FUNCTIONAL ANALYSIS OF ATP BINDING CASSETTE (ABC) TRANSPORTERS IN *CAENORHABDITIS ELEGANS*

by

Zhongying Zhao
B.Sc., Anhui Normal University, 1993
M.Sc., Beijing Normal University, 1996

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

DOCTOR OF PHILOSOPHY

In the
Department
of
Molecular Biology and Biochemistry

SIMON FRASER UNIVERSITY

Spring 2005

APPROVAL

Name: **Zhongying Zhao**

Degree: **Doctor of Philosophy**

Title of Thesis: **Functional Analysis of ATP Binding Cassette (ABC) transporters in *C. elegans***

Examining Committee:

Chair: Dr. Carl Lowenberger
Associate Professor, Department of Biology

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

Supervisor
Dr. Victor Ling, Professor
UBC / British Columbia Cancer Research Centre

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

Supervisor
Dr. Fiona S. L. Brinkman, Assistant Professor
Department of Molecular Biology and Biochemistry

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

External Examiner
Dr. Michael C. Dean, Professor
National Cancer Institute at Frederick, MD, USA

Date Defended/Approved: **April 12, 2005**

SIMON FRASER UNIVERSITY

PARTIAL COPYRIGHT LICENCE

ABSTRACT

This thesis deals with functional analysis of ABC transporter genes in *Caenorhabditis elegans*. ABC transporters constitute one of the largest gene families in *C. elegans*. Sixty ABC genes have been identified and classified into eight different subfamilies. Fifty-seven orthologous pairs can be found between *C. elegans* and *C. briggsae*. As an initial step toward understanding the functions of the ABC genes, I generated transgenic animals containing promoter::GFP/DsRed fusion constructs for the whole family in order to address when and where these genes are turned on *in vivo*. I observed frequent expression of ABC genes in intestine, pharynx and excretory cell, especially for those in subfamilies B and C. Sixteen of the 60 ABC genes are tandemly duplicated genes, forming two 4-gene and four 2-gene clusters. Interestingly, the majority of the promoters from the same clusters drive temporally and/or spatially differential expressions, suggesting active function of these duplicated genes in different tissues or stages rather than their being inactivated, as is typical for duplicated genes. Currently 18 single ABC gene mutants have been created. All of them are wild type under normal laboratory conditions. I used one of the ABC genes, *pgp-12*, an excretory cell-expressing gene, to characterize its transcriptional regulation. A 10 base pair (bp) *cis*-element, Ex-1 and its binding protein C26C6.5a, were isolated. C26C6.5a is also expressed in the excretory cell and binds Ex-1 *in vitro*. RNAi analysis revealed that the C26C6.5a protein functions as a *trans*-activator for excretory cell specific expression. The only member of subfamily E is the ABCE gene that has been annotated as an RNase L inhibitor. This annotation is unlikely to be correct because of the lack of the inhibitor gene in *C. elegans*. Functional analysis of the gene reveals that it is essential for life, possibly involved in transcription and translation, and may function as a nucleocytoplasmic transporter. RNAi data implies that the protein functions in both germline and vulvae.

I would like to dedicate this thesis to my daughter,

Sarah F. Zhao, who has been the comfort

of her mamma and the joy of her father.

ACKNOWLEDGEMENTS

I would like to thank my senior supervisor, Dr. David L. Baillie for his enthusiasm, encouragement, patience and guidance in the course of my research and writing. I would also like to thank the members of my supervisory committee, Drs. Victor Ling, Nick Harden and Fiona Brinkman for their helpful advice, especially during the preparation of this thesis. I also especially thank Drs. Ann Rose, Andy Fire, Yuji Kohara and Robert Barstead for their generosity in providing me with reagents. I am also indebted to the *C. elegans* Genetics Centre and the *C. elegans* Knockout Consortium for the strains they provided. Special thanks to Drs. Robert Johnsen, Jonathan Sheps, and Mrs Christine Beauchamp for their help in the preparation of this thesis. Many thanks to Drs. Wanyuan Ao, John Tyson and Chris Beh for assistance with the yeast one-hybrid screen and Mr. Weiping Shen in Dr. Pio's laboratory for help with the gel shift assay as well as Mrs Rebecca Newbury for photograph work. The support I received from the Baillie's lab, especially Dr. Lynnette Kuervers, Martin Jones, Allan Mah and Carrie Jones was invaluable, and I greatly appreciate their assistance during my research. Finally, I would like to acknowledge the wonderful support from my friends and family, especially my wife, Li Fang.

This work was financially supported by NSERC Canada, Genome British Columbia and Genome Canada.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVATIONS

ABC	ATP Binding Cassette
ABCE	Subfamily E protein of ABC transporter family
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
cDNA	Complementary DNA
CFTR	Cystic fibrosis transmembrane conductance regulator
DNA	Deoxyribonucleic acid
EMS	Ethyl Methane Sulfonate
EMSA	Electrophoresis Mobility Shift Assay
EST	Expressed Sequence Tag
Ex-1	Excretory cell specific 10 bp *cis*-element within *pgp-12* promoter
Ex-L	10 bp *cis*-element to the left of Ex-1 within *pgp-12* promoter
Ex-R	10 bp *cis*-element to the right of Ex-1 within *pgp-12* promoter
Ex-R1	10 bp *cis*-element to the right of Ex-R within *pgp-12* promoter
GFP	Green Fluorescence Protein
IFN	Interferon
IPTG	Isopropyl-β-D-thiogalactoside
kb	Kilo base pairs
MDR	Multidrug Resistance
MRP	Multidrug-resistance related protein
NBD	Nucleotide Binding Domain
OABP	Oligoadenylate binding protein
ORF	Open reading frame
PCR	Polymerase Chain Reaction
PFIC	Progressive familial intrahepatic cholestasis
PGP	P-glycoprotein
RFP	Red Fluorescence Protein/DsRed
RNA	Ribonucleic Acid
RNAi	RNA interference
RNAse L	Ribonuclease L
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SAGE	Serial Analysis of Gene Expression
SL	Splicing leader
TMD	TransMembrane Domain
UTR	Untranslated Region

CHAPTER 1 GENERAL INTRODUCTION

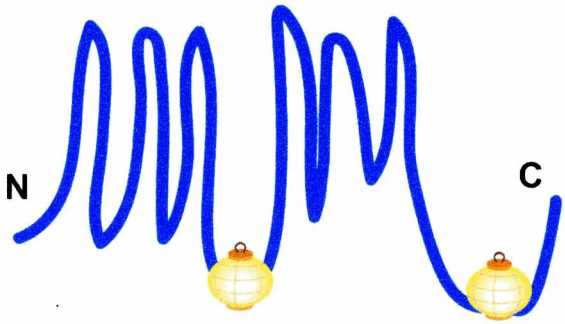
In this thesis, I undertook a functional analysis of the ATP Binding Cassette (ABC) transporter family in *C. elegans*. This family contains members that may not be transporters (ABCE and F subfamilies), but for convenience, I will refer to the gene family as transporters for ease of reading. In this introduction, I will provide an overview of the ABC transporter family and the reasons for choosing *C. elegans* as a model system to investigate ABC transporter genes.

1.1 Overview

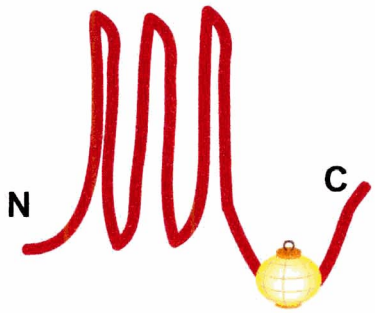
The ATP Binding Cassette (ABC) transporter family comprises one of the largest protein families found in all phyla. Members of the family are involved in transport of various types of substrates across biological membranes. A typical ABC transporter consists of at least one evolutionarily conserved ABC, also known as NBD (Nucleotide Binding Domain) comprising approximately 215 amino acids, as well as a TMD (Trans-Membrane Domain) containing several predicted multi-spanning α-helical segments. An NBD usually contains a Walker A and Walker B motif which are also found in other nucleotide binding proteins and a signature (C) motif, located just upstream of the Walker B site (Runswick and Walker, 1983; Hyde et al, 1990). The C motif usually contains the consensus sequence LSGGQK, which is diagnostic of ABC transporters and distinguishes them from other Walker ATPases. The NBD and TMD are normally arranged in a duplicated forward $(TMD\text{-}NBD)_2$ or reverse $(NBD\text{-}TMD)_2$ configurations, although other

arrangements exist (Higgins 1992 &1995) (Figure1). Many ABC transporters are organized as either full transporters containing two TMDs and two NBDs or as half transporters containing one of each domain, or occasionally only as a solitary NBD (Decottignies and Goffeau, 1997). The half transporters usually become functional by forming a homodimer or heterodimer. ABC proteins have been found in membranes of all cellular compartments except nuclear membranes (Bauer et al, 1999). ABC transporters are grouped into several structural classes, or subfamilies, on the basis of amino acid sequence similarity and domain organization (Croop, 1998). Human ABC transporters are systemically classified into seven subfamilies, i.e., ABCA (ABC1), ABCB (MDR/TAP), ABCC (MRP/CFTR), ABCD (ALD), ABCE (RNAse LI/OABP), ABCF (GCN20) and ABCG (White). Each member of the subfamilies is numbered in consecutive order within the same subfamily with some exceptions. For instance, ABCB7 is also called ABC7 because Savary et al (1997) first isolated the mouse orthologous cDNA and designated it as ABC7 in mouse. These ABC proteins without any TMD may not work as intrinsic membrane transporters, but as regulators for protein translation or gene transcription. For example, members of ABCE and ABCF subfamily contain a pair of NBDs without apparent associated TMDs. Functional analysis of the ABCE protein suggest it has roles in protein translation and gene transcription (Kerr, 2004; Dong et al, 2004; Zhao, et al., 2004a (chapter 3 of the thesis)). One of the human ABCF proteins, ABC50 is involved in promotion of the association of aminoacyl-tRNA with eIF2 (Tyzack et al, 2000) while its yeast homolog GCN20 is involved in activation of eIF2 kinase (Marton et al, 1997).

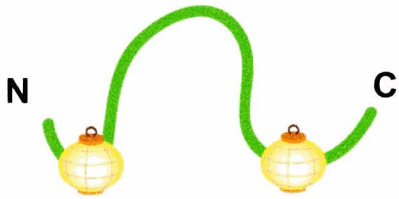
Figure 1. Schematic representation of domain organization for ABC transporters. The transmembrane domain (TMD) consists of six or more predicted transmembrane helices. NBD is represented by a cytoplasmic loop (paper lantern) comprising the ATP binding domain (Walker A and B motifs and the ABC signature motif). Transmembrane and/or linking sequences are differentially coloured for each configuration. A: full-size ABC transporter with TMDs at the amino terminal of NBDs; B and D: half-size ABC transporters consisting of a single TMD and NBD in opposite orientations; C and E: ABC transporters with one or two NBD but no TMD (modified from Decottignies and Goffeau, 1997).



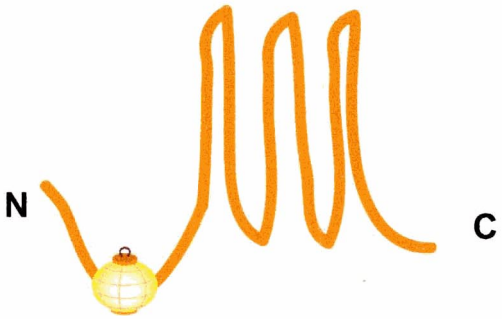
A: (NBD-TMD)2



B: NBD-TMD



C: NBD-NBD



D: NBD-TMD

E: NBD

Since the first ABC gene, MDR1, was cloned and characterized through its capability to confer multidrug resistance to cultured cell lines (Juliano and Ling, 1976; Riordan et al, 1985; Ueda, et al, 1986), ABC genes have turned out to be one of the biggest protein families and well conserved across species. ABC transporters constitute at least 5% of all proteins encoded by the *Escherichia coli* and *Bacillus subtilis* genomes (Blattner et al, 1997), and appear to perform functions affecting processes as diverse as cell division, regulation of translation, elongation of polypeptides, determination of cell volumes (Higgins 199s2; Kerr 2004) in addition to import or export of a wide range of low and a few high molecular weight solutes. ABC transporters are primarily involved in the import of various sugars, vitamins and amino acids into the cell in bacteria. In eukaryotes, many ABC transporters are involved in export compounds across cytoplasm membrane or into an intracellular compartment such as endoplasmic reticulum or mitochondria. No eukaryotic ABC transporter has been found to be involved in import of compounds from outside the cell into its cytoplasm (Saurin W, et al., 1999).

The human genome encodes 48 ABC transporters, many of which are associated with diseases when they are defective (Dean et al., 2001). For example, mutation in ABCC7 leads to cystic fibrosis, which is more common in the Caucasian population than in Africans or Asians (Riordan et al, 1989). ABCB4 transports phosphatidylcholine across the canalicular membrane of hepatocytes (van Helvoort et al, 1996). Mutations in this gene cause progressive familial intrahepatic cholestasis (PFIC) (Deleuze et al, 1996) and are associated with intrahepatic cholestasis of pregnancy (Dixon et al, 2000). Mutations in the ABCA4 gene have been associated with multiple eye disorders (Allikmets, 2000). An analysis of Celera's mouse genome revealed 52 ABC genes in total (Dean 2002). Most of

these genes have orthologs in humans with a few exceptions, such as duplications of ABCA8, ABCB1 and ABCG2 as well as a loss of *ABCA10* in mouse. In total, there are 56 ABC transporter genes in the fruit fly (*Drosophila melanogaster)* genome with at least one representative of each of the known mammalian subfamilies (Dean, et al., 2002). The most intensively investigated ABC gene in the fruit fly is the white gene (CG2759). This gene encodes a protein that transports precursors of eye pigments (guanine and tryptophan) in eye cells (Chen et al. 1996). *Arabidopsis thaliana* contains a total of 129 open reading frames (ORFs) capable of encoding ABC proteins, of which 103 possessed contiguous transmembrane domains and were identified as putative transmembrane proteins. *Arabidopsis* is unusual in its large allocation of ORFs (a minimum of 0.5%) being members of the ABC transporter family (Sanchez-Fernandez et al., 2001). It is possible that the expansion of the ABC transporter family in *Arabidopsis* is related to its unique physiology (plant) compared to other eukaryotic organisms (mostly animals) studied so far. Annotation of the sequence of *Saccharomyces cerevisiae* (yeast) genome predicts the existence of 29 proteins, belonging to the ABC transporter family, which have been classified into six clusters using phylogenetic analysis and detection of conserved amino acid residues (Decottignies and Goffeau, 1997). It has also been proposed that there are 31 distinct ABC proteins in the yeast (Taglicht and Michaelis, 1998). This discrepancy may be the result of different criteria used in defining ABC membership (Sheps, et al, 2004). There are many newly identified human ABC transporters whose functions are not clear. Knowledge of their homologs from other organisms, such as *C. elegans*, would help decipher the functions of these human equivalents. Systematic investigation of the whole ABC family has become realistic with the complete inventory of all members in the family. Functional analysis of the whole

gene family will provide insights into individual ABC protein function within particular subfamilies and the evolutionary dynamics of the whole family. For a rapid and comprehensive investigation of large gene families, such as the ABC transporter family, it will be more efficient and convenient to investigate the family using a well-characterized model organism.

