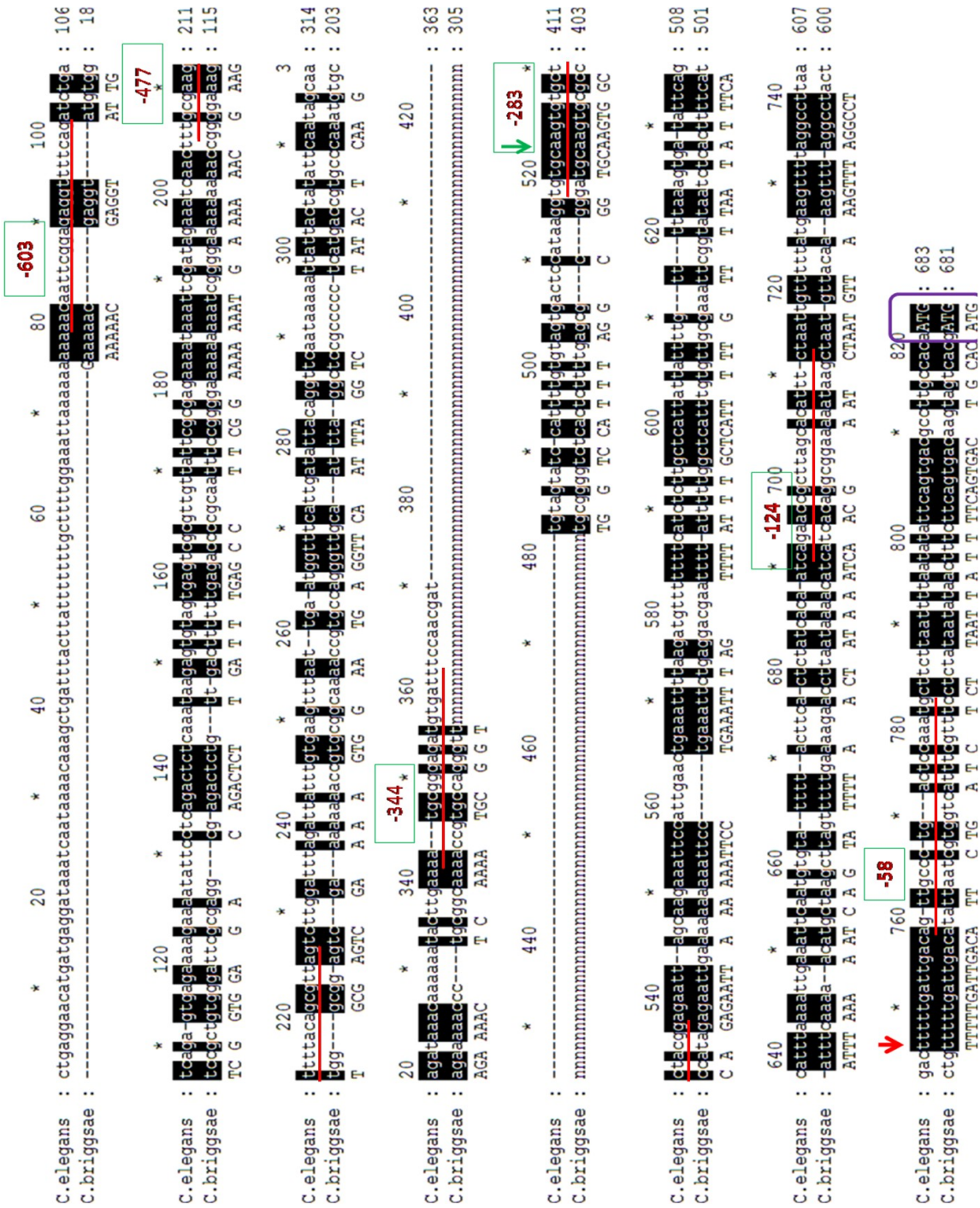


Figure 11. **Sequence alignment of *C. elegans* and *C. briggsae mars-1* gene upstream region. Red lines indicate the primers sequences of the shortest six constructs. Boxed numbers indicate the length of the fused regions using the indicated primers relative to the translation start site (purple box). Green and Red arrows locate the location of the first base of the two conserved regions found using FamilyRelations II. Alignment was done using BioEdit and visualized using GeneDoc.**

3.4 Discussion

In general transcription factors (TFs) bind to specific sequences upstream of a gene's coding sequence. TFs function as expression activators or repressors by alternating the binding strength of RNA polymerase II to its binding site. In this part of my thesis, I identify possible *cis*-regulatory motifs that drive expression of *mars-1*. I have used PCR to generate a range of *let-65*$_{\text{promoter}}$:: *GFP* constructs, then made transgenics with them and analyzed the resulting expression patterns.

Based on methionyl tRNA synthetase role in translation, *mars-1* is expected to be expressed in all tissues. However, for my transgenics I observed GFP florescence in the gut, anterior neurons, pharynx procorpus and isthmus, pharyngeal bulbs and muscles. Expression in other tissues might have been so low that it is not visible.

The expression patterns of promoter fusion for three other tRNA synthetase genes (*kars-1*, *pars-1*, and *tars-1*) are published in WormAtlas (Hunt-Newbury *et al.* 2007, http://gfpweb.aecom.yu.edu/index). *kars-1* expression was observed in the pharynx, intestinal, rectal epithelium, hypodermis, seam cells, and head neurons. For *pars-1*, GFP expression was observed in pharynx, intestinal, and distal tip cell, body wall muscle. For *tars-1,* GFP expressed in pharynx, pharyngeal gland cells, intestinal, anal depressor muscle, and body wall muscles (Hunt-Newbury *et al.* 2007).These expression variations of the investigated AARS genes might be due to differences in genes expression levels.

To identify *cis*-regulatory motifs, promoter constructs were designed to assay GFP expression from positions -2,521bp, -2,053bp, -1,393bp, -684bp and -603 to the *mars-1* translation start site. Variable expression was observed for the -2,521bp, -2,053bp, -1,393bp, and -684bp fragments. For the -603bp fragment, GFP expression was

seen in the gut, anterior neurons, pharyngeal bulbs and muscles. Therefore, the region between -603bp was chosen for finer deletion analysis in order to target reduction or complete loss of GFP expression.

The sequence fragments used to generate the new fusion constructs were from positions -477bp, -344bp, -283bp, -124bp and -58bp to the translation start site. These constructs exhibited mosaic expression that was likely a consequence of somatic loss of the injected extra-chromosomal array.

As shown in table four, in all tissues except the gut, an overall reduction in expression intensity was observed for the -283bp fragment. A very low level pharynx expression was observed in few individuals for the -58bp construct. The *unc-54* 3' UTR used in the *pPD95.67* expression cassette might be causing this weak expression (Boulin *et al.* 2006).

The analysis for predicted transcription factors sites was done by a triple alignment of *C. elegans, C. briggsae* and *C. remanei* using FamilyRelations II to detect conserved sequences that might not be found using BioEdit such as sequences are comprised of inverted sequences and sequences that are not in a matching order relative to other sequences. Two conserved sequence blocks were identified in the three orthologs: AAGTGtGCtCtAcgGAGAATT and TTTTTGATTGACAgTT (lower.case letters are mismatches between *C. elegans* and *C. briggsae*). The second conserved block is in the reverse complementation in *C. elegans* relative to *C. elegans* and *C. briggsae*.

I located these two conserved sequence blocks, using pair-wise sequence alignment of *C. elegans* and *C. briggsae*. The first sequence is located -278bp from the ATG, which places it in -283bp construct which might indicate the presence of a binding

sit for a TF in this region. The second sequence is located at -70 bp from the ATG, which places it in the region before the -58bp construct. The removal of this sequence may well be responsible for the loss of GFP expression for the -58 construct.

By performing a promoter::GFP "cut down" analysis, I found a 66bp window containing a putative *cis* regulatory element that controls the wild type *mars-1* expression. This analysis led to the identification of two conserved DNA sequences for the *mars-1* orthologs regions in three nematode species: *C. elegans, C. briggsae* and *C. remanei.* For future work, additional cut-downs should be conducted every 5bp in the region containing the putative *cis* regulatory elements. Moreover, the identification of possible TFs bound to the found sequences can be experimentally detected using gel-mobility shift assays. For the detected sequence locating in -283bp to +3bp construct, further analysis is required to determine its importance by making constructs using fragments including more bases of the upstream region of this sequence followed by conducting cut-downs every 5bp in the region.

CHAPTER 4: INVESTIGATION OF HAPLO-INSUFFICIENCY PHENOTYPES ASSOCIATED WITH MUTATIONS IN GENES INVOLVED IN TRANSLATION IN *C. ELEGANS*

4.1 Introduction

One of the most extensively studied types of mutants in *D. melanogaster* involves *Minutes* mutants. There are several dominant phenotypes exhibited by *Minute Drosophila* flies including large, rough eyes, wing abnormalities, abdominal segmentation defects, reduced fertility and viability, and small body size (Lambertsson 1998). However, short slender bristles, delayed development, and recessive lethality are the three main common visible phenotypes displayed by individuals with these mutants (Schultz 1929; Lambertsson 1998). *Minutes* have a non-additive phenotypic effect; that is, there is no difference in the severity of phenotypes between flies with a single *Minute* mutation and flies with more than one (Lambertsson 1998).

Minute loci are haplo-insufficient (dominant) and *Minute* loci are dispersed throughout *D. melanogaster's* genome that suggested that *Minute* loci encode different products with related functions (Lambertsson 1998). *Minute* loci were found to correspond to Ribosomal Protein genes (RP) in which reducing the production of a single RP causes the reduction in the number of functional ribosomes that result in impairing protein synthesis (Marygold *et al.* 2007). It has been suggested that some non-ribosomal genes encoding proteins involved in translation (such as aminoacyl tRNA synthetases) may also lead to *Minute* phenotype (Lambertsson 1998).

In this chapter, I discuss my examination of haplo-insufficiency phenotypes (*Minute* like phenotypes) in *C. elegans* associated with mutations in genes involved in translation. In my investigation I used mutant alleles of two genes involved in translation. The first gene is *let-65* which encodes a methionyl- tRNA synthetase (previously discussed in chapter 1). For my investigation, I used two alleles of *let-65*(*s1777* and *s1222).* The mutation found in *s1777* truncates the protein at the 408th amino acid while for *s1222* the mutation is in the first base pair of the first intron and so prevents the splicing of the first intron and causes a premature truncation of the protein at amino acid number 28.

The second gene I used for my investigation is *let-336* which has EMS induced mutations (Johnsen and Baillie 1991). *let-336* encodes *rps-27* which is a small (40s) ribosomal subunit *S27* (The molecular identity was identified by Fernandes, A. R., Simon Fraser University, un-published data). I used two of the three of *let-336* alleles: *s1413* and *s1495* that were most kindly provided to me by Fernandes, A. R., (Baillie lab, Simon Fraser University). Both *let-336(s1413* and *s1495)* prematurely truncate the protein (at 508bp and 515bp respectively) (Fernandes, A. R., Simon Fraser University, un-published data). *let-336(s1413)* changes glutamine to a stop codon at 65aa out of the total 83aa (Fernandes, A. R., Simon Fraser University, un-published data). Similarly, *let-336(s1495)* changes arginine to a stop codon at 68aa out of the total 83aa (Fernandes, A. R., Simon Fraser University, un-published data).

The putative haplo-insufficiency phenotypes that I examined are recessive lethality, small brood size**,** developmental delay, and extended lifespan. Since that the genes I used belongs to two different protein groups (ribosomal proteins and amino-acyl

tRNA synthetases), it was expected that mutations in these genes might not exhibit the same haplo-insufficiency phenotypes.

4.2 Materials and Methods

4.2.1 Maintaining nematode strains and culture conditions

Nematode strains were maintained on Easiest Worm Plate Agar streaked with *E. coli (OP50)* (Brenner 1974). All strains were stored at 15°C. The strains have the genotypes presented in table 5. The control strain BC1270 (see table 5) was used for *let-65* allele's analysis and BC2200 was used as the control strain for *let-336* allele's analysis.

4.2.2 Brood size assay

L4 worms for each tested strain were picked, one per plate, to 20 plates and then incubated at 20°C for 24-hr time periods. After each period, the picked worms were transferred to new plates. F_1 progeny from the P_{0s} were allowed to grow for three days before they were counted. This brooding was continued every 24 hr until the P_{0s} cease to produce progeny. Statistical significant differences between controls and tested animals were determined by using a two-tailed T-test with unequal variance (EXCEL, 2007). Progeny number count for BC2200 was done by Fernandes, A.R., Simon Fraser University, un-published data. *let-336(s1413)* progeny counts were done with collaboration with Fernandes, A. R.. Statistical significance of differences between controls and tested animals was determined by a two-tailed T-test with unequal variance (EXCEL, 2007).

4.2.3 Developmental delay assay and microscopy

Twenty adults from each strain were transferred to *OP50* streaked plates and allowed to lay eggs for four hrs before been removed leaving only unhatched eggs. Hatched animals were cultured for 24, 48 and 72 hrs before being transferred to microscope slides for screening.