1.2 *C. elegans* is an excellent model to investigate ABC transporters

Studies of eukaryotic ABC transporters have primarily been conducted in mammalian systems such as mouse or cultured cell lines. The major drawbacks to studying ABC transporters in mouse are the complexity of the organism and the lack of efficient tools such as those available for *C. elegans* (see below). There are intrinsic disadvantages to investigation of ABC transporters *in vitro* because the knowledge of ABC gene function derived from these investigations may not necessarily reflect its *in vivo* function. In order to investigate systematically the roles of the ABC transporter gene family *in vivo*, I have chosen *C. elegans* as a model system. A major reason for this choice is the simplicity of the multicellular organism. It has a small and transparent body (about 1 mm in length for adult) and thus it is feasible to keep track of gene activity *in vivo* using a variety of reporters such as green fluorescence protein (GFP). *C. elegans*, a multicellular organism, undergoes a complex developmental process, including embryogenesis, tissue morphogenesis and differentiation. Thus, principles of gene functions that are obtained from *C. elegans* may be applicable to other multicellular complex organisms, such as humans, which are somewhat more refractory to laboratory investigation. Other advantages such as short generation time (about 3.5 days at 20°C), hermaphroditism and possibility for long-term storage in liquid nitrogen make *C. elegans*

an ideal model for both forward and reverse genetic analysis. In addition, the complete cell lineage has been worked out (Sulston, et al., 1983), making it possible to accurately keep track of ABC activities within individual cells. RNA interference (RNAi), originally discovered in *C. elegans*, constitutes a powerful means to explore gene functions *in vivo* (Fire, et al., 1998). It is easy to temporarily knockdown a gene function simply by feeding worms with bacteria that expresses double-stranded RNA which is homologous to the gene of interest (Kamath, et al, 2001). Availability of such bacteria library for the whole worm genome makes it practical for high-throughput customized phenotype screening. A major advantage for using *C. elegans* as a model system is that genome sequencing of several closely related nematode species has been completed or nearly completed. Cross-species comparison will greatly improve our understanding of genome dynamics such as gene duplication, gene loss and gene family evolution. This will be very useful for refining gene predictions. For example, 1275 new genes were predicted as a result of a comparison between the *C. elegans* and *C. briggsae* genomes (Stein et al, 2003). Most importantly, the regulatory sequences are usually conserved across related species. Thus, comparison of regulatory sequences among multiple species will significantly facilitate the identification of *cis*-elements that are critical for control of temporal and/or spatial specific gene expression.

It is noteworthy that the ABC transporter family is expanded in *C. elegans*, especially the ABCB subfamily, by gene duplication (Sheps, et al., 2004; Zhao, et al, 2004b). Only a limited number of ABC genes have been functionally characterized. A systematic analysis of these ABC transporters will provide functional insights into the whole family. This will also shed light on evolutionary fates of the duplicated ABC

genes, many of which form tandem duplicates. Comparison of the ABC transporter families between *C. elegans* and *C. briggsae* will provide insights into evolutionary dynamics of the ABC gene family. Transcriptional regulation of ABC genes is poorly understood except for the human MDR1 gene. In order to identify alternative modulators for the multidrug resistance (MDR) protein, studies on transcriptional regulation for this gene has been carried out over the last decade with the majority of these assays have been performed *in vitro*. Tremendous progress has been made in understanding the regulatory network for human MDR1. However, rapid advances in the transcription field indicate that our understanding of MDR1 transcription may still be in its infancy, and dissection of other drug transporters lags far behind (Scotto, 2003). Given the complexity of the gene regulation network, as well as the fact that most of the knowledge regarding human MDR1 has been derived from *in vitro* assays, it is desirable to have a model system that allows *in vivo* study of transcriptional regulation of ABC genes, especially potential drug resistance ones. *C. elegans* was ideal as a model system for in vivo studies of regulation of gene expression. Investigation of transcriptional control of ABC gene expression has become practical because of the large expression dataset available, from which a subset of "co-expressed" genes can be pulled out for prediction and validation of regulatory elements. Transcriptional elements can be predicted or inferred by comparison of regulatory regions among *C. elegans, C. briggsae* and *C. remanei*. The regulatory sequences containing transcriptional elements can be dissected by a combination of PCR fusion and transgenic techniques (see below). Transgenic strains carrying appropriate markers like GFP or DsRed, enable visualization of gene activity in real time *in vivo*. This will eventually lead to the identification of *cis*-elements and their binding *trans*-activator that controls tissue-specific expression, such as exclusive expression of *pgp-12* in the

excretory cell. The ABC transporter family, as a whole, is highly conserved across all phyla. However, typical orthologous pairs between worm and human ABC transporters are not evident as expected, except for members of the ABCE and F subfamilies (Sheps et al., 2004). One example is the ABCE protein, the sole member of the ABCE subfamily, which has obvious orthologs in *D. melanogaster*, *Homo sapiens* and *Saccharomyces cerevisiae*. It is the only ABC protein found in the highly reduced genome of the endosymbiotic *Guilardia theta* (Douglas, et al., 2001), suggesting an essential biological function. However, the ABCE protein has been annotated as an RNase L inhibitor that was first identified in a mammalian antiviral pathway (Bisbal, et al., 1995). This annotation apparently does not make sense in *C. elegans* because RNase L is not found in its genome. Given the good conservation and clear orthlogy of the ABCE gene, it is well worth seeking the proper functions for this gene.

In this thesis, I will address four closely related projects in consecutive chapters.

In chapter 2, I address the evolutionary dynamics of the ABC transporter family between *C. elegans* and *C. briggsae*. I have identified 57 ABC transporter genes in *C. briggsae*. A comparative analysis of ABC transporters between the two species provides insights into evolution of gene family members, especially duplicated ones. Six inversions and seven deletions/insertions were found among the nearest neighbors of the 56 orthologous pairs. Transposase or reverse transcriptase insertions were found within two ABC gene clusters in *C. briggsae*, suggesting a role for transposition in the formation of gene clusters.

In chapter 3, I describe functional characterization of a well-conserved ABC protein, ABCE, i.e., Y39E4B.1. It has been annotated as an RNase L inhibitor in *C.*

elegans because in humans it associates with the RNase L and functions as its inhibitor (Bisbal, et al., 1995). All eukaryotic species show the ubiquitous presence and high degree of conservation of ABCE. However, RNase L is present only in mammals. This indicates that ABCEs may function not only as RNase L inhibitors, but also may have other functions. My preliminary data has shown that ABCE protein may be involved in the control of protein translation and gene transcription, work as a shuttle protein between cytoplasm and nucleus, and possibly as a nucleocytoplasmic transporter.

In chapter 4, I describe the expression analysis of ABC transporters in *C. elegans*. Eighteen ABC gene mutants have been created and a screen for phenotypes of these ABC mutants in the presence of chloroquine and colchicine will be described. As mentioned above, sixty ABC transporter genes have been identified in the *C. elegans* genome and classified into eight subfamilies (Sheps, et al., 2004). I have generated promoter::GFP fusions for the entire family in order to address when and where these genes are turned on *in vivo*. GFP expression was frequently observed in the intestine, excretory cell and pharynx, many of which are consistent with the phenotypes of several ABC transporter knockout mutants (Broeks, et al., 1995 &1996). It is striking that almost all tandemly duplicated ABC genes are expressed at different developmental stages or tissues, which strongly argues against the degeneration of paralogous ABC genes. Another significant observation is that the presence of splicing leader 2 (SL-2) in spliced RNA does not necessarily mean the gene is part of an operon. In other words, an internal promoter in an operon can be functionally independent from its upstream gene and regulatory sequences. Screens of some pgp mutants did not produce significantly increased sensitivity to drugs,

i.e., chloroquine and colchicine, suggesting these two compounds may not be the substrate for these pgp proteins.

In chapter 5, I focus on transcriptional control of gene expression in excretory cell driven by the *pgp-12* promoter. There is emerging evidence that expression of ABC transporters is highly regulated at the level of transcription, suggesting future targets for modulation of the multidrug resistance (MDR) (Scotto, 2003). A *cis*-element and its *trans*-acting binding factor have been isolated for *pgp-12* specific expression. Investigation of regulation of excretory cell specific expression will also help establish transcription network that governs formation of the tubular excretory cell, an excellent model for studying biological tube morphogenesis.

In chapter 6, I present a general discussion of what I have done in the whole thesis as well as the biological significance of my research to the ABC transporter field.

CHAPTER 2 COMPARISON OF ABC TRANSPORTERS BETWEEN *C. ELEGANS* AND *C. BRIGGSAE*

2.1 Abstract

ABC transporters constitute one of the largest gene families in *C. elegans*. Investigation of evolutionary dynamics for the gene family may provide better understanding of biological function of the family members. Given the completion of genome sequencing for both *C. elegans* and *C. briggsae*, I have conducted a comparative analysis of the whole ABC transporter family between the two species. These two worms are morphologically similar but diverged evolutionarily about 100 million years ago. Sixty ABC genes have previously been identified and classified into 8 different subfamilies in *C. elegans*. Annotation of *C. briggsae* ABC transporters has been done by *ab initio* gene prediction in combination with conserved gene order (synteny). The results show strong conservation of ABC transporters between the two species. Fifty-seven putative ABC transporters in *C. briggsae* have been identified, and only three of the *C. elegans* ABC genes do not have an obvious ortholog in *C. briggsae*. Six inversions and seven deletions/insertions have been found among the nearest neighbors of the 57 orthologous pairs. Orthology of ABC genes between the two species has in part been confirmed by the shared expression patterns. Transposase or reverse transcriptase insertions were found within two gene clusters in *C. briggsae,* suggesting a role for transposition in the formation of gene clusters. Compared with human and fruit fly, ABC transporters in *C. elegans* have undergone expansion, especially in subfamily B. Most of the duplicated ABC genes are found on the X chromosome. It remains unclear why local

duplications occur more frequently for ABCs on the sex chromosome than on autosomes in nematodes.

2.2 Introduction

ATP Binding Cassette (ABC) transporters constitute one of the largest protein families in both prokaryotes and eukaryotes. They are widely dispersed in the genomes and show a high degree of amino acid sequence identity among eukaryotes (Dean and Allikmets 1995). Comparative analysis of gene families between closely related species helps to develop an understanding of how gene families evolve. Additionally, this may provide insights into genome dynamics and improve gene prediction, especially for the identification of ambiguous regions such as first exons. Cross-species comparison of intergenic regions of orthologous gene pairs will facilitate identification of regulatory elements responsible for tissue- or stage-specific expression (Prasad and Baillie, 1989; Kennedy et al., 1993; Cui and Han, 2003). Orthologs describe genes in different species that derive from a common ancestor. Orthologous genes tend to have the same function but not necessarily. Paralogs describe genes duplicated within a single species that tend not to have the same functions. Both paralogs and orthologs are used to describe genes of one species in relation to those of other species and a part of homologs, which mean genes of common origin. Homologous genes share a arbitrary level of sequence similarity.

The sequencing of the *C. elegans* genome was the first of its kind for a multicellular organism (the *C. elegans* Sequencing Consortium, 1998). This allowed us to identify 60 ABC genes in the *C. elegans* genome (Sheps, et al., 2004). A comprehensive comparison of worm and yeast genomes revealed that 57% of genes in highly conserved

gene families could be found in orthologous pairs. The ABC transporter family is one of these conserved gene families. However, only 10% of worm ABC transporters were found in orthologous pairs when the comparison was made between worm and yeast genome. This was also the case between the worm and human ABC transporter families from which only 8 of 49 possible pairs (16%) of sister genes contained pair-wise orthologous ABC genes (Sheps, et al., 2004). This situation changes when comparison of ABC transporters between *C. elegans* and *C. briggsae* is made. *C. briggsae*, like *C. elegans* is a soil dwelling nematode that diverged from *C. elegans* approximately 100 million years ago (Coghlam and Wolfe, 2002; Stein et al, 2003). The two organisms are almost morphologically indistinguishable and follow similar developmental programs (Kirouac and Sternberg, 2003). Sequencing of the *C. briggsae* genome has recently been completed by the Genome Sequencing Center (GSC) at the Washington University at Saint Louis (http://genome.wustl.edu/projects/cbriggsae/) and the Sanger Center (http://www.sanger.ac.uk/Projects/C_briggsae/). This provides an excellent opportunity to conduct a comparative analysis of gene families between the two related species on a genome scale. It turns out that over 60% of *C. briggsae* genes have clear *C. elegans* orthologs. The two genomes exhibit extensive synteny (Stein, et al., 2003). I expected to find an even higher percentage of orthologous pairs among ABC genes, given the overall conservation of the ABC transporter family.

Here, I provide the complete set of ABC transporters as well as their classification in both *C. elegans* and *C. briggsae*. A comparison of the whole ABC transporter family between the two species has been made. Orthology and divergence of the family will be discussed.

2.3 Materials and methods

2.3.1 Identification of ABC genes in *C. elegans*

Identification of ABC genes in *C. elegans* was done basically as described previously (Sheps, et al., 2004). One or two (in the case where both half and full ABC transporters are present in the same subfamily) human ABC transporter protein sequence from each individual subfamily was used as query against WormPep by BLASTP with 0.01 as arbitrary expected threshold. A representative hit from each subfamily was used as query against WormPep for additional BLASTP search. This process was repeated until no new ABC proteins were found. Pooling the resulting hits and removing redundant ones made a list of ABC candidate genes. These candidate ABC amino sequences were subject to Pfam scan (Bateman, et al., 2002). Those with the presence of an obvious ABC domain were deemed putative ABC genes. Another BLASTP search was done for each of these ABC domain-containing sequences against the NCBI data set. Only those showing a significant match to other existing ABC proteins were retained as probable ABC genes. Specifically, when a putative *C. elegans* ABC protein sequence was used as a query to seach against NCBI non-redundant protein database by BLASTP, at least top ten hits come out as obvious annotated ABC transporters in other species, which I refer to "Significant match". This led to the identification of 60 ABC genes in *C. elegans*. They all contain at least one putative ABC domain. Some of them were not annotated as ABC transporters in the database but in each case BLASTP search showed obvious similarities with known ABC members, while some other genes were not considered real ABC genes even though they have been annotated as ABC transporters because they align poorly

with all existing ABC members. An example in this category is D2023.6 (data not shown).