Screening of worms was done using a Zeiss Axioscope equipped with a QImaging camera. To immobilize the worms on the 2% moist agarose, 1mM levamisole (in M9) (Sigma, L9756-5G) was applied to the worms immediately prior to imaging. Pictures were taking at each time point with a 10× objective. The body length (head-to-tail) of 20 animals per time point was measured with Image J software (NIH, Bethesda, MD). Statistical significance of differences between controls and tested animals was determined by a two-tailed T-test with unequal variance (EXCEL, 2007).

4.2.4 Lifespan assay

L4 worms from each strain were picked (one each) to 10 plates and incubated at 20 ˚C. Worms were picked to new plates every 24 hrs until P_0s ceased to produce progeny and were scored daily for live/dead worms. An animal was considered dead if it did not respond to prodding with a pick and its pharynx did not pump. The scoring was repeated until all animals were dead. Statistical significance of differences between controls and tested animals was determined with a two-tailed T-test with unequal variance (EXCEL, 2007).

Table 5. Genotypes of strains used in *C. elegans* haplo-insufficiency phenotype investigations.

Strain	Allele	Genotype
BC1270	-	*unc-22(s7) unc-31(e169)/nT1*(IV); +/*nT1* (V)
let-65	*s1222*	*let-65(s1222) unc-22(s7) unc-31(e169)/nT1*(IV); +/*nT1* (V)
let-65	*s1777*	*let-65(s1777) unc-22(s7) unc-31(e169)/nT1*(IV); +/*nT1* (V)
BC2200	-	*dpy-18(e364)/eT1(III); unc-46(e177)/eT1(V)*
let-336	*s1413*	*dpy-18(e364)/eT1(III);let-336(s1413)unc-46(e177)/eT1(V)*
let-336	*s1495*	*dpy-18(e364)/eT1(III);let-336(s1495)unc-46(e177)/eT1(V)*

4.3 Results

4.3.1 Self-progeny brood size

The strains used in this study were *let-65*(*s1222* and *s1777)* and *let-336(s1413* and *s1495)* which were all induced by EMS mutagenesis and all alleles result in prematurely truncated proteins. For *let-65* alleles and its control (BC1270), F_1 individuals with wild type phenotype (25% of the total selfing progeny) were used for the comparison. For *let-366* and its control (BC2200), F_1 individuals with wild type and unc-36 phenotype (about 31% of the total selfing progeny) were used for the comparison. For the punnett squares for *let-56* and *let-366* (the selfing and the control strains) please refer to Appendix H.

Comparison of self-progeny brood sizes with respect to the wild type phenotype of *let-65(s1222)* and the control BC1270, using T-test un-equal variance, showed a significant difference ($P < 0.05$). The brood size mean number (± one standard deviation)

for *let-65(s1222)* was 74± 10 but 88± 8 for the control BC1270. Similarly *let-65(s1777)* showed a mean self-progeny brood size significantly smaller than the control by T-test ($P < 0.05$). The mean self-progeny brood size with respect to the wild type phenotype (± one standard deviation) for *let-65(s1777)* was 54± 17. Thus the self-progeny brood sizes of the mutant strains *let-65(s1222* and *s1777)* are significantly less than that observed of the control (Fig.12).

let-336(s1413 and *s1495)* hermaphrodites showed means self-progeny brood sizes with respect to wild type and unc-36 phenotypes, that are not significantly different ($P > 0.05$) from the control BC2200 (using T-test, un-equal variance). The mean number of the counted progeny (± one standard deviation) for *let-336(s1413)* was 74± 12 and for *let-336(s1495)* was 66± 18. The mean of the self-progeny brood size (one ± standard deviation) for the control was 76± 15. Thus the self-progeny brood size in the mutant strains *let-366(s1413* and *s1495)* is not significantly different from the control's (Fig.13).

4.3.2 Developmental delay

To investigate for developmental delay in the test strains, the body length, from head-to-tail of 20 heterozygous animals of each strain were measured at time points 24, 48, and 72 hrs. The body lengths of *let-65*(*s1222)* and *let-65(s1777)* heterozygous individuals and the control strain BC1270 are not significantly different ($P > 0.05$) using a T-test, un-equal variance at 24 and 48 hr time point (Fig. 14 and 16).

Body length mean (± one standard deviation) for *let-65*(*s1222)* and *let-65(s1777)* at 24 hrs were 0.32± 0.03 mm and 0.30± 0.03 mm respectively. The control BC1270 mean (± standard deviation) was 0.32± 0.03 mm (Fig. 14 and 16).

At 48 hrs, the mean body length (± one standard deviation) was 0.54±0.03 mm and 0.54± 0.04 mm for *let-65(s1222* and *s1777)* respectively. For the control BC1270 at the same time point it was 0.55± 0.04 mm (Fig. 14 and 16).

At 72 hrs, no significant difference was observed for *let-65(s1222)* ($P > 0.05$) compared to the control. Body length means (± one standard deviations) were 0.93± 0.05 mm for *let-65(s1222)* and 0.94± 0.05 mm for BC1270. However, for *let-65(s1777)* body length means (± one standard deviations) was 0.85± 0.06 mm that is significantly different ($P < 0.05$) from the control BC1750 (Fig. 14 and 16).

Comparing body length's of *let-336*(*s1413)* and *let-336(s1495)* heterozygous individuals to the control BC2200 using T-test, un-equal variance at 24 hrs showed no significant difference ($P > 0.05$) with mean body length (± one standard deviation) of 0.27± 0.02 mm and 0.29± 0.02 mm for *s1413* and *s1495* respectively and 0.28± 0.03 mm for BC2200 (Fig. 15 and 16).

At 48 and 72 hr time points, both *let-336* alleles showed significant difference in body length compared to that of the control BC2200 ($P < 0.05$) (Fig. 15 and 16). At 84 hrs, body length means (± one standard deviation) for *s1413* and *s1495* were 0.45± 0.05 mm and 0.40± 0.03 mm respectively. For the control it was 0.53± 0.05.

At 72 hrs the mean body length (± one standard deviation) for *s1413* and *s1495* were 0.82± 0.08 mm and 0.58± 0.03 mm respectively. For the control strain it was 0.99± 0.05 mm.

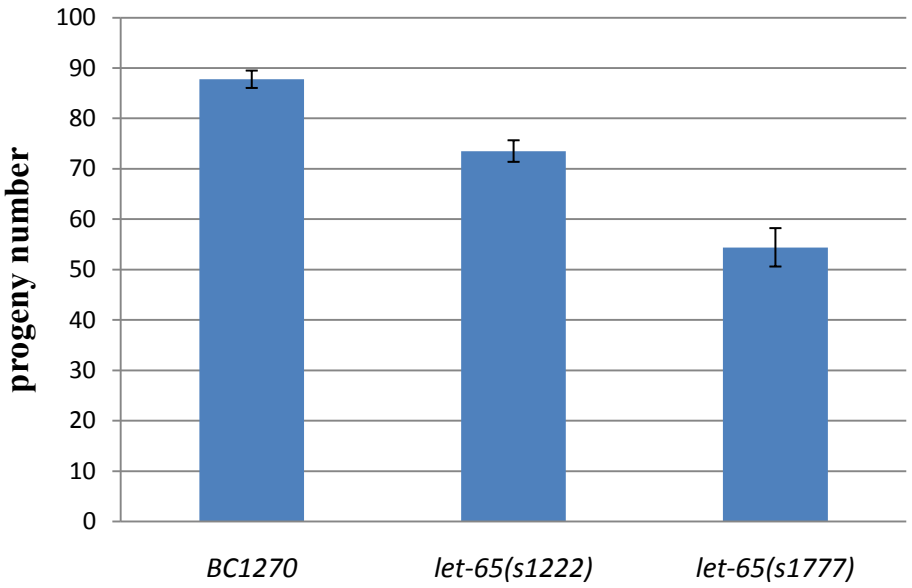


Figure 12. **Average number of wild type progeny for *let-65*(*s1222* and *s1777*) and the control BC1270.**
Mean numbers of progeny for *let-65(s1222* and *s1777)* are significantly different from the control (BC1270) ($P < 0.05$). The error bars represent the standard errors with 95% confidence.

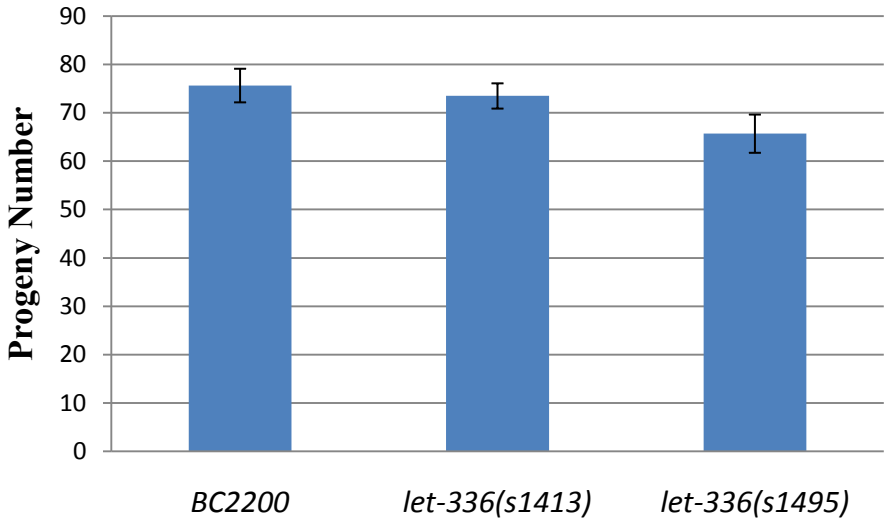


Figure 13. **Average progeny number for *let-336*(*s1413* and *s1495*) and the control BC2200. Mean numbers of progeny for *let-336* alleles showed no significant difference from the control (BC2200) (P > 0.05). The error bars represent the standard errors with 95% confidence.**

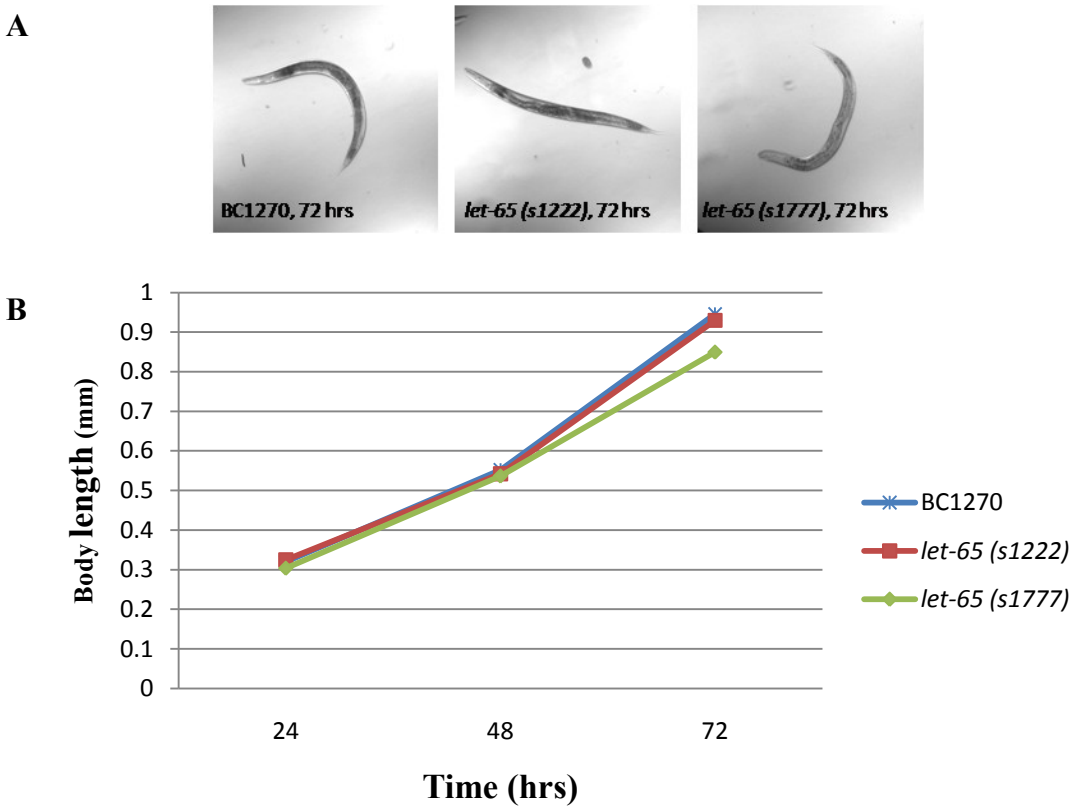


Figure 14. **Body length of *let-65*(*s1222* and *s1777)* at different time points. A. DIC images (10×) of heterozygous *let-65 (s1222* and *s1777)* and BC1270 grown for 72 hrs. B. Average length (mm) of 20 worms from each of *let-65*(*s1222* and *s1777*) and the control strain (BC1270). *let-65* alleles showed no significant difference in length compared to the control strain (BC1270) ($P > 0.05$) at 24 and 48 hrs time point. At 72 hrs, *let-65*(*s1777*) showed significant difference ($P < 0.05$) in length compared to the control.**

A

B

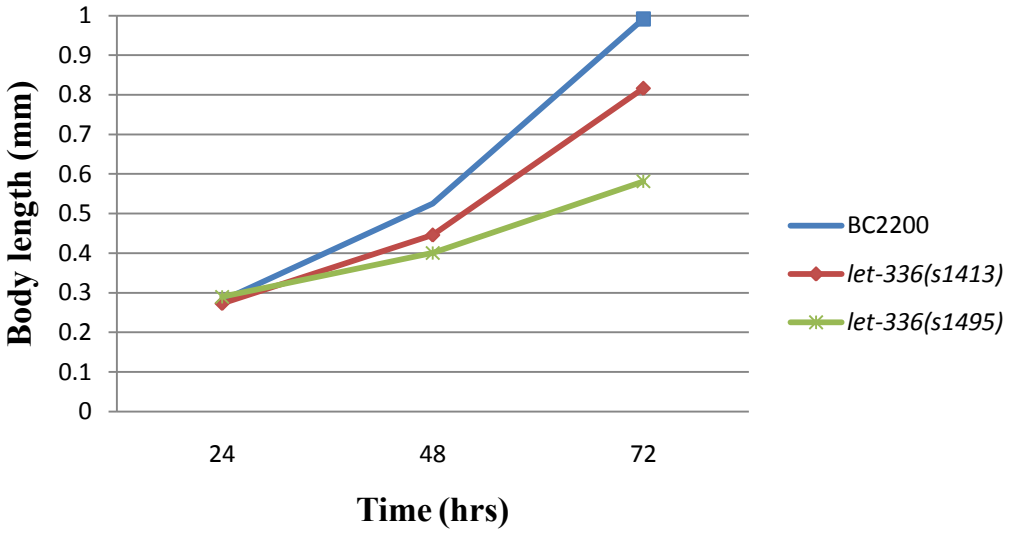


Figure 15. **Body length of *let-336*(*s1413* and *s1495)* at different time points. A. DIC images (10×) of heterozygous *let-336(s1413* and *s1495)* and BC2200 grown for 72 hrs. B. Average length (mm) of 20 worms from each strain of *let-336*(*s1413* and *s1495*) and the used control strain (BC2200). *let-336* alleles showed no significant difference in length compared to the control strain ($P > 0.05$) at 24 hrs. At 48 and 72 hrs, *let-336* alleles showed significant difference ($P < 0.05$) in length compared to the control strain.**

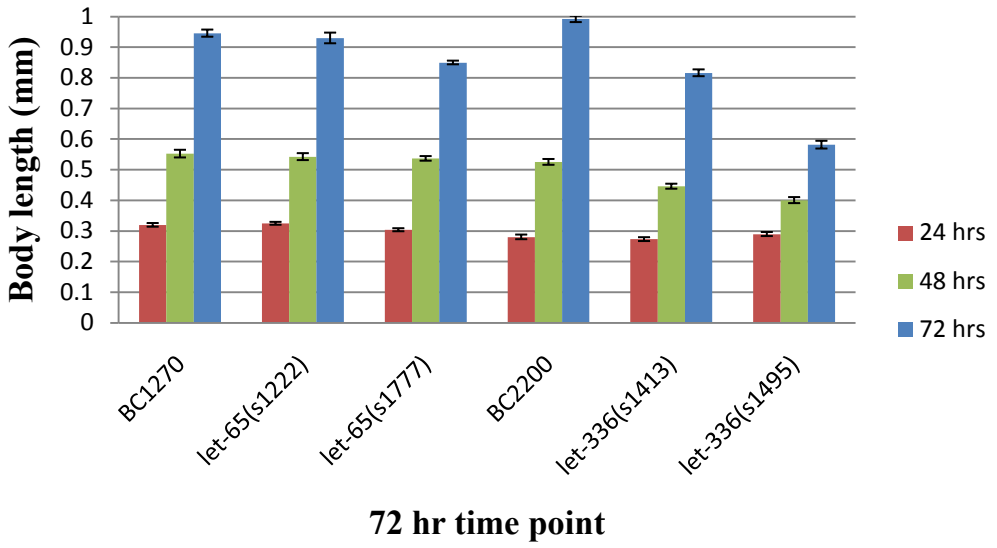


Figure 16. Average body lengths of *let-65(s1222* and *s1777)* and *let-336(s1413* and *s1495)*. **Body length of 20 worms of *let-65(s1222* and *s1777)* and the control strain BC1270 for *let-336(s1413* and *s1495)* and its control strain BC2200) at 24, 48, and 72 hrs. Lines shown are standard error with 95 % Confidence interval.**

4.3.3 Lifespan assay

Lifespan was investigated using ten worms. L4 worms from *let-65(s1222), let-65(s1777)* and their control BC1270 and *let-336(s1413), let-336(s1495)* and their control (BC2200), were picked one each to ten plates and incubated at 20 °C. Worms were brooded every 24 hrs and live individuals were scored until all animals were dead.

let-65(*s1222)* and *let-65(s1777)* heterozygous individuals did not differ significantly from the control (BC1270) using T-test, un-equal variance (Fig. 17 and 18). The mean numbers of days of life at adulthood (± one standard deviation) for *let-65(s1222)* and for *let-65*(*s1777*) were 11± 2.2 days and 11± 1.7 days respectively. For the control BC1270 it was found to be 13± 3.3 days.

Life span of *let-336*(*s1413)* heterozygous individuals was not significantly different from the control BC2200 using T-test, un-equal variance ($P > 0.05$). The mean number of adulthood days alive (± one standard deviation) were 11.7± 2.9 days for BC2200 and 10.7± 2.7 days for *s1413* (Fig. 17 and 18). *let-336(s1495)* is significantly different ($P < 0.05$) from the control with mean value (± one standard deviation) of 23± 2.1 days (Fig. 17 and 18).

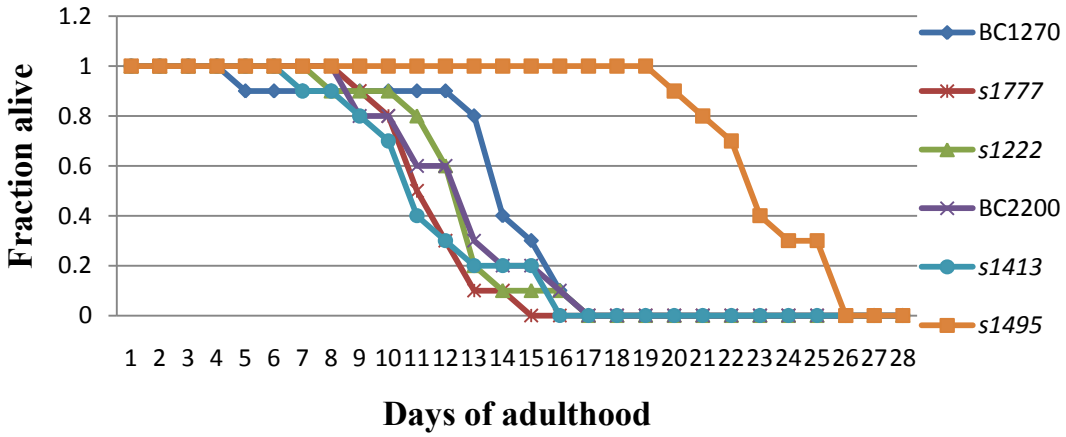


Figure 17. ***let-65(s1222* and *s1777)* and *let-336(s1413* and *s1495)* survival curve.**
Survival times for *let-65*(*s1222* and *s1777*) were not significantly different from the control BC1270 (P > 0.05). Moreover, the survival of *let-336(s1413)* was not significantly different from the control BC2200 (P > 0.05). *let-336(s1495)* live significantly longer than the control BC2200 (P < 0.05).

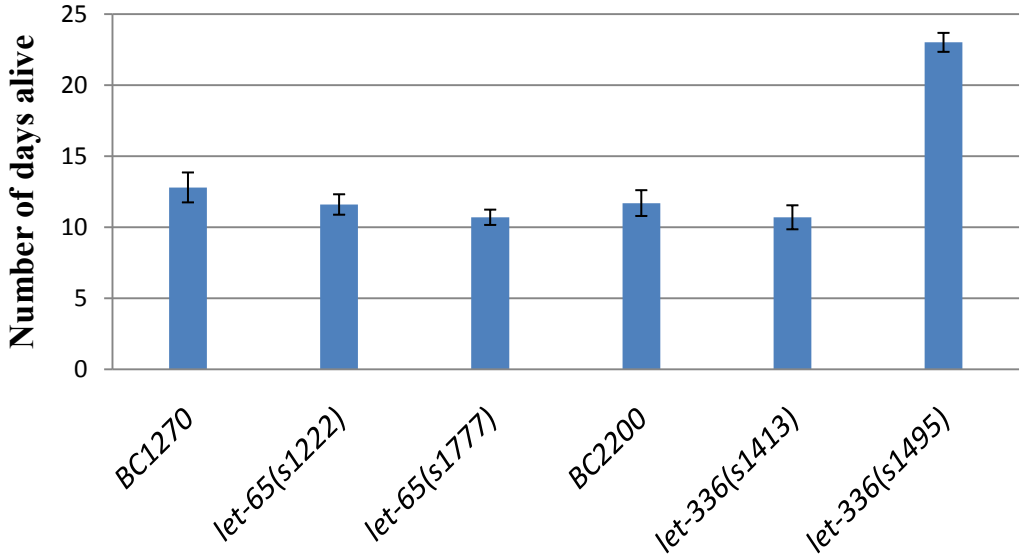


Figure 18. Average number of days survived for *let-65(s1222* and *s1777)* and *let-336(s1413* and *s1495)*.
Number of days of survival for 10 worms of *let-65(s1222* and *s1777)* and their control (BC1270) and for *let-336(s1413* and *s1495)* and their control (BC2200). *let-336(s1495)* showed a significant difference in number of survival days than the control BC2200 ($P < 0.05$). Lines shown are standard errors with 95 % Confidence interval.

4.4 Discussion

In *Drosophila,* mutations in genes encoding RPs cause haplo-insufficient phenotypes. The three main common visible phenotypes are short bristles, delayed development, and recessive lethality (Schultz 1929). Moreover, individuals with *Minute syndrome* also display a wide range of other phenotypes. Reducing the copy number of RPs causes a reduction in the number of functional ribosomes and these results in impairing protein synthesis (Marygold *et al.* 2007).

I investigated the *Minute* like phenotypes in *C. elegans*. In my investigation I used strains for two genes, *let-65(s1222* and *s1777)* and *let-336(s1413* and *s1495). let-336* encodes *rps-27,* a small (40s) ribosomal subunit *S27* protein (identified by Fernandes, A. R., Simon Fraser University, un-published data). Ribosomal protein components are important for ribosome assembly and therefore its translation functions (Brodersen and Nissen 2005). Several human syndromes and diseases result from ribosomal proteins misexpression such as some cancers, Turner syndrome (Ullrich-Turner syndrome), and Diamond-Blackfan anaemia. *In Drosophila*, *rps-27* is likely to be a *Minute* locus, in which a mutation or a deletion in heterozygous individuals displays a *Minute* phenotype (Marygold 2007).

For my investigation, I also used *let-65* that encodes *mars-1*(discussed in chapter 1). *mars-1* is an enzyme that plays a role in cytoplasmic translation through charging $tRNA^{Met}$ with methionine. As part of the *Drosophila Minute* investigations, it was suggested that mutations in non-ribosomal genes involved in translation might display

Minute phenotypes. Moreover, it was also suggested that aminoacyl-tRNA synthetases complete or partial inactivation may lead to a *Minute* phenotype or to a comparable traits (Lambertsson 1998).

For my investigation I used two controls BC1270 and BC2200. BC1270 has the same background genotype as *let-65*(*s1222* and *s1777);* however, it is just missing the EMS induced mutation. Similarly, BC2200 was used as a control for *let336*(*s1413* and *s1495*) since it has the same background genotype but just missing the EMS induced mutation. Heterozygous individuals, of the investigated strains, are visibly wild type and so do not display any uncoordinated movement. All tested allele's exhibit recessive lethality which is one of the three main dominant visible *Minute* phenotypes in *Drosophila*.

One of the *Minute* dominant traits in *Drosophila* is heterozygous reduction in female fertility. Oogenesis was found to be sensitive to ribosomal protein reduction (Lambertsson 1998), In which a small decrease in the ribosomal protein copy number, which causes reduction of complete ribosomes, is likely to affect oogenesis (Lambertsson, 1998).

In *C. elegans* hermaphrodites, the number of self-progeny brood size is limited by the number of sperm produced, which is followed by the gonad switch to oocytes exclusively (Johnsen 1990). In my investigation I tested to see if mutations in genes involved in translational have any effect on brood size. For *let-65* and *let-336* alleles, brood sizes were determined by counting the mature wild type self-progeny produced by hermaphrodites.

let-65(s1222) mean self-progeny brood size was about 16% less than the control (95% confidence limit by T-test). The wild type self-progeny brood size of *let-65(s1777)* was also significantly less than that for the control (approximately 39% less). The significant reduction in self-progeny brood size of *let-65(s1222* and *s1777)* might be a resulted from the protein's haplo-insufficiency.

let-336(s1413) have a mean self-progeny brood size that is about 2% less than the control which is not significantly different (95% confidence limit by T-test). The mean self-progeny brood size of *let-336(s1495)* is also not significantly different than that for the control (less by approximately 10%).