2.3.2 Phylogenetic analysis

Multiple sequence alignment was conducted using ClustalX (Higgins and Sharp, 1988; Jeanmougin, et al., 1988; Thompson, et al., 1997) using whole ABC protein sequences, half ABC protein sequences or only ABC domain sequences with default parameters except for bootstrap label on NODE, not on BRANCH. The multiple alignment was subject to the neighbour-joining tree-making algorithm (Saitou and Nei 1987) and the number at the branch of the nodes represents the value from 100 replications.

2.3.3 Annotation of ABC genes for *C. briggsae*

All the ABC sequences were obtained as follows: run TBLASTN on the Sanger Center *C. briggsae* site (http://www.sanger.ac.uk/Projects/C_briggsae/) against *C. briggsae* contigs using one ABC protein sequence from each subfamily in *C. elegans* as query with default parameters, then combining all hits into one file and deleting redundant hits. Next, retrieve all the genomic sequence for these unique hits and run gene prediction using HMM-based FGENESH (http://www.softberry.com/berry.phtml) with the *C. elegans* option and/or GenScan (Burge and Karlin 1997) with default options. The resulting protein sequences were used to search against WormPep using BLASTP to find putative ABC homologs. If more than five of the top hits are ABC transporters, the query is presumed to be a candidate ABC transporter. In most cases, this query was assigned as

the *C. briggsae* ortholog for the first hit in *C. elegans* and named according to the *C. elegans* cosmid gene name prefixed by "Cb" (see Table 1).

All the putative ABC sequences were subjected to Pfam (Bateman et al. 2002) search. Syntenic information was recorded while performing BLASTP: if a *C. briggsae* ABC gene and at least one of its adjacent open reading frame (ORF) were both found to be orthologous to *C. elegans* ABC genes regardless of order, then synteny was assumed. If the orientation of *C. briggsae* ABC gene and its adjacent one were found to be different from those in *C. elegans*, an inversion was assumed. If three consecutive genes were found missing in the syntenic region, a deletion was assumed.

The resulting fifty-seven putative *C. briggsae* ABC transporters, combined with all the sixty *C. elegans* ABC transporters, were then subjected to similar phylogenetic analysis as mentioned before, and the resulting bootstrapped NJ tree was used to help assign orthology (See appendix figure 1). The *C. briggsae* orthologs were assigned with the following priorities: synteny>phylogenetic tree>alignment.

2.3.4 Pair wise alignment of *C. elegans* and *C. briggsae* syntenic regions containing ABC gene clusters

PipMaker was used to generate the pair wise alignment. It computes alignments of similar regions between two DNA sequences (Schwartz, et al., 2000). The resulting alignments are summarized with a ``percent identity plot (pip)". Alignments were plotted according to the exon/intron position defined in the *C. elegans* sequence file with the "Chaining" option so that PipMaker will identify and plot only matches that appear in the

same relative order in *C. elegans* and *C. briggsae* sequences. *C. elegans* exon positions are derived from Wormbase (WS84).

2.3.5 Construction of promoter::GFP fusion

All promoters were derived from wild type N2 (Bristol) genomic DNA by PCR. Promoter-GFP fusion constructs were built as shown in Figure 2 based on description by Hobert (2002). GFP coding sequence was amplified from A. Fire's vector, pPD95-67, including 5' nuclear localization sequence (NLS). The primer sequences used for GFP amplification were exactly the same as those used by Hobert (2002). Unique DsRed C, D and D* primers were used to amplify DsRed (RFP) (ClonTech) with sequences: 5'-CGCTCATCAAGAGAAAAATGG-3', 5'-AAACGCGCGAGACGAAAG-3' and 5'-GGAAACAGTTATGTTTGGTATATTGGG-3', respectively. In order to obtain putative promoters as accurately as possible for each gene, the following criteria were set for picking up primers to amplify promoter sequences. The 3' primer (B) is positioned as close as possible to, or spanning, the initiation codon (ATG) of the gene of interest. If a primer covers the ATG, it is mutated to ATC in order to avoid unnecessary translation initiation, which might cause a frame shift. I picked an approximate 3kb region upstream from the ATG as a putative promoter if the intergenic region is equal to or more than 3 kb in size, based on the assumption that an average promoter size is ~1kb in *C. elegans*. The full intergenic region was used as a promoter if this region ranged from 250 to 3,000bp; the promoters were arbitrarily extended into UTR or coding region of adjacent genes if the intergenic size was less than 250bp, resulting in a 1-3kb promoter depending on position of its upstream ORF. If no proper primer site could be found around the ATG,

the B primer was shifted upstream, but I tried to include splicing leader 1 (SL1) site within the PCR product (usually TTTCAG if it exists). The 5' overhang (~20bp) of the B primer is the reverse complement of the C primer for GFP or RFP. All primers were picked using the online program Primer 3 (Rozen and Skaletsky, 2000 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi)).

2.3.6 Microinjection

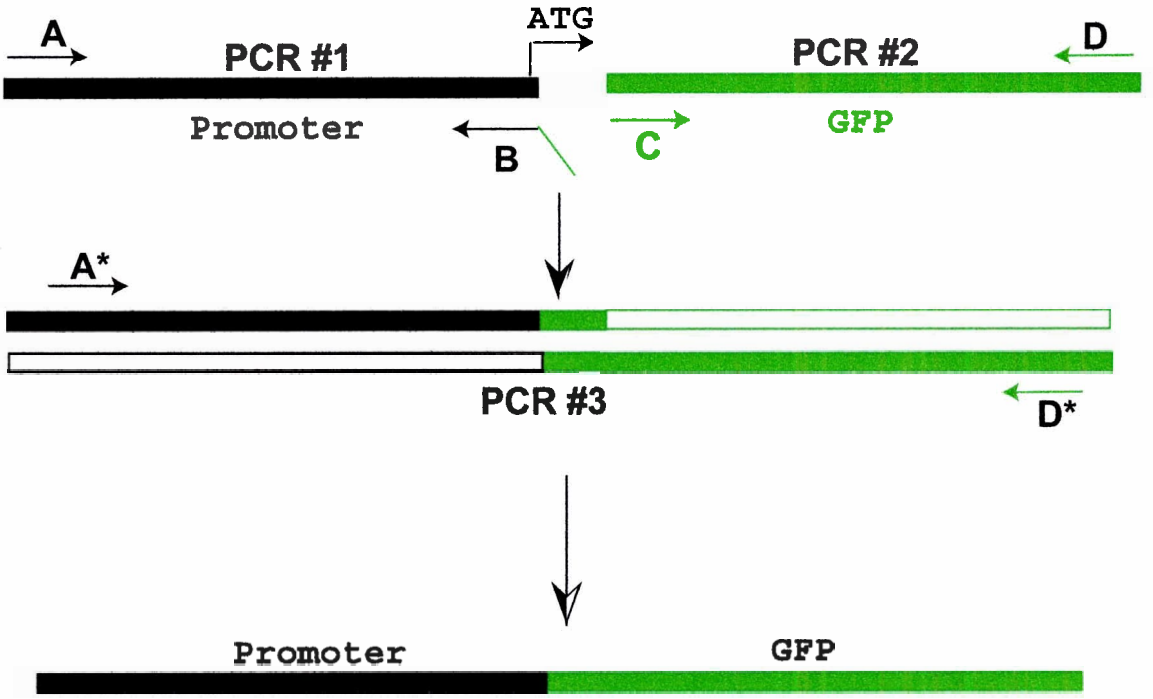
I used the *dpy-5* gene (a gift from Dr. Ann Rose) as a transgenic selection marker to screen for transgenic worms for a number of reasons. First, loss of function of the gene causes strong dumpy (short, fat) phenotype, making it easy to distinct such mutants from wild type worms (Brenner, 1974). Second, it is much easier to locate the syncytial area for microinjection in mutant *dpy-5* worms than those in N2 worms. Selection of transgenic F1 progeny is straightforward. In other words, rescued wild type worms are easily distinguished from the mutant background. The resulting transgenic animals yield a wild type phenotype rather than the twisted one resulting from the *rol-6* marker (Kramer and Johnson, 1993). For transgenic assay in *C. elegans*, the *dpy-5* rescuing plasmid, pCes-361 (a gift from Dr. Ann Rose), is co-injected with the promoter-reporter fusion construct into *dpy-5* mutant worms at concentrations of 100 and 10ng/μl respectively. Wild-type F1 worms are picked after a three-day incubation at 20°C and checked for wild-type F2 progeny after another three-day incubation. Microinjections were attempted up to 5 times with at least 25 injected worms per session until at least two transgenic lines were achieved.

2.3.7 Microscopy

The F2 rescued worms are first examined for embryonic and larval expression using a ZEISS Stemi SV11 dissecting microscope with GFP or RFP filters. All pictures were taken with a QIMAGING digital camera mounted on a ZEISS Axioskop. Only worms showing the same fluorescence pattern as most other reporter-expressing worms were photographed.

Figure 2. Building of promoter::GFP fusion constructs

Primer A and B are used to amplify promoter sequence (Black bar) for the gene of interest (PCR #1). GFP coding sequence (green bar) is derived from Andy Fire's GFP vector pPD95-67 by PCR # 2 with primer C and D. Primer B contains a 5' overhang that is complementary to primer C. Equal molar PCR products from PCR #1 and 2 are used as templates for fusion PCR (PCR #3) with primer A* and D* that are nested to primer A and D respectively. Since primer B and C containing regions are annealing together, PCR product from #1 and #2 can be used as reciprocal templates for elongation into complete fusion PCR product which is subsequently used as template by the primer A* and D*. The resulting PCR product can be directly used for microinjection (modified from Hobert, 2002).



Fusion PCR product ready for injection

2.4 Results

2.4.1 Identification of ABC genes in *C. elegans* and *C. briggsae*

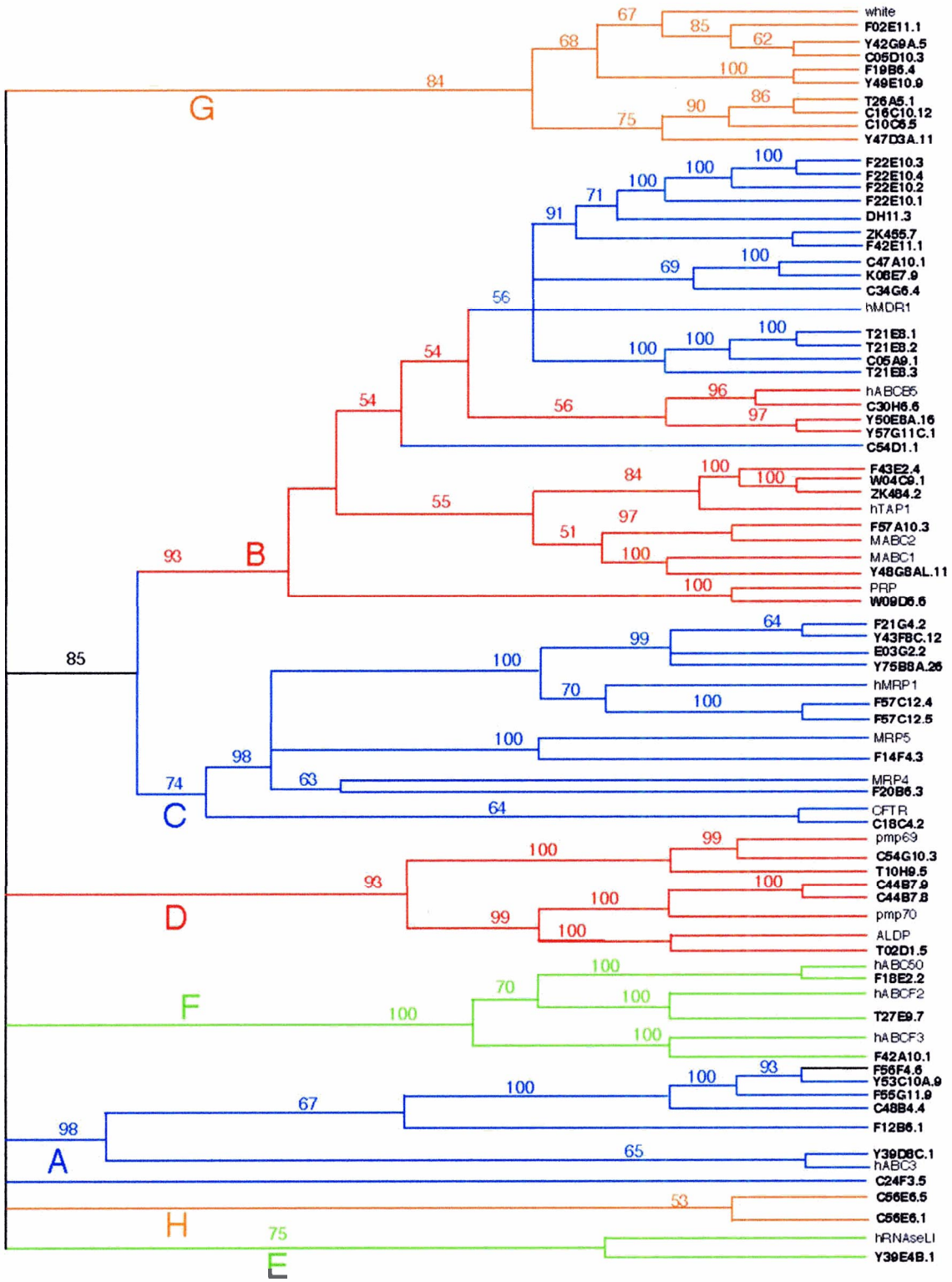
To begin to understand the organization and evolution of ABC transporters in *C. elegans*, the WormBase database (http://www.wormbase.org WS84, August 10, 2002) was searched for ABC transporter protein sequences by a combination of BLAST searches and the annotations already present in the databases. We identified 60 ABC transporters in *C. elegans,* which can be classified into eight subfamilies (Sheps, et al, 2004). The B subfamily is the largest (24 members) and the E and H subfamilies are the smallest ones (one and two members respectively)(Figure 3, Table 1). Classification of ABC transporters in *C. elegans* was done by a phylogenetic analysis combining entire ABC protein sequences (see Methods) with selected human and *Drosophila* ABC protein sequences included for reference. In order to investigate the evolution of large gene families across related species, I annotated all putative ABC transporters (57 in total) in *C. briggsae*, a morphologically similar worm, using its recently released genome sequence. Assignment of putative orthologous ABCs in *C. briggsae* was done by BLASTP search against WormPep in Wormbase (http://www.wormbase.org) with the highest scoring hit assigned as the putative *C. elegans* ortholog. The resulting orthologs were re-evaluated by their synteny, i.e. at least one or more corresponding orthologs lie on either or both sides of the gene of interest. A phylogenetic tree was generated using both *C. elegans* and *C. briggsae* ABC protein sequences, which, in most cases, confirmed the assignment of orthologs mentioned above, i.e., orthologous ABC genes tend to cluster

together (See appendix 1). Only three of the *C. elegans* ABC genes do not find obvious orthologs in *C. briggsae*. Fifty-one of the *C. briggsae* orthologous ABC genes were confirmed by both syntenic and phylogenetic analyses. Six putative *C. briggsae* ABCs do not have syntenic support (Table 1). However, most of them are clustered perfectly with their *C. elegans* equivalents in the phylogenetic tree (See appendix 1).