Although, no *let-336* alleles I used in my counts showed significant differences from the control, when Johnsen, R., (1990) compared the self-progeny brood size of non-balanced *let-336*(*s1413*) to hermaphrodites containing just *dpy-18/+; unc-46/+* as control he found a significant increase of about 15% by a T-test (95% confidence limit) (Johnsen, 1990). Johnsen did his counts in a different genetic background which allowed for a larger brood sizes because all the aneuploids were eliminated. Moreover, Johnsen out-crossed the strain from the balancer and thus cleaned up the genetic background and possibly eliminated other deleterious mutations. Other possibility for the difference is the media on which the strains were maintained. Johnsen used nematode growth media (NGM), whereas I used Easiest worm plate agar. These media are different in their composition in which NGM are made with 30% more NaCl, whereas, Easiest worm plate agar are made with approximately 20% more peptone and 40% more cholesterol (Lee 2009). To investigate the differences, *let-336*(*s1413*) self-progeny brood size count

should be repeated using the same genetic background Johnsen used and using NGM growth media.

In *Drosophila*, one of the three main characteristic of the *Minute* phenotype is delayed larval development (Schultz 1929). Therefore, I sought to evaluate the effect of protein synthesis disruption on development in *C. elegans.* I investigated this by measuring the body length of 20 animals of each tested strain (mutated alleles and controls) at three time points: 24, 48, and 72 hrs after being cultured on Easiest Worm Plate Agar streaked with *E.coli (OP50)*.

let-65(*s1222)* and *let-65(s1777)* heterozygous individuals show no significant difference in mean body length compared to the control BC1270 at 24 hrs and 48 hrs ($P > 0.05$). At 72 hrs, *let-65(s1222)* shows no significant difference in mean body length relative to BC1750 ($P > 0.05$). However, *let-65(s1777)* has about 9% shorter mean body lengths which is significantly shorter than the control ($P < 0.05$). At 72 hrs, eggs were visible in the control's hermaphrodite gonad; however, for *let-65(s1777)* one or two eggs were visible in the gonad indicating that the worms were young early adults at this time point.

The ribosomal gene *let-336* showed a visible developmental delay. At 24 hrs both alleles *let-336(s1413* and *s1495)* show no significant length differences from the control BC2200. At 48 hrs time point, *let-336(s1413* and *s1495)* show about 15% and 25% respectively shorter body length than the control. These values are significantly different from the control ($P <0.05$).

At 72 hrs both, *let-336(s1413* and *s1495)* are significantly shorter ($P <0.05$). *s1413* was about 17% shorter and *s1495* was about 41% shorter. At 72 hrs, eggs were

visible in the control (BC2200) hermaphrodite's gonad; however, for *let-336 (s1413* and *s1495)* no eggs were visible in the gonad indicating that the worms were not adults.

The observed developmental delay represented by the shorter body length is a possible haplo-insufficiency phenotype in *C.elegans*. Even though *let-65(s1777)* and *let-336(s1495)* encoded proteins are truncated with a longer carboxyl end than those encoded by *let-65(s1222)* and *let-336(s1413)*. The alleles producing the longer predicted protein product exhibited a greater developmental delay. This could be because these alleles are partially functional and interfere with the function of non-mutated proteins.

In *C. elegans,* life-span extension was reported to be exhibited as a result of inhibiting translation by of mutations in ribosomal proteins (Hansen *et al.* 2007). Hansen *et al.* (2007) suggested that in *C. elegans* life-span extends is a result of cells shifting from states favouring growth to states favouring maintenance and stress resistance (Hansen *et al.* 2007).

In my investigation, I sought to determine the effect of a ribosomal gene mutation on life-span and life span variations as possible haplo-insufficiency phenotype. This was conducted by observing the survival rate of ten worms incubated at a temperature of 20 $^{\circ}$C. *let-65(s1222* and *s1777),* show mean life-spans that are not significantly different from the control BC1270 ($P > 0.05$). *let-336(s1413)* also has a mean life-span that is not significantly different from the control BC2200 ($P > 0.05$). However, *let-336(s1495)* showed a 49% increase in mean life-span which is significantly different from BC2200. Similarly to what was observed in the developmental delay investigation, *let-336(s1495),* although its encoded protein has a longer carboxyl end than *let-336(s1413)*, it showed the

stronger effect. This could be because the *s1495* encoded protein might be partially functional and interferes with the function of the non-mutated protein.

In this chapter, I investigated the possible existence of haplo-insufficiency phenotypes resulting from mutations in genes that encode proteins involved in translation. My findings suggest that brood size variations, developmental delay and life-span extension are possible haplo-insufficiency phenotypes in *C. elegans*. To confirm these results, more investigations are needed using lethal alleles of essential ribosomal genes. Moreover, the same type of investigation could be conducted using AARS genes and other genes encoding non-ribosomal proteins involved in translation to determine if they exhibit putative haplo-insufficiency phenotypes.

GENERAL DISCUSSION

Essential genes are evolutionary well conserved genes that are necessary for an organism to grow to fertile adulthood and are required for the subsequent generation to reach fertility (Kemphues 2005). In my thesis I demonstrated the molecular identification of the essential gene *let-65,* in *C. elegans*. *let-65* is a large target for mutagenesis and one of four genes with larger number of alleles (nine) in the unc-22 region on the right side of LG-IV. *let-65* has been previously categorized as an essential gene because individuals homozygous for eight of its alleles arrest in a mid larval stage and animals homozygous for the ninth allele are sterile (Rogalski *et al,* 1982; Moerman 1980; Rogalski and Baillie 1985; Clark *et al.* 1988, and Clark 1990).

In Chapter one, I used sequencing and complementation testing to determine the molecular identity of *let-65.* I sequenced the two candidate genes for *let-65 (F58B3.4* and *F58B3.5)* in *let-65* homozygous worms for eight alleles. Each allele had a G/C to A/T transition, which is characteristic of EMS induced mutations. These alleles have amino acid changes in well conserved residues *let-65*(*s254, s694, s1083, s1084, s1154,* and *s1730). let-65(s1777)* truncates the protein prematurely at amino acid 408. *let-65(s1222)*, has a mutation in the first base of the first intron thus preventing intron splicing and therefore, causes a premature truncation of the protein at amino acid 28 (Fig. 3).

To confirm the sequencing results, I conducted a complementation test and rescued two of *let-65* alleles *(s1083* and *s1222)* using the fosmid *WRM0615dH10* that spans the genomic region that contains *mars-1*.

mars-1 encodes a methionyl tRNA synthetase (*MARS-1*). *MARS-1* is one of at least 20 cytoplasmic aminoacyl tRNA synthetase enzymes in *C.elegans* (AARS). AARS enzymes represent a central component of the cellular information transfer system in translation by catalyzing aminoacylation of their cognate tRNAs (O'Donoghue and Luthey-Schulten 2003).

MARS-1 in *C. elegans* is a 917 amino acids long protein that is well conserved from that I propose that this is the reason it is a large mutagenesis target. *MARS-1* consists of multiple functional domains: the catalytic core domain distinguished by a Rossman fold, a tRNA binding domain, and an anticodon-binding domain. Seven out of the eight alleles have mutations in the core domain (*s254, s694, s1088, s1084, s1154,* and *s1730*). *let-65(s1777*) truncates the protein such that it excludes the tRNA binding domain and the anticodon-binding domain. In *C. elegans* there are 36 AARS genes that are distributed non-uniformly on autosomal chromosomes while none are on the X-chromosome. This could be because X-linked genes are transcriptionally silenced during mitosis and early meiosis (Kamath *et al.* 2003)

Inside a living cell, proteins are directed to their finale sub-cellular destinations by specific amino acid signals. There are two forms of AARS enzymes. One form functions in the cytoplasm and the other functions in the mitochondria. Both forms function in translation.

In chapter two, I determined the sub-cellular localization of MARS-1 by making constructs of which GFP encoding sequence was fused, in-frame, with *mars-1*. This showed that MARS-1 localizes in the cytoplasm and may therefore functions in translation with cytoplasmic ribosomes. Moreover, I used computational methods (BlastP

and WoLF PSORT), to determine the sub-cellular localization of all known AARS enzymes. My final determination of the locations of the AARS proteins was based on BlastP *S. cerevisiae* matches. To confirm the computational approach, I determined the sub-cellular localization of *sars-1* and *sars-2* by constructing GFP constructs. From this I determined that *sars-1* (previously named *srs-2*) encoded protein localize to the cytoplasm and *sars-2* (previously named *srs-1*) encoded protein localize to the mitochondria.

As a result of this work, *C. elegans* AARS genes were renamed from xrs- to xars-, to bring them into line with their homologs in other organisms. Moreover, to standardize the AARS gene nomenclature, genes encoding cytoplasmic proteins were renamed "xars-1" and genes encoding mitochondrial proteins were renamed to be "xars-2".

Regulation of genes takes place in multiple levels: Pre-transcriptional, transcriptional, translational, and post-translational level (Lee 2009). Transcription is regulated by activator and repressor proteins known as transcription factors that regulate the binding of Pol II to the gene's promoter region. In the third chapter, I discuss my examination of the transcriptional regulation of *mars-1* by identifying an upstream *cis*-regulatory motif containing region for *mars-1*. This was done by generating reporter constructs of GFP encoding sequence fused to sequential *mars-1* upstream cut downs. These constructs were injected into worms to generate transgenic animals and these were examined for the presence or absence of fluorescence indicating genes expression. I narrowed the region containing the putative *cis*- regulatory elements responsible for *mars-1* expression to a 66bp window. Moreover, using FamilyRelations II and by

aligning the region, I found two blocks of sequences which are also in *C. briggsae* and *C. remanei,* as well as *C. elegans* that could contain *cis*- regulatory elements.

In *Drosophila* mutations in genes encoding ribosomal proteins cause haplo-insufficient *'Minute'* phenotypes. Individuals with *Minutes* syndrome exhibit three visible phenotypes; short slender bristles, delayed development, and recessive lethality (Schultz 1929). These individuals could also exhibit other phenotypes. The common *Minutes* variable phenotypes exhibited by RPs could be because they are required in equimolar numbers; therefore, reducing the production of one RPs causes reduction in the number of functional ribosomes resulting in impaired protein synthesis (Marygold *et al.* 2007).

In chapter four, I investigated if mutations in genes that are involved in translational result in haplo-insufficient phenotypes in *C. elegans.* In my investigation I used two alleles of the ribosomal protein *let-336(s1413* and *s1495)*. Moreover, since it was suggested by Lambertsson (1998); that a complete or partial inactivation of aminoacyl-tRNA synthetases may lead to a *Minute* or a phenotype similar to *Minute* phenotype, I also used *let-65(s1222* and *s1777)* which encode abnormal *MARS-1* proteins in my investigation. As controls for my experiments, I used BC1270 (*let-65* control) and BC2200 (*let-336* control) that share the same genotypes of examined alleles but without the EMS induced point mutations.

The first putative haplo- insufficient phenotype I looked at was the self-progeny brood size. Although no significant reduction in brood size was observed for the two *let-336* alleles, *let-65(s1222* and *s1777)* both exhibited a significant brood size reduction. The second putative haplo-insufficiency phenotype I investigated was the developmental delay which I did by measuring the body lengths of 20 animals of each strain at three

time points of development (24, 48, and 72 hrs). Significantly smaller body length was observed for *let-65(s1777)* at 72 hrs compared to that observed for the control strain BC1270. For *let-336(s1413* and *s1495)* both alleles showed significantly smaller body lengths at 48 hrs and 72 hrs. Since in *Drosophila*, one of the three main characteristic of the *Minute* phenotype is delay in larval development, these results indicate that developmental delay is a possible haplo-insufficiency phenotype in *C. elegans.*

As part as my investigation I considered life-span extension as a putative haplo-insufficiency phenotype. By observing the survival rate of 10 worms each per strain incubated at 20°C, I found that, with the exception of *let-366*(*s1495),* there is no significant difference in the life span between the strains and their controls. *let-336(s1495)* showed a significant increase of 49% in mean life-span to its control. More sever phenotypes observed for the alleles with the least sever mutational changes (*let-65(s1777)* and *let-336(s1495)*) could be because these mutated proteins might be partially functional and may interfere with the function of the non-mutated protein.

APPENDICES

Appendix A. *let-65* alleles and their lethal phenotypes.

For information on the isolation, mapping and characterization of alleles from Reference 1 , see (Rogalski *et al.* 1982, Moerman, 1980), for isolation and mapping of alleles from reference 2, see (Rogalski and Baillie 1985), reference 3 is (Clark *et al.* 1988) and for reference 4 see (Clark 1990).

Gene	Strain	Allele	Genotype	*let-x/let-x* arrest phenotype	Reference
let-65	BC1898	*s174*	*let-65(s174)unc-22(s7)/ nT1;+/nT1*(V)[*lin-(n300)* on *nT1)]*	larva arrest	1
	BC00962	*s254*	*let-65(s254) unc-22(s7)/nT1*(IV); +/*nT1*(V)	larva arrest	1
	BC1121	*s694*	*let-65(s694) unc-22(s7)/+;* + (IV)	Sterile	2
	BC2028	*s1154*	*let-65(s1154) unc-22(s7) unc-31 (e169)/nT1*(IV); +/nT1 (V)	larva arrest	3
	BC2116	*s1222*	*let-65(s1222) unc-22(s7) unc-31 (e169)/nT1*(IV); +/nT1 (V)	larva arrest	3
	BC1909	*s1083*	*let-65(s1083) unc-22(s7) unc-31 (e169)/nT1*(IV); +/nT1 (V)	larva arrest	4
	BC1910	*s1084*	*let-65(s1084) unc-22(s7) unc-31 (e169)/nT1*(IV); +/nT1 (V) [*lin- (n300)* on *nT1)*	larva arrest	4
	BC3258	*s1730*	*let-65(s1730) unc-22(s7) unc-31 (e169)/nT1*(IV); +/nT1 (V)	larva arrest	4
	BC3305	*s1777*	*let-65(s1777) unc-22(s7) unc-31 (e169)/nT1*(IV); +/nT1 (V)	larva arrest	4

Appendix B. *F58B3.4* and *mars-1* wild type sequence and primers used for sequencing *let-65* alleles.

>*F58B3.4* (*IV*:11628417,11632118)

gcttctttttttctaagtttctcataatatttaagttgttttcaattgttgaccgtttgggactcgtttt
tatttatttattcttcatagccttccgatttatattcgctgattttgtatcccattttgaattcatcaatt
catatcggttcagtctctttttgggcacgttttctattatcaagacataaagtttattgtttcctgtattt
ttatgtagttactgaattattaatagtggtcggattttcactcaaaatacttttttccaactttcaaattg
gaaaagaatctatagagcattcagaaaatgggaaaaaagaagtctggcaaggcaccagctgggctgaatga
tgaacgcttctcgcaaatcaaatcaaatccgatgtttgtcggattgaagaactcggagaagaaggtaaaaa
tgcatcattttcactcctgtcctgaagtttgaaaaaaaattaaaactttaattcataggatttttacatt
ttttttgatttttcgcgttttccgaatttactgagtgataattcctaatttttgtgatgttaaacacc
ttttgtcaattttctgagatttttccgtataaaaactataacccaaaagaaattctctgaaatttcaaaat
gttcctctctaataagttaacttaaattaataatttttttgagaaaaattgtctcaaaattgacaaacggt
gttatagaaaaatgaaggaaaaccaataaatatagccggaaaactcgaaaatgacaaatttttaatctaaa
aaatatttgaaaatttggaaaaatttaataattggtaaacagttatacttggtcatcaaggtccattttat
taatatttaaagcaaaatttcaaaatgactctactccacattgaaactttcaattttcaggtagtaattga
caaaagatttgctgctgctctcactgatgagcggttttcaactcgagcaaaagtggatatgcgaggacgaa
aacagaagaaaactgtcggaaatagtatgttagatctgtacgaattggaagaggatgaggaagttcaacca
ttgaaaaaagaaaaaatatcgaaaaagaccacaaaagatgaagatgatgaactggatgatttctttgatga
gaaagatgatgatgaagatgcattggaagatgctgatgatgaagaggacgagatttcagatgaagagaatg
aagatgaagaggaagaagaaaatgaaaaaataggaatgaacggattcaaaagactcgatttggctcgtgga
gaaggaaatgttgattcatcttcagatgatgattcaagtgaagatgaaggtgaagatgttgaatttgatga
gaaggaaggtggaatcgagctcgatttagctaatcttgataggtaaaaatttgcaacaataagtacacaag
taatttttaaataatttcagagaagtagaccaagttgaatggacatcaaaccgtctggcagtatgtaatc
ttgaatgggacactatgaattgcgaagatattctgatgcttgtcaaatcatttgtgttagtttttcattta
aagaggagtgaccgatcaaaatttttggttttattatactcaaatttgagctcaaaatcaagggtgtcgt
caaaaaattctcgatttctatttttcagggaaaatcgaaaaaaatttttttaaaaatatgttgaaaaaaggt
gaaaactatacatttatcgaaactgttgattttgatgagagaattgagaagtgactcaaaagtattcaaag
ttaaaaaaaaacgaaatatcagacagtagaaagttttggataatccgatattcgaatatatttatgatttt
caaaaaaaaaaaatcgaaaaaccaacatctttgctcaaaatgacctcaaaagattggattttaaacgaagc
ttttcacaaatttctcaaaaaaaaatttatgactgataataattaggctatttttccgaaattctgaaccg
tcataaaacatttttttagaagtctcgttacaaaatttggtatttttcgggccattttgagtataataaag
ctagaaatttttttatagtccttattgtatttaattaaaccttcttcagaccacaaggtggatctgtaattt
ccgttggaatctatctttcggattttggaaaaggacaattgagtaaagaagaaaaaactggaccacttcta
aaattggcaaaacctgtggaagaatacaaagaagatgaaatggatgagtaagatttattgtttttttactct
aagaacaaaaaatgttattctataatttcagtgaaacgagaacagcagtacgggagtatttagtcaacaag
ctcaagtactattattcagttatcacttttgattcaattccgtcggcagttgcagtttacgaggaatgtga
tggttttcaatttgaagaaaccggtctgaaaatggatatgcgatttattccagatgacatggattttgagg
taaattaaaatattttcataatatctatatatatattgaataaaaattcaggaagatcgagtaaaagaatt
tctgaatgcggaagatgtgaatttgacgaaatataaagcaaaaaagaaatcgaaaagtgcaattatctcaa
ccggagccaaaattctatgggacgaggatgatccacaaagaaagaagaagtttcttgaagctttcaatggt
gatgaagatgctggaaaagatttaattgtggactctgatgaatcggatggtgatgaagcaaatcggaaaac
acttatggctctgttgaataccgatgaacgaaagtcaaaattagatgttgattgggaaggagaagagaaga
atggaagtgattcctctgatggagaatatgttaaagtagatgatgatgatgaagagattggagtaaagcaa
aagaaaaagaaaggatctgatgatgatgaagatgaaaaagaagaggtagaagaaaagaaagaagatgtaaa
gctcacaggatacaaagcatataagaagaaacaaaaacaaaagctgatggaaaataagttgaaaagaaagg
gaacatcaaaagaagctgagactaacatcaagaccgttgctgcagccgacagtatctcgaaagacgaccgt
ttctctgctttattcaccgattctgcatatgctattgagccaagttccaagaaattcaagggatcgcttct
tgtgacgaagcaggctgaacaaaagtctaaaggatctagcactgtcatcgaaacgaagaaacccgaggatc
tcgttcagaagttgaagaaacaagctgataaatggaataagaagaagtctgtgaaaaactgattttatttt
tatgtttttgttctgttatttttttatcgttacgttatgaattgttttccattgcttaacaatttcttggta
ataggaatatgaccatcttgcaaataatttggaaccattgaccatttcattacaatgcggaagataagaaa
ttgggctcgtttgataagcatttccgatttgttaagagtgtcacgaatcggattatagtgaagatggaaag
atcgcattggattttgagtgtcaatttccagttttcaacttaaaatttattaccctgacggggataaatac

acacttttgtgaattaaaaagtttttttcacaatttgtaaaaatgtaaatatgtgttttattgtttttat
cgaaaaaaaa

Underlined: gene

>*mars-1* (*IV*:11631554,11635810)

caaagataaacaaaaaatacttgaaaatgcgggagatgtgatttccaacgattgtagtatccattttgtgt
agtgactccataaggtgtgcaagtgtgctctacggagaattagcaagaaattccattgaactgaaatttta
agatgtttttttcatctcttgctcattattattttgtttttaaagtgatattcagcatttaaaattgaaat
tcaatgtgtattttacttcactctatcacaatcagaaccgcttagcacatttctaattgtttttatgaagt
tttaggccttaagacttttgattgacagttgccctgactccaaatgcttcttaattttaatatattcagt
gacgccttgccacaatgggtcacgacttggcggatataaaaaaatctttcgaggcaagtttgccgggatat
gttgagaagaaagatccaaaaaggtaattgggcaacttttttcgaagttaaatgtttaataatttataatt
caagtattctaccgcaaccaggaaagcgaaacattttgattactgctgcattgccatacgtgaacaatgtt
ccacatctcggaaacattattggatgtgttctcagtgctgatgtctttgcccgttattgtaatctacgagg
acatcaaacctttacgttggaggaacagatgagtatggaacggcaacagaaacaaaagcacttcaagagg
gatgtactccacgagagttatgtgacaaatatcatgccattcataagggaatctatgagtggttcggaatt
gatttctctcatttcggaagaaccaccactgatcatcaaactgagtaggtttaattgaatttaagttttat
attattcaattgatttcagaatttgtcaagacatgtttttgaaacttcacaaaaatggctacacatcatcg
caatcggtggatcagctgtactgcaaccaatgcgaaaaattcttggctgatcgattcgtcaccggaacttg
tccaatgtgcgcttatgatgatgctcgtggagatcagtgtgatggatgtggaaagttgatcaatgctgtgg
atcttaaggacgcgaagtgtcatatgtgtaaagcaactccagaagtgaaacagtccacgcacatcttctta
tcacttgacaagcttcaacagaaaacaacagaacacttggatcgtgaattagctaaagaagataatcgctg
gtcatcaaatgctgttggaattaccaaggcttggatgaagttaggattggatcctcgttgcatcacaagag
atctcaaatggggaacggctgttccattggatggattcgagaagaaagtgttctacgtatggtttgacgcc
ccaattggttatctttcaatcacaaaatgtgtactcggtgataactggactaaatggtggaagaatccaga
gaatgttgagctgttcaatttcgtcggaaaggacaatgtagctttccacgccgtcatgttcccttgttctc
aattgggtgccaatgacaactatacagttgttaataatctctgtgccactgaatatctcaactatgaggac
acaaaattcagcaagtcacgtggaaccggaattttcggagatgcagctcaaggaactgaaattccagcaga
tatctggcgattttacctcttatacatgagacccgaaagccaagatactgcttttagctgggatgactttg
tgttgaaagttaacagcgaattgctgaataatcttggaaatttcatcaacagagcactttcgttcgtcgca
aactcttttggaggagttgttccagaaatgaatttgacaaatgatgacgcagaagttctttctgaaattca
caatgaatgtatgcaatgggataagcagttcgacggagttcatctcaaagatgctgtgaagactattctaa
atgtttctcgtcttggtaaccaatatatgcaagctcaaactccatgggtgctcatgaaaaaggatgaagaa
ggaaagaagagagctggaacaatcatcggagtggctgctaacatcgcctatcatgtttctgttcttttata
cccaatcatgccaacaatttctgcaacaatccgtgaacagtgtggtttaccagcacttcctctcttcactc
cattcccgatttgttatcttaaagccggtcacaaaatcgggcaaccatcccctcttttccaaaaactcgat
cctgcacaaattgccgaattcaaagcaaagtttggtggttctcaagatgcccagagctccgctccaaaaac
tgccgaaaagccaaaacaacagaagaagcaagcacccaccaaagataaaaagggtaattaaaatgcttggg
gagtcactgctagatatcgcttttgtctgatgtaatgtgtttatgcgaatcggtgctttctcagcaagcgt
cttttagttaattatgtagatgtagattcgtagtcatttaccgtattttaggtggagaatggtggagtcgg
cgagctgagaagtacactgagagctgcacacaatgcattcaactgcgcatggattgcatgccgaaatttta
ctatttttgttgtttcaactctccctttctctacttcccgttcaatttatcaacctccccgcgcagtctt
ctgcttcataacggttttattattacaggagacaaaaagatggcgtcaactgctgcttttgtggaacttga
acaaggagccaaggttatttcgcaactcatcgctcaaaatctgaaaaagtttgatcaagccagtacgttat
ttatatagctgccctatatttagttcatccataaattttcagaggccctcttcacccgaaatcagctccag
agattggatggagagaacaagcagcttacgattgatgttaaaacacttcagcatcaattgattgaactgga
aactgctgccggaatcaagcaagtgccaaaaccagttgtttcgtgtacaccaacaccaacttctactccag
cttctggaattatcacagaagctccaaagaaagaggctccttcaactccagcaccttccgagccgaagaaa
gctaaggagcagaagaaaggaaagggtggggctgctgcggctccagtcgacgatacaatcgatgtgggaag
actcgatatgcgcgttggtcgaatcatcaagtgtgagaagcatcctgatgcagatgctctttacgtggaac
agatcgatgttggagagtcagccccgcgaactgtagtttctggactcgttcgtcacgttccactcgatcaa
atgcagaaccgccttgtcgttgtgctttgcaatctgaaaccagcaaaaatgcgtggagtagagtcccgagc
catggttatgtgtgcttcatcacctgacaaagttgagattatggaagttccggctgattccaaaccaggta
ctccagtcgtctgcccaccatatacacacagacctgatgagcaattgaatccaaagaaaaaggtaaaaatt
ttaatatttttttttaatgatagaatatctgattattgaaaattattaaacaatatttttttccagatctggg

agacagttgcagaggacttgaaggtttcggctgaaggcttcgcagagtggaaaggacaaccgcttctcatt
ggaagtgaatccaaaatgactgccccaacgcttcgtggagtccatgtgaaataattggttttctgcattta
gcttctttttttctaagtttctcataatatttaagttgttttcaattgttgaccgtttgggactcgtttt
tatttatttattcttcatagccttccgatttatattcgctgattttgtatcccattttgaattcatcaatt
catatcggttcagtctctttttgggcacgttttctattatcaagacataaagtttattgtttcctgtattt
ttatgtagttactgaattattaatagtggtcggattttcactcaaaatacttttttccaactttcaaattg
gaaaagaatctatagagcattcagaaaatgggaaaaaagaagtctggcaaggcaccagctgggctgaatga
tgaacgcttctcgcaaatcaaatcaaatccgatgtttgtcggattgaagaactcggagaagaaggtaaaaa
tgcatcattttcactcctgtcctgaagtttgaaaaaaaattaaaactttaattcataggatttttacatt
ttttttgatttttcgcgttttccgaatttactgagtgataattcctaatttttgtgatgttaaac