It is apparent that there is an expansion of subfamily B in the worm genome as compared to human and fly (Table 2). Both *C. elegans* and *C. briggsae* have more than twice the number of subfamily B members as human and fly. Most of these expansions were achieved by tandem duplication. For example, ten tandemly repeated ABCB genes were found on *C. elegans* X chromosome, forming two four-gene clusters and one two-gene cluster (see below). Exploration of the mechanisms by which the worm maintains multiple copies of ABC members will provide insights into the function and evolutionary dynamics of duplicated genes. The single member of ABCE subfamily suggests it has an essential function across species, which will be addressed in detail in the subsequent chapter.

Figure 3. Phylogenetic analysis of ABC transporters in *C. elegans*

The full protein sequences of all ABC transporters in *C. elegans* and those of some established human and fly ABCs were used for multiple alignment by ClustalX. The tree was generated using the neighbour-joining method implemented in PAUP*4.0b2a. Bootstrap value from 100 replications was shown on branch node. Subfamily assignment is based on human ABC, i.e., those clustered with specific human ABC are presumed to be in the same subfamily. Branches were colored by distinct domain organization as indicated in figure 1 (Courtesy of Jonathan Sheps).



Gene names indicate genome of origin by font: Homo **Caenorhabditis**

Table 1. List of ABC genes in *C. elegans* and *C. briggsae*

Subfamily	*C. elegans*					*C. briggsae*			
	Gene name	ORF name	Protein size	1st exon EST	ORF name[1]	Protein size	Identity%[2]	Synteny[3]	Inversion
A	*Abt-2*	F12B6.1	1547	No	CbF12B6.1	2296	82%-1012	Yes	No
(7)[4]	*Ced-7*	C48B4.4	1689	Yes	CbC48B4.4	1689	82%-1689	Yes	No
	Abt-3	F55G11.9	1431	No					
	N/A	F56F4.6	260	No	CbF56F4.6	1407	85%-138	Yes	No
	Abt-5	Y53C10A.9	1564	No	CbY53C10A.9[9]	1546	58%-939	No	N/A[5]
	Abt-4	Y39D8C.1	1802	No	CbY39D8C.1	1798	86%-1082	No	N/A
	Abt-1	C24F3.5	1429	No	CbC24F3.5	1567	58%-458	Yes	No
B	*Pgp-10*	C54D1.1	1283	No	CbC54D1.1	1247	89%-703	Yes	No
(24)	*Pgp-11*	DH11.3	1239	No	CbDH11.3	1209	85%-1178	Yes	No
	Pgp-1	K08E7.9	1321	No	CbK08E7.9	1868	90%-1321	Yes	Yes
	Pgp-9	C47A10.1	1294	No	CbC47A10.1	1294	89%-1294	Yes	Yes
	Pgp-2	C34G6.4	1265	No	CbC34G6.4	1851	94%-660	Yes	No
	Pgp-3	ZK455.7	1268	No	CbZK455.7	1268	92%-1268	Yes	No
	Pgp-4	F42E11.1	1266	Yes	CbF42E11.1	1281	88%-1266	Yes	No
	Pgp-12	F22E10.1	1318	No	CbF22E10.1	1299	85%-1113	Yes	No
	Pgp-13	F22E10.2	1291	No	CbF22E10.2	1224	85%-958	Yes	No
	Pgp-14	F22E10.3	1327	Yes	CbF22E10.3	1327	95%-1327	Yes	Transposase
	Pgp-15	F22E10.4	1270	No	CbF22E10.4	1192	92%-707	Yes	No
	Pgp-8	T21E8.3	1238	Yes	CbT21E8.3	1988	61%-1170	Yes	RTase[6]
	Pgp-5	C05A9.1	1283	No					
	Pgp-6	T21E8.1	1225	No	CbT21E8.1	1222	80%-1225	Yes	No
	Pgp-7	T21E8.2	1269	No	CbT21E8.2	1253	61%-1170	Yes	No
	Haf-2	F43E2.4	761	Yes	CbF43E2.4	829	80%-763	Yes	No
	Haf-4	W04C9.1	787	Yes	CbW04C9.1	868	95%-787	Yes	No
	Haf-9	ZK484.2	815	Yes	CbZK484.2	816	97%-816	Yes	No
	Haf-7	Y50E8A.16	807	Yes	CbY50E8A.16	795	79%-808	Yes	No
	Haf-8	Y57G11C.1	633	No					
	Haf-6	Y48G8AL.11	565	Yes	CbY48GAL.11	481	93%-311	Yes	D/I[7]
	Haf-1	C30H6.6	586	No	CbC30H6.6	972	91%-586	Yes	Yes
	Haf-3	F57A10.3	733	No	CbF57A10.3	1175	91%-668	Yes	No
	Haf-5	W09D6.6	704	Yes	CbW09D6.6	886	95%-363	Yes	D/I
C (9)	*Mrp-5*	F14F4.3	1400	Yes	CbF14F4.3	1424	84%-1424	Yes	No
	Mrp-6	F20B6.3	1421	No	CbF20B6.3	630	66%-630	Yes	No
	Mrp-2	F57C12.4	1525	Yes	CbF57C12.4	454	88%-454	Yes	No
	Mrp-1	F57C12.5	1534	Yes	CbF57C12.5	1548	85%-1140	Yes	No
	Mrp-3	E03G2.2	1398	No	CbE03G2.2	1377	88%-713	Yes	D/I
	Mrp-4	F21G4.2	1573	No	CbF21G4.2	2137	88%-889	Yes	No
	Mrp-7	Y43F8C.12	1153	No	CbY43F8C.12	961	93%-329	Yes	Yes
	Mrp-8	Y75B8A.26	1114	Yes	CbY75B8A.26	758	88%-631	No	N/A
	N/A	C18C4.2	1247	Yes	CbC18C4.2	823	87%-390	Yes	D/I

Table 1. (cont'd)

Subfamily	*C. elegans*					*C. briggsae*			
	Gene name	ORF name	Protein size	1st exon EST	ORF name[1]	Protein size	Identity%[2]	Synteny[3]	Inversion
D	N/A	T10H9.5	598	Yes	CbT10H9.5	701	86%-596	Yes	No
(5)	N/A	C54G10.3	305	Yes	CbC54G10.3	305	93%-305	Yes	No
	N/A	T02D1.5	734	Yes	CbT02D1.5	733	92%-734	Yes	No
	N/A	C44B7.8	665	Yes	CbC44B7.8	622	94%-622	Yes	No
	N/A	C44B7.9	661	Yes	CbC44B7.9	622	94%-622	Yes	No
E (1)	N/A	Y39E4B.1	610	Yes	CbY39E4B.1	610	95%-601	Yes	No[8]
F	N/A	T27E9.7	622	Yes	CbT27E9.7	141	98%-141	Yes	Yes
(3)	N/A	F42A10.1	712	Yes	CbF42A10.1	712	96%-712	Yes	No
	N/A	F18E2.2	622	Yes	CbF18E2.2	714	90%-622	Yes	No
G	N/A	Y42G9A.6	620	No	CbY42G9A.6	132	72%-128	Yes	No
(9)	N/A	C05D10.3	598	Yes	CbC05D10.3	649	91%-594	Yes	No
	N/A	F02E11.1	658	No	CbF02E11.1	520	85%-470	No	N/A
	N/A	F19B6.4	695	Yes	CbF19B6.4	708	73%-688	No	N/A
	N/A	Y49E10.9	454	No	CbY49E10.9	679	61%-243	No	N/A
	N/A	Y47D3A.11	547	Yes	CbY47D3A.11	728	81%-547	Yes	No
	N/A	C16C10.12	610	No	CbC16C10.12	409	81%-388	Yes	D/I
	N/A	C10C6.5	610	Yes	CbC10C6.5	595	86%-610	Yes	No
	N/A	T26A5.1	608	No	CbT26A5.1	574	82%-319	Yes	Yes
H (2)	N/A	C56E6.1	1677	No	CbC56E6.1	1431	72%-463	Yes	No
	N/A	C56E6.5	595	Yes	CbC56E6.5	463	75%-592	Yes	No

1 All *C. briggsae* ABC gene assigned the same name as its putative *C. elegans* equivalents prefixed by "Cb".
2 Identity% is derived from BLASTP search against WormPep with *C. briggsae* ABC protein sequences as query. Only the first alignment is taken into account.
3 Synteny means that at least one or more ortholog was found within 3 adjacent genes on either side of the target gene.
4 The number of members in the subfamily
5 Not applicable
6 RTase: reverse transcriptase
7 D/I denotes deletion or insertion
8 The sole member of the ABCE subfamily that is well conserved across eukaryotes
9 CBG20290 derived from wormbase (WS139)

Table 2. Subfamily distribution of ABC transporters in humans, flies, and worms

	ABCA	ABCB	ABCC	ABCD	ABCE	ABCF	ABCG	ABCH	Total
Homo sapiens*	12	11	12	4	1	3	5	0	48
D. melanogaster*	10	8	14	2	1	3	14	5	57
C. elegans	7	24	9	5	1	3	9	2	60
C. briggsae	6	22	9	5	1	3	9	2	57

* Sheps et al. 2004

2.4.2 Local duplication and transposition

There are two four-gene clusters in subfamily B in *C. elegans*, i.e. F22E10.1, 2, 3, & 4 and T21E8.1, 2, 3 & C05A9.1 (adjacent to T21E8.3 but on another cosmid). ORF and gene names are interchangeable across the thesis, especially F22E10.1, whose gene name, i.e., *pgp-12* will be intensively used in chapter 5. Most of them find obvious orthologs in *C. briggsae* except for C05A9.1 (Figures 4 & 5). Interestingly, both of the clusters have insertions within the syntenic regions in *C. briggsae*. A transposase and G-repeat-like protein are inserted between CbF22E10.2 and CbF22E10.3, suggesting the local duplication may have been achieved by transposition (Figure 4). In addition, the *C. briggsae* orthologs for F22E10.1 and F22E10.2 assigned by synteny do not agree with what was observed from the phylogenetic tree generated using ABC proteins sequences from both species (See appendix 1). The ortholog assignment by synteny is supported by the shared expression pattern between the two genes (Figure 9). Further refinement of the gene prediction needs to be done to clarify whether this results from transposition, gene conversion, or improper gene prediction. However, assignment of ortholoy between F22E10.1 & 2 and CbF22E10.1 & 2 is not clearcut and should be treated with caution due to lack of exact functional support. There is an RTase insertion between CbT21E8.2 and CbT21E8.3 orthologs in *C. briggsae* (Figure 5), implying the role of retro-transposition in the formation of the gene clusters. CbT21E8.3 and CbC05A9.1 appear to be fused in *C. briggsae,* both in GenScan and FGENESH predictions, into a single gene containing 3 ABC domains (rather than two genes with a total of four ABC domains (two ABC domains each), as in *C. elegans*). Alignment of the two orthologous DNA sequences by PipMaker (Schwartz et al. 2000) shows a poor match between them (the first three and a

half exons of *C. elegans* C05A9.1 are missing from *C. briggsae*) (Figure 7). However, the sequence contains the equivalent length of nucleotides, suggesting active evolution within the local ABC gene cluster. All the eight ABC genes within the two clusters belong to the subfamily B. There is another two-gene cluster consisting of ZK455.7 (pgp-3) and F42E11.1 (pgp-4), which are members from subfamily B. So, in total, there are ten tandem ABCB members in the *C. elegans* genome, which in part explains why the subfamily B shows an apparent expansion as opposed to other metazoans (Table 2).

Figure 4. GenScan prediction of the *C. briggsae* syntenic region for the *C. elegans* cosmid F22E10

The resulting open reading frames (ORFs) are used as query to search against *C. elegans* Wormpep by BLASTP in Wormbase. *C. briggsae* orthologs were assigned as described in methods. Transposase and G-repeat containing genes are assigned by the highest scores from BLAST hits against WormPep. These two genes were not found in *C. elegans* syntenic regions, in which the four ABC genes are well conserved.

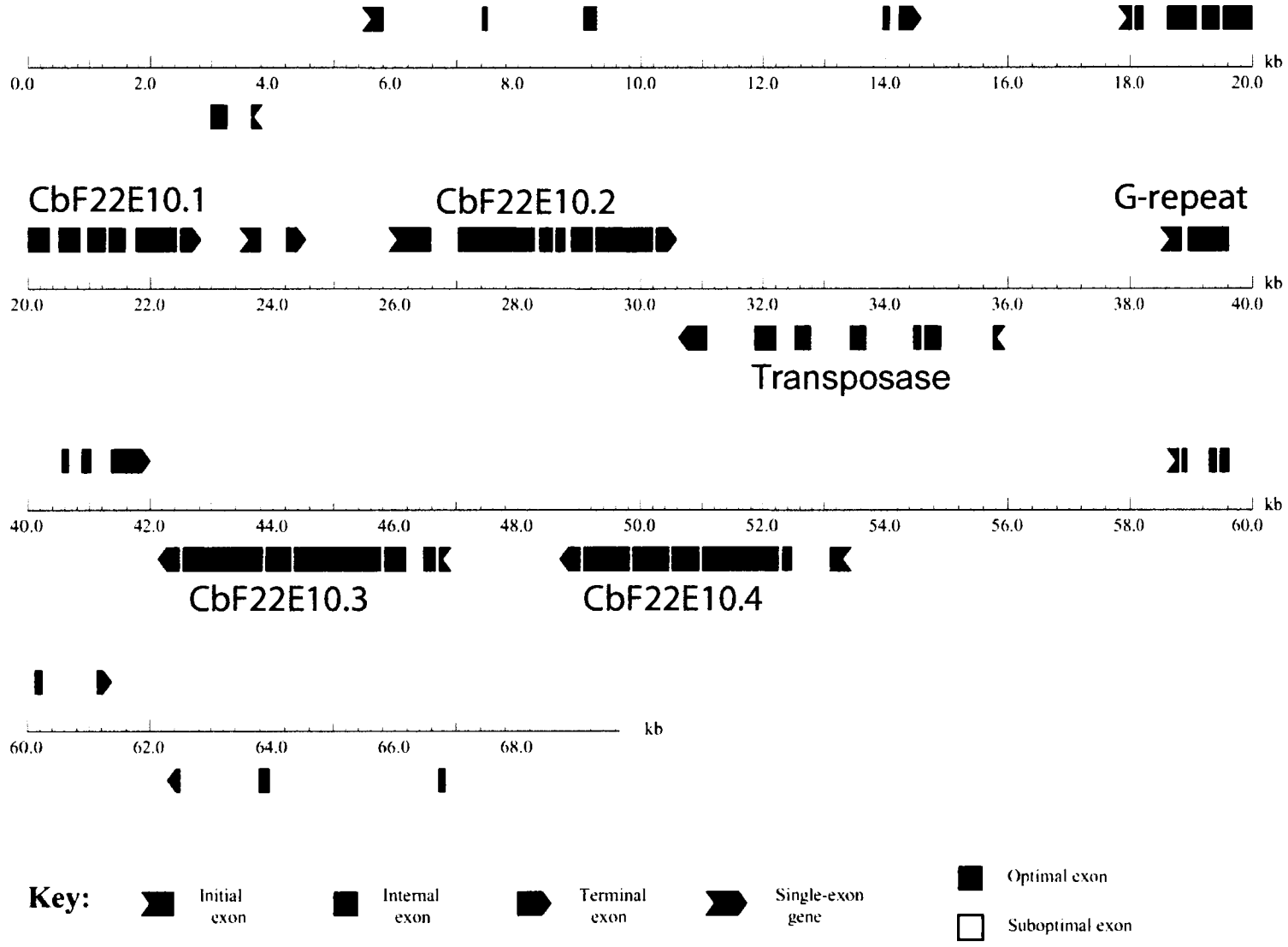
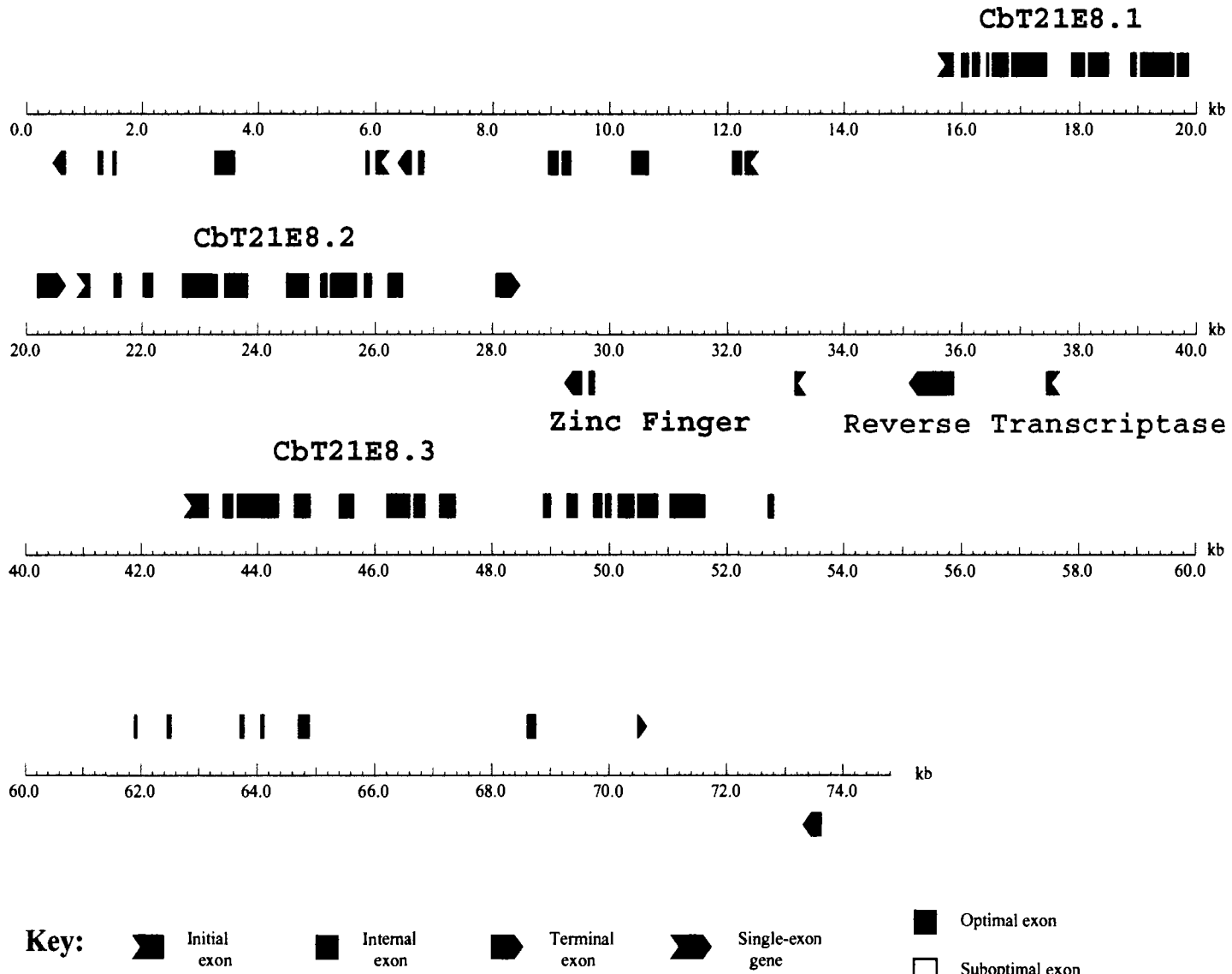

kb
0.0
2.0
4.0
6.0
8.0
10.0
12.0
14.0
16.0
18.0
20.0
CbF22E10.1
CbF22E10.2
G-repeat
kb
20.0
22.0
24.0
26.0
28.0
30.0
32.0
34.0
36.0
38.0
40.0
Transposase
kb
40.0
42.0
44.0
46.0
48.0
50.0
52.0
54.0
56.0
58.0
60.0
CbF22E10.3
CbF22E10.4
kb
60.0
62.0
64.0
66.0
68.0
Key:
Initial exon
Internal exon
Terminal exon
Single-exon gene
Optimal exon
Suboptimal exon

Figure 5. GenScan prediction of *C. briggsae* syntenic region of *C. elegans* T21E8-C05A9.1

Both GenScan and FGENESH predict the *C. briggsae* syntenic region of *C. elegans* T21E8.3 and C05A9.1 as a single ORF (data from FGENESH not shown). An ORF encoding reverse transcriptase was found between CbT21E8.2 and CbT21E8.3 but not in its *C. elegans* syntenic region. Orthologous gene names in *C. briggsae* were assigned as described in methods.


CbT21E8.1
0.0
2.0
4.0
6.0
8.0
10.0
12.0
14.0
16.0
18.0
20.0
kb
CbT21E8.2
20.0
22.0
24.0
26.0
28.0
30.0
32.0
34.0
36.0
38.0
40.0
kb
Zinc Finger
Reverse Transcriptase
CbT21E8.3
40.0
42.0
44.0
46.0
48.0
50.0
52.0
54.0
56.0
58.0
60.0
kb
60.0
62.0
64.0
66.0
68.0
70.0
72.0
74.0
kb
Key:
Initial exon
Internal exon
Terminal exon
Single-exon gene
Optimal exon
Suboptimal exon

2.4.3 Divergence of ABC transporters between *C. elegans* and *C. briggsae*

There are three ABC genes in *C. elegans*, F55G11.9, Y57G11C.1 and C05A9.1, for which no obvious orthologs were found in *C. briggsae*. Potential *C. briggsae* orhtolog of C05A9.1 is possibly fused with CbT21E8.3 since the latter is predicted to contain three ABC domains while all the remaining members within the cluster contains only two ABC domains *C. briggsae*, which is typical of full-sized ABC transporters.

Some ABC orthologs only have synteny on one side but not on the other side, a situation that may result from transposition or mis-assembly of the genome. The *C. briggsae* phylogenetic tree may not be perfect since all its protein sequences are derived only by *ab initio* gene prediction. Possible problems like improper exon-intron boundary and mis-prediction of the first exon could happen in the prediction of *C. briggsae* ABC gene structure. These cannot be improved without accessory data such as ESTs, SAGE probes as well as cross-species alignments with *C. elegans* or other nematode genomic sequence. Such refinement of ABC gene predictions *in C. briggsae* is, however, beyond the scope of this thesis.

The T21E8 orthologous region in *C. briggsae* was subject to gene prediction using both GenScan and FGENESH, and both programs give similar results, i.e., CbT21E8.1 and CbT21E8.2 are properly predicted, but the T21E8.3 and C05A9.1 orthologous region forms a single ORF in both cases, which probably represent a gene fusion event in *C. briggsae*. A putative reverse transcriptase insertion was found between CbT21E8.2 and CbT21E8.3.

Synteny was found between CbW09D6.6 and its *C. elegans* ortholog, but several genes seem deleted from the orthologous upstream region in *C. briggsae*. It is not certain whether this represents mis-assembled genomic sequence or a real deletion. The situation is similar for CbY48GAL.11, where there is a ~4kb deletion downstream of the syntenic region relative to that in *C. elegans*. For CbY43F8C.12, it looks like an independent translocation but no adjacent transposase found. Specifically, CbY43F8C.12 and CbY43F8C.13 are in reverse order in relation to their *C. elegans* orthologs, but no surrounding synteny is found outside these two genes. There is an obvious gene fusion event affecting CbF21G4.2 with a unusual length of amino acids, which can be easily found based on the multiple alignment of this gene with other members of the ABCC subfamily (data not shown).

2.4.4 Pair-wise alignment of genomic sequences for the two four-gene clusters

PipMaker yields a near-perfect exon-level match between F22E10 and its *C. briggsae* orthologous regions. This is not the case between the T21E8-C05A9 orthologous regions (Figure 6). It is worth noting that alignment of F22E10 and its *C. briggsae* orthologous region shows excellent conservation of exons but poor conservation in introns. There are some exceptions. For example, exons 8 and 9 of F22E10.2 look like one exon including intronic sequence between them. Apparently, there are obvious fingerprints, i.e., mosaic patterns with regions of high similarity (Webb et al. 2002), found within both 5' and 3' untranslated region (UTR)s (Figure 6). These small conserved regions may contain important regulatory motifs or so-called *cis*-elements

responsible for specific gene expressions. Mapping of tissue specific *cis*-elements with the aid of cross-species comparison will be described in detail in chapter 5 of this thesis.

Similar alignment has been done for the T21E8, C05A9 and their orthologous regions, which reveals potential error by gene-prediction (Figure 7). For instance, exon 9 of T21E8.2 looks like an intron, while a big portion of its 3' UTR like an exon. Also, the first exon does not seem real for T21E8.3, which is also short of EST match. As mentioned before, neither does GenScan and nor FGENESH predict the C05A9.1 orthologous region as an independent gene in *C. briggsae* but rather as a fusion with CbT21E8.3. This is in agreement with the pair wise aligment, i.e., the first three and a half exons for C05A9.1 show poor alignment to its corresponding sequences in *C. briggsae*, which probably causes the fusion of ORF during gene prediction. Intriguingly, the 3' UTR region shows several separate segments with good conservation. Whether these conserved regions represent alternative spliced exons, a real 3' UTR, or a 3' regulatory region need further investigation. Interestingly, the fingerprints are not as significant in the nearby UTRs of these four genes as those of F22E10 orthologous regions but some small conserved regions could still be found a little farther upstream of T21E8.1 and downstream of T21E8.3, suggesting the shared regulatory elements between the two species (Figure 7).

2.4.5 Chromosomal distribution of ABC transporters

As shown in Figure 8, ABC genes are widely dispersed over the *C. elegans* genome. Local gene duplication is commonly seen in the nematode genome. As mentioned above, there are two adjacent four-gene clusters (F22E10.1, 2, 3, 4 and

T21E8.1, 2, 3, C05A9.1) and one adjacent two-gene cluster (ZK455.7 and F42E11.1) of B subfamily on chromosome X, indicating active local duplication of B subfamily members on the sex chromosome. The mechanisms for biased ABC duplication on chromosome X remains to be determined. Another three two-gene clusters are also found over the genome for C, D and G subfamily respectively. In addition, several other non-adjacent gene clusters can be found on chromosome I, III and IV within 100 kb from each other. A similar ABC gene distribution is assumed in *C. briggsae* given the extensive synteny is observed between the two nematode species. As in figure 5 and 6, both orientation and coding regions of clustered genes are well conserved between the two species.

2.4.6 Experimental validation of ortholog assignment by promoter::GFP transgenic assay

In order to demonstrate the accuracy of ortholog assignment, I fused with GFP promoter regions of both *C. elegans* gene and its orthologous sequence in *C. briggsae* by fusion PCR. The resulting PCR products were used in transgenic assay to examine whether the putative orthologous genes share expression patterns. The chosen ABC genes include F22E10.1 and Cb-F22E10.1, ZK455.7 and CbZK455.7, F42E11.1 and CbF42E11.1 as well as C18C4.2 and CbC18C4.2. As expected, all the tested gene pairs gave the same expression patterns mostly in excretory cell at all developmental stages except for CbF42E11.1, which gave no observable expression (Figure 9). The same expression patterns between gene pairs supports the orthology assignment between the two species. Alignment of the two promoter sequences from F42E11.1 and CbF42E11.1

did not show any obvious conservation though the orthology were assigned both sequence alignment and sytenic information, suggesting the faster divergence of the orthologous pair compared to that of other ABC orthologous pairs. F42E11.1 is a tandemly repeated ABC with ZK455.7. Conservation of the similar expression patterns between the local duplicated ABC gene in *C. elegans* but not between its *C. briggsae* ortholog provides an opportunity for investigation the evolution of gene expression regulation.

Figure 6. Alignment of *C. elegans* F22E10 and its *C. briggsae* orthologous region by PipMaker with "Chaining" option

Orientation and coding regions are well conserved between the two species. Conservation can also be observed in 5'- and 3'-UTR but not in other intergenic and intron sequences. The exons 8 and 9 of F22E10.2 look like one exon.