Underlined: gene

Gene	Sequencing Primer	Sequence
F58B3.4	Forward-1	TTTGTCAATTTTCTGAGATTTTTCC
	Forward-2	CGAATTGGAAGAGGATGAGG
	Forward-3	ATGGACATCAAACCGTCTGG
	Forward-4	GTATTTTTCGGGCCATTTTG
	Forward-5	CGATTTATTCCAGATGACATGG
	Forward-6	TGGAAGTGATTCCTCTGATGG
	Forward-7	TGTCATCGAAACGAAGAAACC
	Reverse-1	AATGAAATGGTCAATGGTTCC
	Reverse- 2	CGGTCTTGATGTTAGTCTCAGC
	Reverse-3	GCTCCGGTTGAGATAATTGC
	Reverse-4	TGTATTCTTCCACAGGTTTTGC
	Reverse-5	TCTCATCAAAATCAACAGTTTCG
	Reverse-6	CAACATTTCCTTCTCCACGAG
	Reverse-7	TTAATAAAATGGACCTTGATGACC
	Reverse-8	CTTCAATCCGACAAACATCG
F58B3.5 (Part 1)	Forward-1	TGAATGGTTTCATTGATATTACAGG
	Forward-2	GCCGGGATATGTTGAGAAGA
	Forward-3	ATTCGTCACCGGAACTTGTC
	Forward-4	CCGTCATGTTCCCTTGTTCT
	Reverse-1	AGCCACTCCGATGATTGTTC
	Reverse-2	AGTTTGCGACGAACGAAAGT
	Reverse-3	CTGTTGCCGTTCCATACTCA
F58B3.5 (Part 2)	Forward-1	AGCAGTTCGACGGAGTTCAT
	Forward-2	GTTTATGCGAATCGGTGCTT
	Forward-3	AGCTCCAGAGATTGGATGGA
	Forward-4	TCGTCTGCCCACCATATACA
	Reverse-1	GAGTCCCAAACGGTCAACAAT
	Reverse-2	GCTTCTCACACTTGATGATTCG
	Reverse-3	CGGGGAGGTTGATAAATTGAA
	Reverse-4	CACTGTTCACGGATTGTTGC

Appendix C. BlastP (scores and E-values), and WoLF PSORT localization scores for AARS localizations predictions.

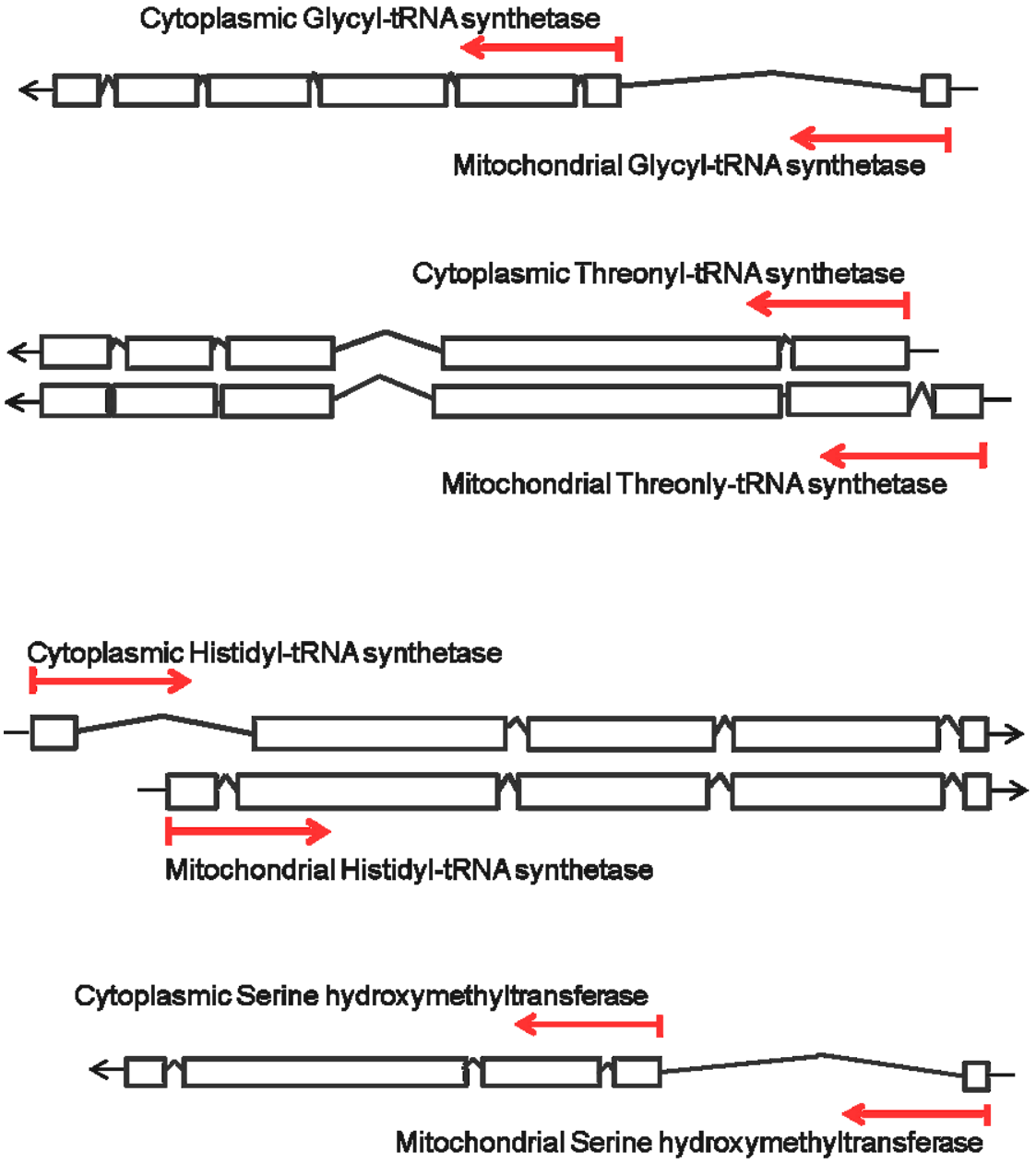
AARS enzyme	Previous genes and sequence Nomenclature	BlastP scores	BlastP E-value	Wolf PSORT scores
Alanyl (A) ARS	*ars-1, (W02B12.6)*	506	2e-142	cyto: 22.0, nucl: 7.0
Alanyl (A) ARS	*ars-2, (F28H1.3)*	907	0	cyto: 21.0, mito: 4.0, plas: 3.0, nucl: 3.0
Cysteinyl (C) ARS	*crs-1, (Y23H5A.7a)*	552	3e-156	plas: 14.0, mito: 10.0, nucl: 3.0, E.R.: 3.0
Cysteinyl (C) ARS	*crs-2, (Y23H5A.1)*	-	-	-
Aspartyl (D) ARS	*drs-1, (B0464.1)*	532	2e-150	cyto: 18.0, cyto_nucl: 14.8, cyto_pero: 12.2, nucl: 6.5, pero: 4.5
Aspartyl (D) ARS	*drs-2, (F10C2.6)*	276	3e-73	mito: 28.5, cyto_mito: 15.5
Glutaminyl (Q) ARS	*ers-1, (Y41E3.4)*	529	3e-149	cyto: 15.5, cyto_nucl: 12.5, nucl: 6.5, pero: 3.0, cysk: 3.0
Glutamly (E) ARS	*ers-3, (T07A9.2)*	258	7e-68	cyto: 18.0, mito: 8.0, pero: 3.0
Glutamly (E) ARS	*ers-2, (ZC434.5)*	515	9e-145	mito: 14.0, extr: 6.0, cyto: 5.5, cyto_nucl: 5.0, nucl: 3.5
Phenylalanyl (F) ARS	*frs-1, (T08B2.9)*	489	2e-137	cyto_nucl: 15.5, cyto: 15.0, nucl: 14.0
Phenylalanyl (F) ARS	*frs-3, (Y60A3A.13)*	279	2e-74	mito: 21.0, nucl: 8.0
Phenylalanyl(F) ARS	*frs-2, (F22B5.9)*	538	3e-152	cyto: 28.5, cyto_nucl: 15.0
Glycyl (G) ARS	*grs-1, (T10F2.1)*	611	7e-174	mito: 29.0
Histidyl (H) ARS	*hrs-1, (T11G6.1)*	394	1e-108	cyto: 20.0, cyto_nucl: 12.7, cyto_plas: 11.7, pero: 5.0, nucl: 3.0

AARS enzyme	Previous genes and sequence Nomenclature	BlastP scores	BlastP E-value	Wolf PSORT scores
Isoleucyl (I) ARS	*irs-1, (R11A8.6)*	1118	0	cyto: 16.0, cyto_nucl: 12.3, cyto_mito: 9.8, nucl: 7.5, pero: 3.0, mito: 2.5
Isoleucyl (I) ARS	*irs-2, (C25A1.7)*	433	3e-120	mito: 17.0, nucl: 12.0
Lysyl (K) ARS	*Krs-1, (T02G5.9)*	622	2e-177	cyto: 18.5, cyto_nucl: 15.0, nucl: 6.5, pero: 3.0
leucyl (L) ARS	*lrs-1, (R74.1)*	838	0	nucl: 18.0, cyto_nucl: 17.0, cyto: 14.0
Leucyl (L) ARS	*lrs-2, (ZK524.3)*	405	6e-112	nucl: 16.5, cyto_nucl: 12.5, mito: 8.0, cyto: 7.5
Methionyl (M) ARS	*mrs-1, (F58B3.5)*	610	1e-173	cyto: 16.0, nucl: 12.0, cyto_mito: 8.8, cyto_plas: 8.8
Methionyl (M) ARS	*Y105E8A.20*	249	2e-65	cyto: 21.5, cyto_nucl: 15.8, cyto_pero: 12.3, nucl: 6.0
Asparaginyl (N) ARS	*nrs-1, (F22D6.3)*	558	3e-158	cyto: 16.5, cyto_nucl: 13.0, nucl: 6.5, mito: 4.0, cysk: 4.0
Asparaginyl (N) ARS	*nrs-2, (F25G6.6)*	280	2e-74	cyto: 20.5, cyto_nucl: 13.0, mito: 5.0, nucl: 4.5
Asparaginyl (N) ARS	*Y66D12A.23*	209	4e-53	nucl: 26.0, cyto: 5.0
Prolyl (P) ARS	*prs-1, (T20H4.3)*	590	1e-167	cyto: 14.5, cyto_mito: 8.7, nucl: 7.0, pero: 4.0, mito_pero: 3.8, plas: 2.5, extr_plas: 2.5
Prolyl (P) ARS	*prs-2, (T27F6.5)*	194	2e-28	mito: 21.0, cyto: 7.0, pero: 2.0

AARS enzyme	Previous genes and sequence Nomenclature	BlastP scores	BlastP E-value	Wolf PSORT scores
Arginyl (R) ARS	*rrt-1, (F26F4.10)*	180	3e-44	cyto: 18.5, cyto_mito: 10.7, nucl: 7.0, mito_pero: 2.8, pero: 2.0
Arginyl (R) ARS	*rrt-2, (C29H12.1)*	212	6e-54	cyto: 17.0, mito: 10.0, pero: 3.0
Seryl (S) ARS	*srs-1, (W03B1.4)*	196	2e-49	nucl: 21.0, mito: 7.0, cyto: 4.0
Seryl (S) ARS	*srs-2, (C47E12.1)*	414	1e-114	cyto: 20.5, cyto_nucl: 14.8, cyto_pero: 11.7, nucl: 7.0
Threonyl (T) ARS	*trs-1, (C47D12.6)*	793	0	mito: 26.0, cyto_mito: 16.0, cyto: 4.0
Valyl (V) ARS	*vrs-1, (ZC513.4)*	518	1e-145	cysk: 25.0, cyto: 5.0
Valyl (V) ARS	*vrs-2, (Y87G2A.5)*	928	0	nucl: 20.5, cyto_nucl: 16.0, cyto: 10.5
Tryptophanyl (W) ARS	*wrs-1, (Y80D3A.1)*	449	2e-125	cyto: 27.5, cyto_pero: 15.0
Tryptophanyl (W) ARS	*wrs-2, C34E10.4*	201	2e-50	nucl: 30.0
Tyrosinyl (Y) ARS	*yrs-1, K08F11.4*	166	2e-40	mito: 25.0, cyto: 4.0
Tyrosinyl (Y) ARS	*Y105E8A.19*	450	2e-86	cyto: 14.0, cyto_nucl: 9.8, pero: 9.0, cyto_mito: 7.8, nucl: 4.5

Appendix D. *gars-1*, *hars-1*, *tars-,1* and *mel-32* gene models.