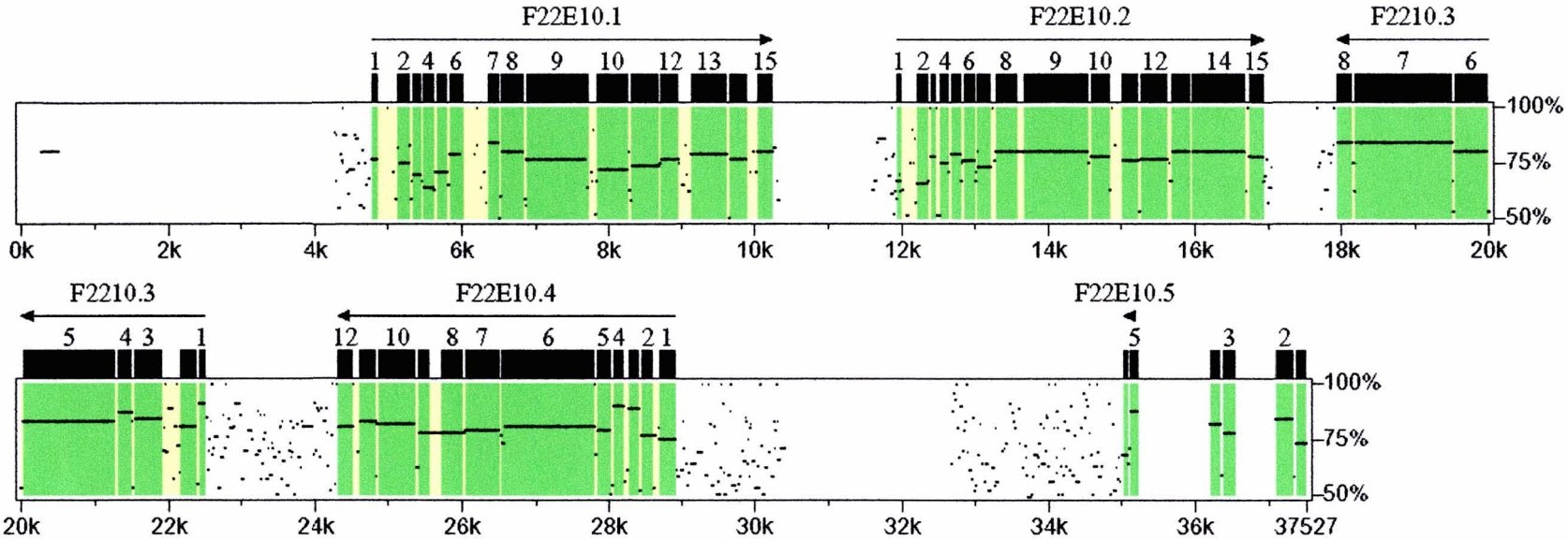
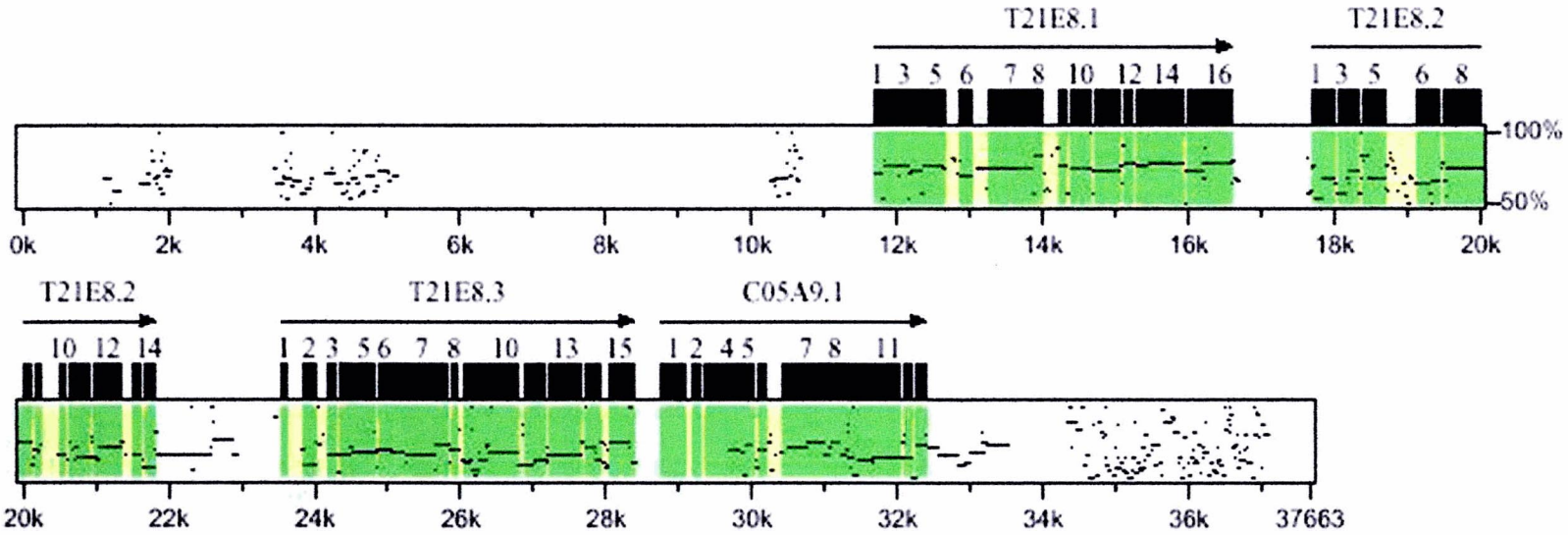

F22E10.1
F22E10.2
F2210.3
1 2 4 6 7 8 9 10 12 13 15
1 2 4 6 8 9 10 12 14 15
8 7 6
100%
75%
50%
0k 2k 4k 6k 8k 10k 12k 14k 16k 18k 20k
F2210.3
F22E10.4
F22E10.5
5 4 3 1
12 10 8 7 6 5 4 2 1
5 3 2
100%
75%
50%
20k 22k 24k 26k 28k 30k 32k 34k 36k 37527

Figure 7. Alignment of *C. elegans* T21E8-C05A9.1 and its *C. briggsae* orthologous region by PipMaker.

C05A9.1's orthologous region in *C. briggsae* seems fused with CbT21E8.3. Orthologous UTRs are not as conserved as those between F22E10 orthologous regions.


T21E8.1
T21E8.2
1 3 5 6 7 8 10 12 14 16
1 3 5 6 8
100%
50%
0k
2k
4k
6k
8k
10k
12k
14k
16k
18k
20k
T21E8.2
T21E8.3
C05A9.1
10 12 14
1 2 3 5 6 7 8 10 13 15
1 2 4 5 7 8 11
20k
22k
24k
26k
28k
30k
32k
34k
36k
37663

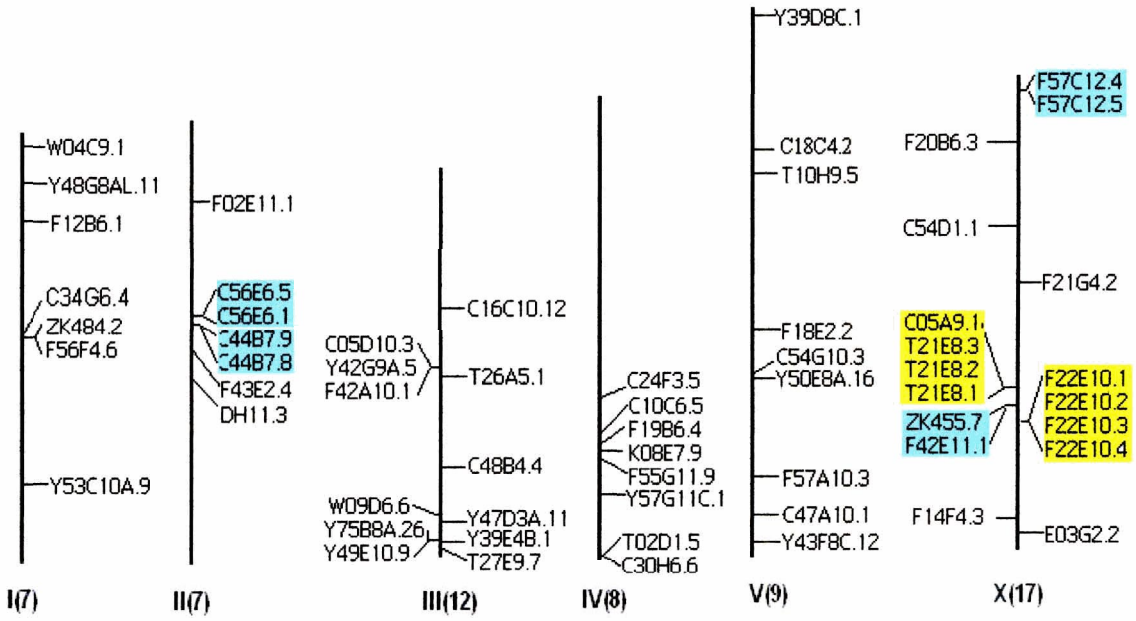


Figure 8. Genomic distribution of ABC transporters in *C. elegans*

Each vertical line represents individual chromosome indicated by Roman number down below. The number of ABC genes on each chromosome in included in bracket. Two 4-gene clusters are highlighted in yellow and four 2-gene clusters in light blue. Ten ABC genes from subfamily B are clustered on X chromosome.

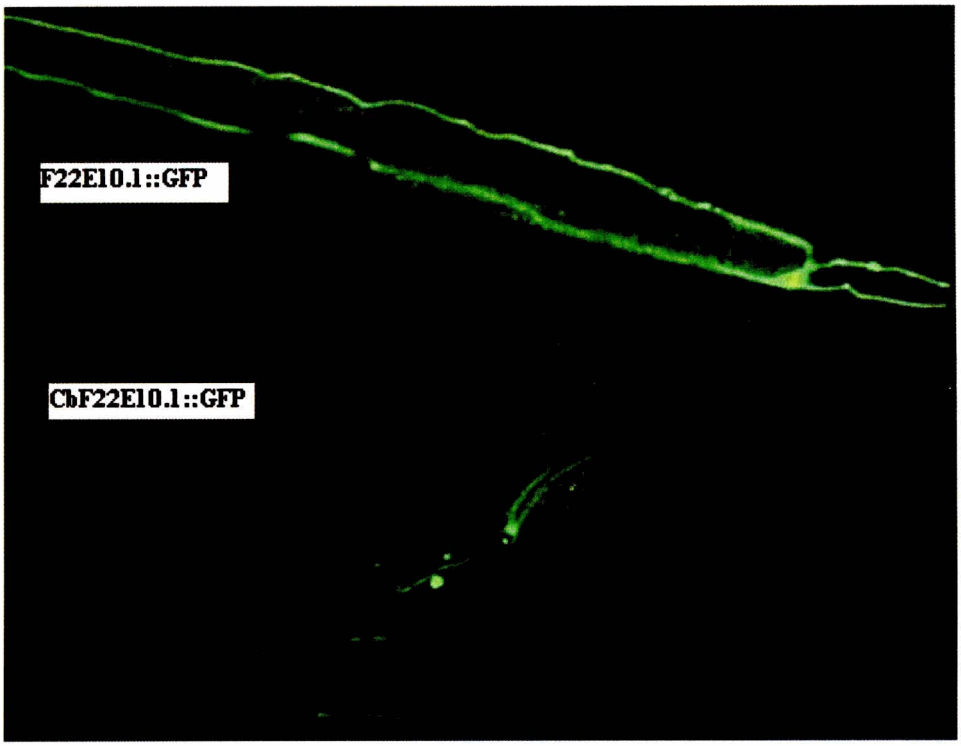


Figure 9. Promoter of both F22E10.1 (top) and CbF22E10.1 (bottom) drive GFP expression in the excretory cell

Only expressions in adults were shown.

2.5 Discussion

An investigation of both paralogous and orthologous members of large gene families can provide good insights into genome evolution. In this investigation, I conducted a comparative analysis of the whole ABC gene family between *C. elegans* and *C. briggsae*. The *C. elegans* genome contains 60 ABC genes while the *C. briggsae* genome has 57 members. Many of ABC genes occur in clusters, especially on the X chromosome, and the number and organization of the family are well conserved between the two species. Transposase or reverse transcriptase insertions were observed within two syntenic gene clusters in *C. briggsae* as compared to those from *C. elegans*.

2.5.1 The ABC genes are well conserved between the two nematode species

In general, ABC genes are well conserved between the two species. Only three ABC genes in *C. elegans* cannot find obvious orthologs in *C. briggsae*. It seems there has been a fusion event in the case of C05A9.1 since the pair-wise alignment of the genomic DNA shows many exons are still conserved between *C. elegans* and *C. briggsae*, suggesting this is a hot spot for evolution. The orientation of all ABC gene clusters is also conserved. Unfortunately, refined gene predictions cannot be made for *C. briggsae* at this moment due to the lack of EST and other supporting data, and this may result in some miscalled genes. For example, CbF12B6.1 is a possible fusion compared to F12B6.1, while CbF20B6.3 looks like a truncated product with reference to F20B6.3. However, taking synteny information into account can greatly increase confidence in the assignment of orthology. It is possible I have missed some potential *C. briggsae* ABC genes, especially those which have diverged too much from those in *C. elegans* to be identified

employing the techniques used for ABC identification in *C. briggsae* (see methods). Availability of genome sequences for both *C. elegans* and *C. briggsae* makes it feasible to examine evolutionary dynamics of a big gene family like ABC transporter family. The correlation of functional coupling and physical coupling seems to be prevalent in bacterial and archaea genomes (Dandekar et al. 1998; Tamames et al. 1997). Thus, the analysis of conserved gene orders in the same family among different genomes may provide insight into functional annotation of individual members. As mentioned previously, sixteen *C. elegans* ABC genes form two four-gene clusters and four two-gene clusters. It is interesting to note that 15 out of 16 tandem duplicated ABC genes (paralogs) can find obvious ortholog in *C. briggsae* with conserved gene order (synteny). Given the divergence of the two species about 100 million years ago (Stein, et al., 2003), it is hard to imagine these paralogous ABC genes are functionally redundant. Functional determination of these tandemly duplicated ABC genes will help develop a general understanding of why organisms maintain multiple copies of duplicated genes in the genome. F55G11.9 and Y57G11C.1 do not have any obvious orthologs in *C. briggsae*. Absence of both EST tags and RT-PCR product by ORFome project (Lamesch, et al 2004) suggest that they might be pseudogenes.

2.5.2 Local duplication and transposition of ABC genes in both species

ABC genes are distributed throughout the *C. elegans* genome (Figure 8) and many of them tend to occur in local clusters (16 out of 60), but few of these are present within operons except for T27E9.7 (Blumenthal et al. 2002), suggesting a different mechanism for regulation of duplicated ABC genes. Adjacent ABC gene clusters are more often seen

on X chromosome. 12 of 17 ABC genes on the chromosome have adjacent ABC genes, most of which occur in the same orientation. All the adjacent ones tend to cluster together on the phylogenetic tree (Figure 3), implying they are resulted from tandem duplication. Ten out of twelve tandemly duplicated ABC genes belong to subfamily B. It's yet to be clear why tandem duplication has become more frequent on the sex chromosome for ABCB subfamily. Local duplication can also be found on the other chromosomes. Tandem duplication is also observed in ABC transporter gene family of C. *briggsae*. A comparison of two 4-ABC gene clusters between the two species reveals the potential roles of transposition in the generation of duplicated genes. The mechanism that preserves a large proportion of duplicate genes for a long period of time remains controversial. Unequal crossover is believed to be responsible for tandem duplication of local genes (Tachikui, et al., 2003). The classical model predicts that duplicated genes initially have fully overlapping, redundant functions, such that one copy may shield the second copy from natural selection, if gene dosage is not critical (Watterson, et al., 1983). However, new models have been proposed to explain frequent preservation of duplicated genes. For example, Force et al (1999) described a duplication-degeneration-complementation (DDC) model, predicting that (1) degenerative mutations in regulatory elements can increase rather than reduce the probability of duplicate gene preservation and (2) the usual mechanism of duplicate gene preservation is the partitioning of ancestral functions rather than the evolution of new functions. It will be interesting to investigate whether the tandemly duplicated ABC genes are maintained to subdivide the roles of its ancestor by, for instance, examining their expression patterns. Subfamily B (24 members) has significantly expanded compared to other subfamilies (Figure 3, Table 2). The worm dwells in soil that is full of a variety of environmental challenges, such as heavy metal,

and natural toxins produced by plants and microbes. Some of the B subfamily members have been demonstrated to be involved in resistance to such compounds (Broeks et al, 1995). It is possible that expansion of the B subfamily probably provides a more efficient protection against these challenges.

2.5.3 Orthology and divergence

The *C. elegans* genes appear to have a faster rate of rearrangement compared to these of *Drosophila* (Coghlan and Wolfe 2002). I identified 12 rearrangements between ABC genes of *C. elegans* and *C. briggsae*, with an average of five genes per rearrangement. This is roughly the same ratio as previously reported, i.e. 4030 rearrangements for the whole genome between the two worms, estimated by whole genome comparison. As shown in Table 2, subfamilies D, E and F have a similar number of ABC members between worms, fly and humans. These family members may play core biological roles as has been postulated for orthologous genes conserved between worm and yeast (Chervitz et al. 1998). In the set of worm and human ABC transporters, only 8 of 49 possible pairs (16%) of sister genes contained a single human protein and its nematode ortholog. Of these 8 orthologous pairs, half of them involve ABC proteins that are predicted not to be transporters at all (subfamilies E and F, see below). Similarly, 10% of ABC transporters were found in orthologous pairs when the comparison was made between worm and yeast genomes (Sheps, et al., 2004). A comprehensive comparison of worm and yeast genomes revealed that 57% of genes in highly conserved gene families were found in orthologous pairs, and the study suggests that such gene families provide a conserved 'core' proteome which forms the basis of eukaryote biochemistry (Chervitz et

al. 1998). Clearly, ABC transporter evolution has not been typical of strongly conserved gene families, and this might imply that ABC transporter-mediated metabolism differs radically among eukaryotes, yet this is one of the most ancient and widespread gene families known. Perhaps ABC transporters must also be regarded as the most functionally divergent gene family in the eukaryotic core proteome. However, subfamily A size looks notably smaller in worms (7 compared to 12 in human, see Table 2). Several human ABCA members participate in cellular lipid homeostasis in specialized environments (Kaminski et al. 2000 & 2001) which may not be necessary for worms. While for the B, C and G subfamilies, the number of ABC genes is expanded in worms, especially for subfamily B (24 or 22 as compared to 12 in humans). This can be partly explained by the fact that many members of these subfamilies are expected to function in drug or toxin exclusion, which is critical for worms that survive in toxic soil environments. Within the Pgp related ABCB subfamily the only clear orthologous relationship between *C. elegans* and human ABC genes is that of W09D6.6 and MTABC3, a half-transporter involved in iron homeostasis (Sheps et al, 2004; Mitsuhashi et al. 2000). Its rat orthologue is overexpressed in a rat model of hepatocarcinogenesis (Furuya et al. 1997). Among ABCC molecules, whose range of functions broadly overlaps with P-gps, only human MRP5, a transporter of nucleotide analogues (Jedlitschky et al. 2000; Wijnholds et al. 2000) and worm F14F4.3 are sister genes. Within subfamily ABCA, Y39D8C.1 and human ABC3 are orthologues, but again no functional data is available for either gene. There is only one member in E subfamily for all these species, suggesting an important function for ABCE across eukaryotes. All four members of subfamilies E and F (Fig.3) have unambiguous and strongly supported orthologues among nematodes, yeast, fly and human (Sheps, et al., 2004). I would predict that the nematode versions perform the same roles as

their human (and indeed yeast) counterparts. Their strong conservation argues for involvement in something indispensable, at least on an evolutionary timescale. The yeast orthologue of the RNAse L Inhibitor protein (ABCE subfamily) is essential for growth, as is YER036c, the yeast homologue of T27E9.7/ABCF2 (Cherry et al. 2002). On the other hand GCN20, the yeast version of F42A10.1/ABCF3, is non-essential, though mutants do have specific defects in translation. The three genes in subfamily F, which lack transmembrane domains, are generally regarded as forming ribosome-associated proteins involved in regulation of mRNA translation, rather than transporters. The RNAse L Inhibitor, also known as the Oligoadenylate Binding Protein (OABP), is thought to be involved in the regulation of both the interferon-induced antiviral response (Le Roy et al. 2000) and muscle differentiation (Bisbal et al. 2000) in mammals. However, lack of Rnase L in *C. elegans* makes ABCE protein unlikely as its inhibitor, which will de detailed in the next chapter.

Increasing experimental data from functional genomics provide direct information about the implications of coupling of genes, especially those from cDNA microarrays (Brown and Botstein 1999), localization assays (focus of chapter 4 in this thesis) and protein-protein interaction analysis (Li, et al., 2004). Integration of computational prediction and systematic experimental investigation will finally unveil the complex networks at a functional level.

CHAPTER 3 CHARACTERIZATION OF ABCE PROTEIN IN *C. ELEGANS*[1]

[1]This chapter has been published in Biochemical and Biophysical Research Communications 2004b under the co-authorship of Lily Fang, Robert C. Johnsen and David L Baillie

3.1 Abstract

The ABCE gene (Y39E4B.1) has been annotated as an RNase L inhibitor in eukaryotes. All eukaryotic species show the ubiquitous presence and high degree of conservation of ABCEs. However, RNase L is present only in mammals. This indicates that ABCEs may function not only as RNase L inhibitors in mammals, but also as something else that has not been determined. As an initial step to search for novel functions of ABCE, I characterized the gene in *C. elegans* by a combination of data mining and functional assays. ABCE's promoter drove GFP expression in hypoderm, muscle, pharynx, vulvae, head & tail neurons at all developmental stages. Early embryonic expression was also observed. Promoter plus partial genomic region drove GFP in both cytoplasm and nucleus. The products of three genes, *rpl-4, nhr-91* and C07B5.3 were previously found to interact with ABCE. The expression data showed overlapping expression patterns of ABCE with *rpl-4* and *nhr-91*, but not C07B5.3. RNAi against ABCE resulted in embryonic lethality, slow growth as well as defects in vulvae, germline and nerve system. Taken together, ABCE is an essential gene that might be involved in control of translation and transcription, work as shuttle protein between

cytoplasm and nucleus, and possibly as a nucleocytoplasmic transporter. In addition, RNAi data suggest that ABCE and NHR-91 may function in vulva and germline development as well as molting pathways *in C. elegans*. Furthermore, my data showed that ABCE, along with its interacting components, functions in a well-conserved pathway.

3.2 Introduction

As mentioned previously, 60 ABC transporter genes have been identified in the nematode *C. elegans'* genome (Sheps et al, 2004). They are classified into 8 subfamilies, i.e., ABCA to ABCH. Many of these subfamilies include multiple duplicated paralogues. However, only one member of the ABCE subfamily, i.e., Y39E4B.1 has been identified in *C. elegans*, which is referred to as ABCE hereafter. Unlike typical ABC transporter proteins, the ABCE protein, together with ABCF proteins, contains two ATP binding domains but no obvious transmembrane domains, suggesting it is not an intrinsic transmembrane transporter. Sequences of ABCE proteins are highly conserved among all eukaryotic species, especially in their ATP binding domains, which show over 90% identity across all eukaryotes (Kerr 2004). These proteins are also comparable in length, which contain around 600 amino acids. Similar to *C. elegans*, ABCE has been found to have only one subfamily member in most eukaryotes, except in *A. thaliana*. No obvious ABCE homologues were found in bacteria, suggesting a fundamental role in eukaryotes.

ABCE genes of eukaryotes were commonly annotated as RNase L inhibitors, which were first identified in human (Bisbal et al, 1995). Human RNAse L inhibitor has implicated in the antiviral mechanisms of interferon (IFN) and might regulate RNA turnover and stability. IFN induces four different forms of human 2-5A-synthetases

which, upon activation by double-stranded RNA (dsRNA), convert ATP into an unusual series of oligomers known as 2-5A. 2-5A-dependent RNase L is the effector enzyme of this system. Its activation by subnanomolar levels of 2-5A leads to the inhibition of protein synthesis by cleavage of mRNA at the 3' side of UpNp sequences (Floyd-Smith et al, 1981; Kerr and Brown 1978; Zhou et al, 1993). RNase L appears to have homologues only in mammals such as human, mouse, and rat. No apparent RNase L homologues have been identified in other eukaryotes or in the archaea, which possess a candidate ABCE gene (Kerr, 2004). Classically, ABCE's main function was thought to work as an RNase L inhibitor in all eukaryotic organisms. However, non-mammalian eukaryotic organisms do not contain RNase L. Therefore, the ABCE genes very likely have yet to be identified functions other than as RNAse L inhibitor. Search for the novel functions of these proteins constitutes the main focus of this chapter.

In contrast to the *in vitro* studies of mammalian ABCE proteins, *C. elegans* as a model system provides an opportunity to study ABCE functions within the context of a developing multi-cellular organism. As an initial step into identifying ABCE's functions, I characterized this gene in *C. elegans* by a combination of data mining and expressional and functional assays. The results suggest that ABCE is an essential gene, possibly involved in translation and transcription control, growth regulation and vulvae development, and very possibly functions in well-conserved pathways among all eukaryotes.

3.3 Materials and methods

3. 3. 1 Strains

All strains were maintained and cultured using standard techniques. N2 (Bristol); KR3532, dpy-5(e907); GR1373 eri-1(mg366) IV; BC10798, dpy-5(e907), sEX10798[dpy-5(+) + rCesY39E4B.1-GFP + pCes361]; BC06518, dpy-5(e907), sEX1050[dpy-5(+) + rCes rpl-4::GFP + pCeh361]; BC06519, dpy-5(e907), sEX1051[dpy-5(+) +rCesC07B5.3-GFP + pCes361]; BC06495, dpy-5(e907), sEX1029[dpy-5(+) + rCes-nhr-91-RFP + pCes361]; BC12778, dpy-5(e907),sEX1277 8 [dpy-5(+) + rCesC26C6.5-GFP + pCes361]; BC11724, dpy-5(e907), sEX11724 [dpy-5(+) + rCesF36H1.2-GFP + pCes361].

3.3.2 Sequence analysis, data mining and phylogenetic analysis

All protein sequences were derived from NCBI, EMBL or Wormbase (WS126). The ABCE protein sequences from different species were also subject to multiple alignments using CLUSTALX with Gonnet protein weight matrix. A 100 times bootstrapped neighbour joining tree was generated from the resulted alignment edited using GeneDoc (Nicholas et al, 1997). SAGE data were retrieved from BC Genome Science Centre SAGE site: http://elegans.bcgsc.bc.ca.

3.3.3 Construction of promoter::reporter fusion PCR products

Promoter or full-length Green Fluorescence Protein/DsRed Fluorescence Protein (GFP/Ds-Red) fusion constructs were built basically as described in chapter 2. DsRed specific C, D and D* primers were used to amplify DsRed with sequences: 5'- CGC TCA

TCA AGA GAA AAA TGG-3', 5'-AAA CGC GCG AGA CGA AAG-3' and 5'-GGAAACAGTTATGTTTGGTATATTGGG-3' respectively. The full-length protein fusion with DsRed was generated using similar methods, with DsRed cDNA fused in-frame at the end of Y39E4B.1.

3.3.4 Microinjection and microscopy

Microinjection and microscopy were performed as described in chapter 2.

3.3.5 RNAi by feeding using the *eri-1* (*mg366*) mutant

RNAi by feeding was done with mutant *eri-1*(*mg366*) (Kennedy et al, 2004) as described previously (Fire et al, 1998). RNAi bacteria (kindly provided by Dr. J. Ahringer) were streaked onto NGM plates containing ampicillin, tetracycline and isopropyl-β-D-thiogalactoside (IPTG), and incubated overnight at room temperature. Ten L4 *eri-1* (mg366) worms were seeded onto the plates and incubated at room temperature for another 24 hours. Three *eri-1* (*mg366*) worms were transferred onto the 3 replica plates and scored for RNAi phenotypes every 8 hours for 3 consecutive days. The RNAi worms were subjected to second round RNAi using the same bacteria to score the phenotypes.

3.3.6 RNAi by injection

PCR primers were designed so that 615 bp PCR product covers the first two exons of ABCE with T7 or T3 overhang on either side. It was used as template to make single stranded RNAs using Ambion MEGAscript T3/T7 High Yield Transcription Kit (Cat. No. 1338) according to manufacture's specifications. The resulting RNAs were combined

together and purified with Invitrogen Micro-to-Mini Total RNA Purification System (Cat. No. 12183-018). The purified RNAs were annealed by incubation at 68 °C for 10 minutes and 37 °C for 30 seconds. The annealed RNAs were quantified on 1% agrose gel. The injection mixture contained 20 ng/μl double stranded RNA, 4 mM KPO4, 0.6 mM K Citrate and 0.4% PEG8000. L4 or young adult were used for microinjection. The injected worms were placed on fresh plates for six hours and transferred to another fresh plates every 24 hours for consecutive three days. The plates containing eggs for the first six hours were discarded. Phenotypes were scored on the remaining plates.

3.4 Results

3.4.1 ABCEs are well conserved across all eukaryotic species

There is over 65% identity between human and worm ABCE protein sequences. They also have comparable numbers of amino acids in each protein, ranging from 594 to 610 (Table 3). Surprisingly, the ABCE subfamily has only one member in most species except for *A. thaliana*, which has two members. This indicates there have been few gene duplications or that, in general, no functional redundancy or separable novel function ABCE proteins have evolved. The *C. elegans* homolog with highest similarity to ABCE is a member of the ABCF subfamily, i.e., F55G11.9. However, the percentage of identity between worm ABCE and F55G11.9 is significantly lower than that between worm ABCE and its orthologues. In addition, F55G11.9 encodes 1431 as opposed to 610 amino acids for worm ABCE and only 169 amino acids can be aligned between worm ABCE and F55G11.9 (Table 3). Non-redundancy and high-degree of conservation suggest the significant functions of ABCE among all eukaryotes.

Table 3. Conservation of ABCE across eukaryotes

Protein sequences are retrieved from GenBank or EMBL, aligned against *C. elegans* ABCE protein sequence by online pair wise BLAST program from NCBI with default parameters except the Filter is turned off.