The shown models are obtained from WormBase, WS215. The models do not represent the proportional sizes of the genes.



Appendix E. Forward primers used in promoter amplification reactions to generate *let-65*$_{promoter}$::GFP transgene constructs.

Distance from ATG start site (bp)	Forward primer
-2875	GGAACATCGTCTTCATCATCATT
-2484	GATGTTTATCGGCTGCTTCAT
-1923	TTGTCTGGAAGTCAGCATCG
-1248	TTCGCTGGTCCGATATAAGG
-603	AAACAATTCGGAGAGGTTTTCAG
-526	GTGAGTCGCGTTGTTATTCG
-477	TTGCGAAGTTTTACAGCGTTAG
-344	AATGCGGGAGATGTGATTTC
-283	TGTGCAAGTGTGCTCTACGG
-124	CAGAACCGCTTAGCACATTTC
-58	AGTTGCCCTGACTCCAAATG

Appendix F. Genotypes of the generated *mars-1*$_{\text{promoter}}$::GFP strains.

Strain	Genotype
BC8708;sEX2677	*dpy-5(e907)/dpy-5(e907)[F58B3.5*$_{p(-2,521)}$*::GFP+pCeh361]*
BC8709;sEX2678	*dpy-5(e907)/dpy-5(e907)[F58B3.5*$_{p(-2,053)}$*::GFP+pCeh361]*
BC8710;sEX2679	*dpy-5(e907)/dpy-5(e907)[F58B3.5*$_{p(-2,053)}$*::GFP+pCeh361]*
BC8711;sEX2680	*dpy-5(e907)/dpy-5(e907)[F58B3.5*$_{p(-1,393)}$*::GFP+pCeh361]*
BC8967; sEX2927	*dpy-5(e907)/dpy-5(e907)[F58B3.5*$_{p(-603)}$*::GFP+pCeh361]*
BC8964; sEX2924	*dpy-5(e907)/dpy-5(e907)[F58B3.5*$_{p(-477)}$*::GFP+pCeh361]*
BC8965; sEX2925	*dpy-5(e907)/dpy-5(e907)[F58B3.5*$_{p(-477)}$*::GFP+pCeh361]*
BC8960; sEX2921	*dpy-5(e907)/dpy-5(e907)[F58B3.5*$_{p(-344)}$*::GFP+pCeh361]*
BC8961; sEX2920	*dpy-5(e907)/dpy-5(e907)[F58B3.5*$_{p(-344)}$*::GFP+pCeh361]*
BC8968; sEX2928	*dpy-5(e907)/dpy-5(e907)[F58B3.5*$_{p(-283)}$*::GFP+pCeh361]*
BC8966; sEX2926	*dpy-5(e907)/dpy-5(e907)[F58B3.5*$_{p(-283)}$*::GFP+pCeh361]*
BC8969; sEX9929	*dpy-5(e907)/dpy-5(e907)[F58B3.5*$_{p(-124)}$*::GFP+pCeh361]*
BC8995; sEX2955	*dpy-5(e907)/dpy-5(e907)[F58B3.5*$_{p(-58)}$*::GFP+pCeh361]*

Appendix G. Upstream regions of *mars-1* in *C. elegans*, *C.briggsae,* and *C. remanei*.

C. elegans

>IV: *mars-1* (- 683bp to +3bp upstream region sequences)

Ctgaggaacatgatgaggataaatcaataaaacaaagctgatttacttatttttttgcttttggaaattaa
aaaaaaaaacaattcggagaggttttcagatctgatcagagtgagaaaagaaaatattcctcagactctca
aataagagtgtagtgagtcgcgttgttattcgcgagaaaataaattcgatagaaatcaactttgcgaagtt
ttacagcgttagtcttggatttagattatttgtgaagtttaattgaatggtttcattgatattacaggttc
aataaaaattattactatattcaatagcaaagataaacaaaaaatacttgaaaatgcgggagatgtgattt
ccaacgattgtagtatccattttgtgtagtgactccataaggtgtgcaagtgtgctctacggagaattagc
aagaaattccattgaactgaaattttaagatgtttttttcatctcttgctcattattattttgtttttaaa
gtgatattcagcatttaaaattgaaattcaatgtgtattttacttcactctatcacaatcagaaccgctta
gcacatttctaattgtttttatgaagttttaggccttaagacttttgattgacagttgccctgactccaa
atgcttcttaattttaatatattcagtgacgccttgccacaATG

C. briggsae

>IV: CBG06108 (-862bp to +3bp upstream region sequences)

gtatgtggtccgctgtgggattcgcgaggcgagactctgttgacttttttgagaccccgcaatttccgggg
aaaaaaaaatcggggaaaaaaaaaccggggaaagtgggcggagtcgaaaaaaaaccgtgcggcaaaaccgt
gccagggttgcaatttaggctccgccccctcatgaccgtgcccaaatgtgcagaaaaaccctgcggcaaaa
ccgtgccagggttnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnntgcggggtctcactcttttgagcgcgggatgcaa
gtgcgccccatagagaattgaaaaaaaaaattcctgaaattctgaggacgaattttatttttggctcattt
gtgttgcgaaattcggtataatctcacttttcatatttcaaaaacatgctaagcttagtttttgaaaagaa
ccttaataaaacatcatcacaggcggaaaaataagctaatgttacaaaagtttaggcctactctgtttttg
attgacatattaatcgtggtcatttcgtttcctctataatataactttcttcagtgacaagtagtcacgAT
G

C. remanei

> IV: CRE12616 (-683bp to +3bp upstream region sequences)

aaacttgcaaaggtttgttgttaagattttcagaacaaatttcagagatctcaaagagctttatgattcag
ccggttaactcattgcttttgtgtcgctgagattgcgtactcatctgaaaatcaaaaaaagcaaaaaccttt
gatatgatagttatttttttcagaaatggatcgacaacgatgcggaaatctgcaccaaactccaatgtgg
tttaagcgcccgttcagatgtttatctgctcaattttgttgaggaattttcagatcgaattgtgtcagtta
acttaactttcgacatcttaattcaggttaatttatcgattttccaggtaccaaaacgagaacgaagcgag
aatttggacaaacaacgactcgaagcacattcttctgagattttttcattcagtacgaaggcttgagtgggg
acgaatacctcacttgctccatcatcccttccaatttggaaaagtctgaatatgatctcagtacggattgg
acagttggannnnnnnnnnnacgcagcgtctaaagttcacggtcgcttatttttgcggagttggagagtgcg
ctccgttgagaatcgttgaagtttgcactgaatttcaatggaaaaaatttgtcaagttctgcaggttctta
cgaaattttcagaacatttttttcaacattttcaaaacaaccttgttcgaatattcatcgaaatttgataa
actttatttttatgttagatacaaagccgttggtttctgttagttttaggccctgtgtttcctgactgtttt
attttgcagcatgtatttatgtgttatgtattactatgattttgaatttttcagtgacttccagtcacaAT
G

Appendix H. Punnett square tables of selfed *let-56*, *let-366* and the controls (BC1270 and BC2200).

BC1270*: unc-22, unc-31;+/ nT1(IV), nT1(V)*

	***unc-22, unc-31;* +**	***unc-22, unc-31; nT1*(V)**	***nT1*(IV); +**	***nT1(IV); nT1(V)***
***unc-22, unc-31;* +**	*unc-22, unc-31;+/ unc-22, unc-31;+* **(Twitchers)**	*unc-22, unc-31;+ / unc-22, unc-31;nT1*(V) **(Dead)**	*unc-22, unc-31;+/ nT1*(IV), + **(Dead)**	*unc-22, unc-31;+/ nT1*(IV), *nT1*(V) **(Wild type)**
***unc-22, unc-31; nT1*(V)**	*unc-22, unc-31;nT1*(V)/ *unc-22, unc-31;+* **(Dead)**	*unc-22, unc-31,nT1*(V)/ *unc-22, unc-31, nT1*(V) **(Dead)**	*unc-22, unc-31, nT1*(V)/ *nT1*(IV), + **(Wild type)**	*unc-22, unc-31;nT1*(V)/ *nT1(IV), nT1(V)* **(Dead)**
***nT1*(IV); +**	*nT1*(IV), +/ *unc-22; unc-31,+* **(Dead)**	*nT1*(IV), +/ *unc-22, unc-31; nT1*(V) **(Wild type)**	*nT1*(IV), +/ *nT1*(IV), + **(Dead)**	*nT1*(IV), +/ *nT1(IV), nT1(V)* **(Dead)**
nT1(IV), nT1(V)	*nT1(IV), nT1(V)/ unc-22, unc-31,+* **(Wild type)**	*nT1(IV), nT1(V)/ unc-22, unc-31, nT1*(V) **(Dead)**	*nT1(IV), nT1(V)/ nT1*(IV), + **(Dead)**	*nT1(IV), nT1(V)/ nT1(IV), nT1(V)* **(Sick)**

let-65 (s1222 and *s1777) :let-65, unc-22, unc-31;+/ nT1(IV), nT1(V)*

	***let-65, unc-22; unc-31;*+**	***let-65; unc-22; unc-31;nT1*(V)**	***nT1*(IV), +**	***nT1(IV), nT1(V)***
let-65, unc-22; unc-31, +	*let-65, unc-22; unc-31, +/* *let-65, unc-22; unc-31, +* **(Larval lethal)**	*let-65, unc-22, unc-31;+ /* *let-65, unc-22, unc-31;nT1*(V) ***(Dead)***	*let-65, unc-22, unc-31;+/* *nT1*(IV), + **(Dead)**	*let-65, unc-22, unc-31;+/* *nT1*(IV), *nT1*(V) **(Wild type)**
let-65, unc-22; unc-31, nT1(V)	*let-65, unc-22; unc-31, nT1*(V)/ *let-65, unc-22, unc-31;+* **(Dead)**	*let-65, unc-22, unc-31; nT1*(V)/ *let-65, unc-22, unc-31;nT1*(V) **(Dead)**	*let-65, unc-22, unc-31;nT1*(V)/ *nT1*(IV), + **(Wild type)**	*let-65, unc-22, unc-31; nT1*(V)/ *nT1(IV), nT1(V)* **(Dead)**
***nT1(IV),* +**	*nT1*(IV), +/ *let-65, unc-22, unc-31;+* **(Dead)**	*nT1*(IV), +/ *let-65, unc-22; unc-31;nT1*(V) **(Wild type)**	*nT1*(IV), +/ *nT1*(IV), + **(Dead)**	*nT1*(IV), +/ *nT1(IV), nT1(V)* **(Dead)**
nT1(IV), nT1(V)	*nT1(IV), nT1(V)/ let-65, unc-22, unc-31;+* **(Wild type)**	*nT1(IV), nT1(V)/ let-65, unc-22, unc-31;nT1*(V) **(Dead)**	*nT1(IV), nT1(V)/ nT1*(IV), + **(Dead)**	*nT1(IV), nT1(V)/ nT1(IV), nT1(V)* **(Sick)**

BC2200*: dpy-18/ eT1(III); unc-46/ eT1(V)*

	dpy-18; unc-46	***dpy-18; eT1*(V)**	***eT1*(III)*; unc-46***	***eT1*(III); *eT1*(V)**
dpy-18; unc-46	*dpy-18; unc-46/ dpy-18; unc-46* **(dpy, unc)**	*dpy-18; unc-46/ dpy-18; eT1*(V) **(Dead)**	*dpy-18; unc-46/ eT1*(III)*; unc-46* **(Dead)**	*dpy-18; unc-46/ eT1*(III); *eT1*(V) **(Wild type)**
***dpy-18; eT1*(V)**	*dpy-18; eT1*(V)/ *dpy-18; unc-46* **(Dead)**	*dpy-18; eT1*(V)/ *dpy-18; eT1*(V) **(Dead)**	*dpy-18; eT1*(V)/ *eT1*(III); *unc-46* **(Wild type)**	*dpy-18; eT1*(V)/ *eT1*(III); *eT1*(V) **(Dead)**
eT1*(III); *unc-46	*eT1*(III); *unc-46/ dpy-18; unc-46* **(Dead)**	*eT1*(III); *unc-46/ dpy-18; eT1*(V) **(Wild type)**	*eT1*(III); *unc-46/ eT1*(III); *unc-46* **(Dead)**	*eT1*(III); *unc-46/ eT1*(III); *eT1*(V) **(Dead)**
***eT1*(III); *eT1*(V)**	*eT1*(III); *eT1*(V)/ *dpy-18; unc-46* **(Wild type)**	*eT1*(III); *eT1*(V)/ *dpy-18; eT1*(V) **(Dead)**	*eT1*(III); *eT1*(V)/ *eT1*(III); *unc-46* **(Dead)**	*eT1*(III); *eT1*(V)/ *eT1*(III); *eT1*(V) **(unc-36)**

let-336(s1413 and *s1495): dpy-18 / eT1(III); let336, unc-46/ eT1(V)*

	dpy-18; let336, unc-46	***dpy-18; eT1*(V)**	***eT1*(III); *let-336, unc-46***	***eT1*(III); *eT1*(V)**
dpy-18; let336, unc-46	*dpy-18; let336, unc-46/* *dpy-18; let336, unc-46* **(Dead)**	*dpy-18; let336, unc-46/* *dpy-18; eT1*(V) **(Dead)**	*dpy-18; let336, unc-46/* *eT1*(III); *let-336, unc-46* **(Dead)**	*dpy-18; let336, unc-46/* *eT1*(III); *eT1*(V) **(Wild type)**
***dpy-18; eT1*(V)**	*dpy-18; eT1*(V)/ *dpy-18; let336; unc-46* **(Dead)**	*dpy-18; eT1*(V)/ *dpy-18; eT1*(V) **(Dead)**	*dpy-18; eT1*(V)/ *eT*(III); *let-336, unc-46* **(Wild type)**	*dpy-18; eT1*(V)/ *eT*(III); *eT1*(V) **(Dead)**
eT1*(III); *let-336 unc-46	*eT1*(III); *let-336, unc-46/* *dpy-18; let336, unc-46* **(Dead)**	*eT1*(III); *let-336, unc-46/* *dpy-18; eT1*(V) **(Wild type)**	*eT1*(III); *let-336, unc-46/* *eT1*(III); *let-336, unc-46* **(Dead)**	*eT1*(III); *let-336, unc-46/* *eT*(III); *eT1*(V) **(Dead)**
***eT1*(III); *eT1*(V)**	*eT1*(III); *eT1*(V)/ *dpy-18; let336; unc-46* **(Wild type)**	*eT1*(III); *eT1*(V)/ *dpy-18; eT1*(V) **(Dead)**	*eT1*(III); *eT1*(V)/ *eT1*(III); *let-336, unc-46* **(Dead)**	*eT1*(III); *eT1*(V)/ *eT1*(III); *eT1*(V) **(Unc-36)**

REFERENCE LIST

Akins, R. A., and A. M. Lambowitz, 1987 A protein required for splicinggroup I intron in Neurospora mitochondria is mitochondrial tyrosyl-tRNA synthetase or a derivative thereof. Cell. **50(3):** 331-345.

Altschul, S. F., J. C. Wootton, E. M. Gertz, R. Agarwala, A. Morgulis, A. A. Schäffer, and Y. K. Yu, 2005 Protein database searches using compositionally adjusted substitution matrices. FEBS J. **272:** 5101-5109.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and. D. J. Lipman, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25(17):** 3389-402.

Anderson, P., 1995 Mutagenesis. Methods cell biol. **48:** 31-58.

Blackwell, T. K. and A. K. Walker, 2006 Transcription mechanisms In *WormBook*, ed. The *C. elegans* Research Community, doi/10.1895/wormbook.1.121.1, http://www.wormbook.org.

Boulin, T., J. F. Etchberger and O. Hobert, 2006 Reporter gene fusions In *WormBook*, ed. The *C. elegans* Research Community, doi/10.1895/wormbook.1.106.1, http://www.wormbook.org."

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

Brodersen, D. and P. Nissen, 2005 The social life of ribosomal proteins. FEBS J**. 272:** 2098-2108.

Brown, C. T., Y. Xie, E. H. Davidson, and R. A. Cameron, 2005 Paircomp FamilyRelationsII and Cartwheel: tools for interspecific sequence comparison. BMC Bioinformatics. **6:** 70.

Chalfie, M., Y. Tu, G. Euskirchen, W. Ward, and D. Prasher, 1994 Green fluorescent protein as a marker for gene expression. Science. **263(5148):** 802-805.

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

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

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

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

Deniziak, M. A. and J. Barciszewski, 2001 Methionyl-tRNA synthetase. Acta Biochim Pol. **48(2):** 337-50.

Eriani, G., M. Delarue, O. Poch, J. Gangloff, and D. Moras, 1990 Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature. **347:** 203-306.

Ferguson, E. L., and H. R. Horvitz, 1985 Identification and characterization of 22 genes that affect the vulval cell lineages of *Carnorhabditis elegans*. Genetics. **91:** 491-520.

Fisk, D., C. Ball, K. Dolinski, S. Engel, E. Hong, L. Issel-Tarver, K. Schwartz, A. Sethuraman, D. Botstein, and J. Cherry, 2006 The *Saccharomyces cerevisiae* S288C genome annotation: a working hypothesis. Yeast **23:** 857–865.

Gabius, H. J., G. Graupner, and F. Cramer, 1983 Activity Patterns of aminoacyl-tRNA synthetases, tRNA methylases, arginyltransferase and tubulin: tyrosine ligase during development and ageing of *Caenorhabditis elegans*. Eur J Biochem. **131:** 231-234.

Hansen, M., S. Taubert, D. Crawford, N. Libina, S. Lee, and C. Kenyon, 2007 Lifespan extension by conditions that inhibit translation in *Caenorhabditis elegans*. Aging Cell. **6:** 95–110.

Herrmann, J. M. and W. Neupert, 2000 What fuels polypeptide translocation? An energetical view on mitochondrial protein sorting. Biochimica et Biophysica Acta. **1459:** 331-338.

Hobert, O., 2002 PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans*. Biotechniques. **32(4):** 728-730.

Horton, P., K. J. Park, T. Obayashi, N. Fujita, H. Harada, C. J. Adams-Collier, and K. Nakai, 2007 WoLF PSORT: protein localization predictor. Nucleic Acids Res. **35 (Web Server issue):** 585-587.

Hunt-Newbury, R., R. Viveiros, R. Johnsen, A. Mah, D. Anastas, L. Fang, E. Halfnight, D. Lee, J. Lin, A. Lorch, S. McKay, H. Okada, J. Pan, A. Schulz, D. Tu, K. Wong, Z. Zhao, A. Alexeyenko, T. Burglin, E. Sonnhammer, R. Schnabel, S. Jones, M. Marra, D. Baillie, and D. Moerman, 2007 High-throughput in vivo analysis of gene expression in *Caenorhabditis elegans*. PLoS Biol. **5(9):** e237.

Jakubowski, H., 2001 Translation Accuracy of Aminoacyl-tRNA Synthetases Implications for Atherosclerosis. American Scociety for Nutritional Sciences. **131:** 2983S-2987S.

Johnsen, R., 1990 Genetic analysis of the left half of linkage group V in Caenorhabditid elegans. PhD Thesis, Simon Fraser University, Burnaby, B.C., Canada.

Johnsen, R.C., D.L. Baillie (1991) Genetic Analysis of a Major Segment [LGV(left)] of the Genome of Caenorhabditis elegans. Genetics. **129:** 735-752.

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

Jones, M. R., S. J. Maydan, S. Flibotte, D. Moerman, and D. L. Baillie, 2007 Oligonucleotide Array Comparative Genomic Hybridization (oaCGH) based characterization of genetic deficiencies as an aid to gene mapping in *Caenorhabditis elegans. BMC Genomics*. **8**:402.

Kamath, R. S., A. G. Fraser, Y. Dong, G. Poulin, R. Durbin, M. Gotta, A. Kanapin, N. Le Bot, S. Moreno, M. Sohramann, D. P. Welchman, P. Zipperlenm, and J. Ahringer 2003 Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. Nature. **421:** 231-237.

Kemphues K. 2005 December 24 Essential Genes In *WormBook*, ed. The *C. elegans* Research Community, doi/10.1895/wormbook.1.57.1, http://www.wormbook.org.

Kiriyama, Y., M. Kubota, T. Takimoto, T. Kitoh, A. Tanizawa, Y. Akiyama, and H. Mikawa, 1989 Biochemical characterization of U937 cells resistant to L-asparaginase: the role of asparagine synthetase. Leukemia, **3(4):** 294-297.

Lambertsson, A., 1998 The Minute genes in Drosophila and their molecular functions. Adv Genet. **38:** 69-134.

Lee, D., 2009 Molecular identification of *let-56* and characterization of its regulatory elements in *Caenorhabditis elegans*. M.Sc Thesis, Simon Fraser University, Burnaby, B.C., Canada.

Lee, S. S., Lee, R. Y. N., Freaser, A. G., Kamath, R. S., Ahringer, J., and Ruvkun, G. 2002 A systematic RNAi screen identifies a critical role for mitochondria in *C. elegans* longevity.

Lee, S. W., B. H. Cho, S. G. Park, and S. Kim, 2004 Aminoacyl-tRNA synthetase complexes: beyond translation. Journal of Cell Science. **117:** 3725-3734.

Lingappa, V. R., J. R. Lingappa, and G. Blobel, 1979 Chicken ovalbumin contains an internal signal sequence. Nature. **281:** 117- 121.

Maeda, I., Y. Kohara, M. Yamamoto, and A. Sugimoto, 2001 Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. Curr Biol. **11:**171-6.

Marchler-Bauer, A., J. Anderson, F. Chitsaz, M. Derbyshire, C. DeWeese-Scott, J. Fong, L. Geer, R. Geer, N. Gonzales, M. Gwadz, S. He, D. Hurwitz, J. Jackson, Z. Ke, C. Lanczycki, C. Liebert, C. Liu, F. Lu, S. Lu, G. Marchler, M. Mullokandov, J. Song, A. Tasneem, N. Thanki, R. Yamashita, D. Zhang, N. Zhang, and S. Bryant, 2009 CDD: specific functional annotation with the Conserved Domain Databas. Nucleic Acids Res. **37:** 205-210.

Marra, M. A. and D. L. Baillie, 1994 Recovery of dupications by drug resistance selection in *Caenorhabditis elegans*. Canadian Journal of Genome. **37:** 701-705.

Marygold, S., J. Roote, G. Reuter, A. Lambertsson, M. Ashburner, G. H. Millburn, M. P. Harrison, Z. Yu, N. Kenmochi, T. C. Kaufman, S. J. Leevers, and K. R. Cook, 2007 The ribosomal protein genes and *Minute* loci of *Drosophila melanogaster*. Genome Biology. **8(216):** 1-26.

McGhee, J. D., and M. W. Krause, 1997 Transcription Factors and Transcriptional Regulation. In: Riddle, D. L., T. Blumenthal, B. J. Meyer, and J. R. Priess, (Eds.), *C. elegans* II (pp. 147–184). NY: C Cold Spring Harbor.

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

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

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

Ochsner, U. A., X. Sun, T. Jarvis, I. Critchley, and N. Janjic, 2007 Aminoacyl-tRNA synthetases: essential and still promising targets for new anti-infective agents. Expert Opin. Investig Drugs. **16(5):**573-593.

O'Donoghue, P. and L. Luthey-Schulten, 2003 On the Evolution of Structure in Aminoacyl-tRNA Synthetase. Microbiology and Molecular Biology Reviews. **67(4):** 550-573.

Paetzel, M., A. Karla, N. C. J. Strynadka, and R. E. Dalbey, 2002 Signal Peptidases. Chem. Rev. **102:** 4549-4579.

Palade, G. and P. Siekevitz, 1956 Liver microsomes: an integrated morphological and biochemical study. J. Biophys. Biochem. Cytol. **2:** 171-201.

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

Rogalski, T. M. and D. L. Baillie, 1985 Genetic organization of the *unc-22* IV gene and the adjacent region in *Caenorhabditis elegans*. Mol. Gen. Genet. **201 (3):** 409-414.

Rozen, S and H. J. Skaletsky, 2000 Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (Eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology (pp 365-386). Totowa, NJ: Humana Press.

Sanni, A., M. Mirandes, J. Ebel, Y. Boulanger, J. Wallerg, and F. Fasiolo, 1988 Structure and expression of the genes encoding the alpha and beta subunits of yeast phenylalanyl-tRNA synthetase. The journal of Biological chemistry. **263(30):** 15407-15415.

Schein, J. E. 1994 Deficiency mapping and correlation of the genetic and physical maps in the u*nc-22(IV)* region of *Caenorhabditis elegans*. M.Sc Thesis, Simon Fraser University, Burnaby, B.C., Canada.

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

Schultz, J. 1929 The Minute reaction in the development of *Drosophila melanogaster*. Genetics. **14:** 366-419.

Simmer, F., C. Moorman, A. M. van der Linden, E. Kuijk, P. V. van den Berghe, R. S. Kamath, A. G. Fraser, J. Ahringer, and R. H. Plasterk, 2003 Genome wide RNAi of *C. elegans* using the hypersensitive rrf-3 strain reveals novel gene functions. PLoS Biol. **1:** 77-84.

Sönnichsen, B., L. B. Koski, A. Walsh, P. Marschall, B. Neumann, M. Brehm, A. M. Alleaume, J. Artelt, P., Bettencourt, E. Cassin, M. Hewitson, C. Holz, M. Khan, S. Lazik, C. Martin, B. Nitzsche, M. Ruer, J. Stamford, M. Winzi, R. Heinkel, M.

Röder, J. Finell, H. Häntsch, S. J. Jones, M. Jones, F. Piano, K. C. Gunsalus, K. Oegema, P. Gönczy, A. Coulson, A. A. Hyman, and C. J. Echeverri, 2005 Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans.* Nature. **434 (7032):** 462-9.

Vatcher, G. P., C. M. Thacker, T. Kaletta, H. Schnabel, R. Schnabel, and D. L. Baillie, 1988 Serine Hydroxymethyltransferase Is Maternally Essential in *Caenorhabditis elegans*. J Biol Chem. **273(11):** 6066-73.