	Protein length	Length of aligned region [a]	Identity	No. of paralogs
C. elegans	610	610	100%	1
C. briggsae	610	610	95%	1
D. melanogaster	611	603	67%	1
M.musculus	599	594	66%	1
H.sapiens	599	594	65%	1
S. cerevisiae	608	605	63%	1
A. thaliana [b]	605	599	65%	2
C.elegans [c]	1431	169	24%	3

[a] Alignment against *C. elegans* sequence [b]the sequence used for alignment is TrEMBL: Q8LPJ4, otherwise retrieved from GenBank [c] *C. elegans* homolog (F55G11.9) with highest identity to ABCE

3.4.2 ABCEs function as more than RNase L inhibitors

ABCEs in all eukaryotic species are commonly annotated as RNase L inhibitors because they were first isolated from an expression library by its binding to 2-5ApCp, association with and inhibition of RNase L (Bisbal et al, 1995). Given the ubiquitous nature and high conservation of ABCEs across eukaryotes, I expected the proteins involved in the same pathway with ABCE to be conserved and co-localized. However, BLAST searches with human RNAse L sequence as a query against the non-redundant sequence database identified hits only in human, mouse, and rat using low complexity filtering to mask the ankyrin repeats of the RNase L. No putative RNase L homologues were identified in any other eukaryotes in which the ABCE gene was identified (Kerr 2004). From this I hypothesize that ABCE might not function as an RNase L inhibitor in all eukaryotes because RNAse L is absent in the species in which ABCE is present. Presence of RNase L in mammals suggests an immune system that has evolved to fight viruses unique to these species. So ABCE may function partly in an antiviral pathway in mammals, but it may have other functions common to all eukaryotes, which have not been determined.

3.4.3 Expression patterns of ABCE in *C. elegans*

As an initial step in investigating the role of ABCE, a promoter::GFP fusion construct was generated by PCR (see Material and Methods). The promoter contained a 2965 bp sequence upstream of the start ATG. The resulting PCR product was co-injected

with *dpy-5(+)* plasmid into dpy-5 hermaphrodite to generate a transgenic line. The 2965 bp ABCE promoter drove GFP expression in the majority of tissues and developmental stages, including hypoderm, anterior and posterior neurons, muscle, pharynx and vulva in adults, larvae and embryos (Figure 10 and Table 4). ABCE promoter plus genomic region covering the first two exons drove GFP expression in the similar tissues, and GFP was clearly seen in both cytoplasm and nucleus from body wall muscle cells (Data not shown). However, transgenic animals carrying full length ABCE fused to RFP did not show any observable expression. I speculate that transgenic animals did not carry the fusion constructs because the construct is so toxic that the concentration used for injection mixture had to be 10 times lower than other constructs. It is possible that transgenic animals with the constructs are selected against, so that only animals without the construct can survive as transgenic lines. The worm homologue with the lowest E-value for human RNase L is ankyrin protein (F36H1.2), with unknown function. Similarly, promoter::GFP transgenic lines were generated for the F36H1.2. It expressed mostly in the excretory cell and the intestine, but it had few overlapping patterns as those of ABCE (Table 4), suggesting different pathways for the two genes in *C. elegans*.

3.4.4 Proteins that interact with ABCE are completely or partially co-localized

A large-scale yeast two-hybrid screen was done in *C. elegans* (Li et al, 2004), in which several proteins (RPL-4, NHR-91 and C07B5.3) were found to interact directly with worm ABCE protein. This significantly advanced the understanding of the potential functions of ABCE. The interactions of ABCE with two of the three proteins were supported by the transgenic report assay that showed worm ABCE is completely or

partially co-expressed with *rpl-4*, *nhr-91* but not C07B5.3 (Figure 10). Both ABCE and RPL-4 can be seen in hypoderm, muscle, pharynx and neurons. *rpl-4* encodes a large ribosomal subunit L4 protein. The interaction and co-expression of ABCE with RPL-4 strongly suggests that ABCE is involved in protein translation control. The exact biochemical mechanisms of ABCE in protein translation regulation require further investigation.

A promoter: DsRed construct of *nhr-91* gave expression in embryos, larvae and adults. DsRed was observed in adult head & tail neurons, hypoderm, vulvae, seam cells and excretory duct cell. These observations were consistent with what was reported previously (Table 4). Expression was also observed in larval intestine. Both ABCE and *nhr-91* were expressed in vulvae, hypoderm and neurons (Figure 10). Interaction and co-expression of ABCE and *nhr-91* suggest that the two genes are involved in the same conserved pathway. Since NHR-91 is a transcription factor, interaction of ABCE protein and NHR-91 suggests the potential role of ABCE in gene transcription.

Figure 10. Localization of ABCE and its interacting components by transgenic reporter assay.

Normaski and fluorescent (GFP/DsRed) pictures were taken and superimposed to each other (Merge). ABCE promoter drove GFP expressions in hypoderm, vulvae, muscle, pharynx, head and tail neurons. Rpl-4 promoter drove GFP expressions in hypoderm, muscle, pharynx as well as head and tail neurons. Nhr-91 promoter drove DsRed expressions in hypoderm, vulva, seam cell, excretory duct cell as well as head and tail neurons. C07B5.3 promoter only drove weak GFP expressions in posterior intestine. Overlapping expression patterns were observed between ABCE, rpl-4 and nhr-91, but not C07B5.3.

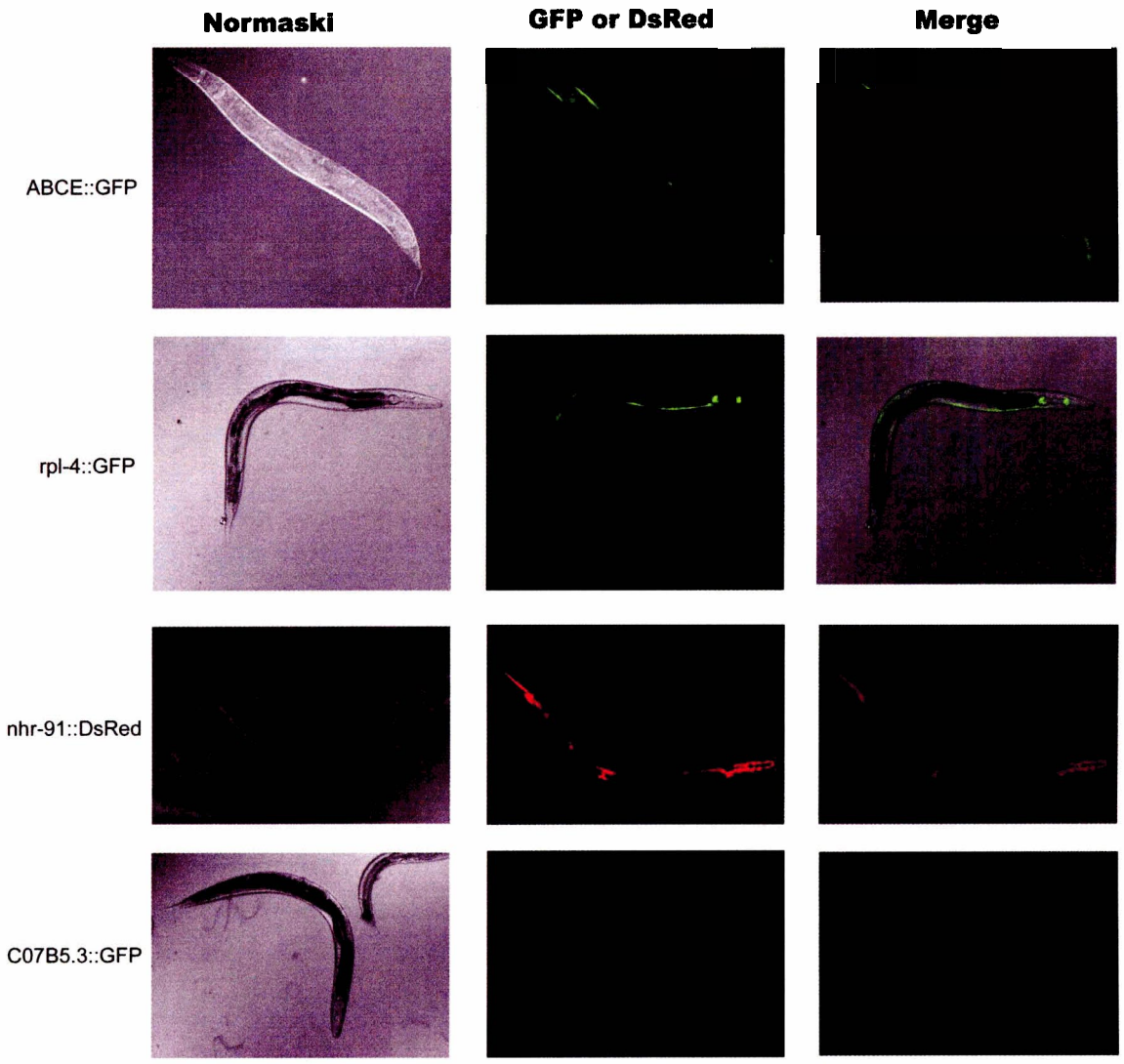

Normaski
GFP or DsRed
Merge
ABCE::GFP
rpl-4::GFP
nhr-91::DsRed
C07B5.3::GFP

Table 4. Comparison of expression pattern and RNAi phenotypes of Y39E4B.1 and its potential interactors

Gene Name	Expression patterns	RNAi phenotypes	No of SAGE tags [c]
Y39E4B.1	larvae, adult hypoderm, muscle, pharynx, neurons, vulva	Emb[d] Gro[e]	1
rpl-4[a]	hypoderm, muscle, pharynx and neurons	Emb Ste[f]	42
nhr-91[a]	neurons, vulva, seam cells, excretory duct cell	WT[g] Emb Gro Vul[k]	5
C07B5.3[a]	weak in larvae and adult gut	WT	0
F36H1.2[b]	embryo, larvae, adult gut excretory cell	Dpy[h] Pvl[i] Rup[j]	3

[a]Genes that interact with Y39E4B.1 by Li et al (2004), [b]highest score BLASTP hit in WormPep using human RNAse L sequence as query, [c]Number of long SAGE tags in embryo (McKay et al, 2003) in position +1, [d]embryonic lethality, [e]slow growth, [f]sterile, [g] wild type, [h]dumpy, [i]protruding vulva, [j]exploded body, [k]Phenotypes (in bold font) were observed by RNAi with *eri-1* only

All nuclear receptors (NR) share a common structure with specific, conserved domains that function in DNA binding, ligand binding, and transcriptional activation (Gronemeyer and Laudet, 1995). The exact part of NHR-91 that interacts with ABCE protein remains to be determined. Given the interaction of ABCE protein with both NHR-91 and RPL-4, it is possible that ABCE may function not only in protein translation control, but also in communications between gene transcription and translation, possibly encoding a shuttle protein between cytoplasm and nucleus. The results from PSORT (Gabaldon et al, 2004) prediction of worm ABCE show that the protein has a 37% possibility in nucleus and 20% in cytoplasm. Given the conservation of ABCE protein and NHR-91, it is plausible that the pathway in which the two proteins are involved would be well conserved across all other eukaryotic species.

No obvious orthologues were identified for C07B5.3 in other phyla (data not shown), suggesting its unique roles in nematodes. It is expressed weakly in adult posterior gut (Figure 10). Few overlapping expression patterns were seen between ABCE and C07B5.3, suggesting a possible false positive interaction between the two gene products in yeast two-hybrid screen. But this is a tentative conclusion because the GFP expression data for C07B5.3 is inconsistent with previous *in situ* hybridization patterns (Birchall PS et al, 1995). I have yet to determine the reason for the discrepancies.

3.4.5 RNAi using *eri-1* mutant captured extra phenotypes for *nhr-91*

Both ABCE and RPL-4 are essential for normal development based on existing RNAi data (Kamath et al, 2003). In order to further investigate the potential roles of

ABCE and its interacting proteins, RNAi studies was undertaken for ABCE and its interacting proteins using *eri-1(mg366)* due to its RNAi susceptibility in nerve systems (Table 4) (Kennedy et al, 2004). In addition to the phenotypes previously observed, additional phenotypes were observed for *nhr-91* (Table 4 in bold font). The progeny of worms after RNAi against *nhr-91* showed severe slow growth: 70% of eggs reached sizes of L2 or L3 stage after 3 day incubation on RNAi plates in 20 °C. In addition, slight embryonic lethality was also observed for the gene. Slow growth and embryonic lethality suggest that NHR-91, together with ABCE protein are involved in transcription control of genes involved in development and growth. Microscopic examination of the morphology of slow growing worms demonstrated that many of them had defects in vulvae development and molting (Figure 11). Embryonic lethality was also observed as previously seen (Gronemeyer et al, 1995). Expression of both ABCE and NHR-91 in vulvae and hypoderm suggest the two proteins function in molting and vulvae development pathways in *C. elegans*.

3.4.6 ABCE is an essential gene that regulates germline development

Since high conservation of ABCE proteins is observed among eukaryotic species as well as non-redundancy, ABCE is likely an essential gene. I decided to investigate this possibility by RNAi with injection because ABCE was expressed intensively in neurons, which are refractory to RNAi by feeding (Fire, et al, 1998). I performed RNAi by injection with double stranded RNA targeting first two exons of ABE gene, which shows least homology in the worm genome. As expected, RNAi phenotypes by injection showed much higher penetrance (100%) as opposed to that by feeding (20%), clearly indicating

that ABCE is an essential gene. The phenotype is even more pronounced in *C. briggsae* when *C. elegans* ABCE double stranded RNA was injected (0-1% laid eggs developed into adult) (Figure 12). In *C. elegans*, approximately 5% eggs from the injected worms developed into sick sterile adults at 20 °C after one week (slow growth). Most of these hatched embryos arrested at L1 to L2 stage. These small "healthy looking" *C. elegans* adults frequently showed defects in germline in both male and hermaphrodite, with vacuolated nuclei. The dying adult looks disintegrated (Figure 13B). The *C. briggsae* G16 germline accumulated liquids with a few vacuolated nuclei (Figure 13 G, H and I). Given the fact that yeast ABCE interacts with nuclear pore complex protein Nup116, which is involved in nucleocytoplasmic transport in yeast (Ito et al, 2001), I performed RNAi on a transgenic strain generated with full-length HIS-72 protein fused with GFP. ABCE seems affect the nuclear distribution of HIS-72 (Figure 13 J, K and L).

Figure 11. RNAi phenotypes for *nhr-91* by feeding

RNAi against *nhr-91* was done with *eri-1* mutant for consecutive two generations (see text). (A) A slow-growing worm after three-day incubation at 20 °C, underdeveloped vulvae were shown. (B) Worm with defects in molting was shown. (C) A dead embryo was shown inside the uterus of a young adult. (D) Cysts were observed around the intestine for a young adult after RNAi.

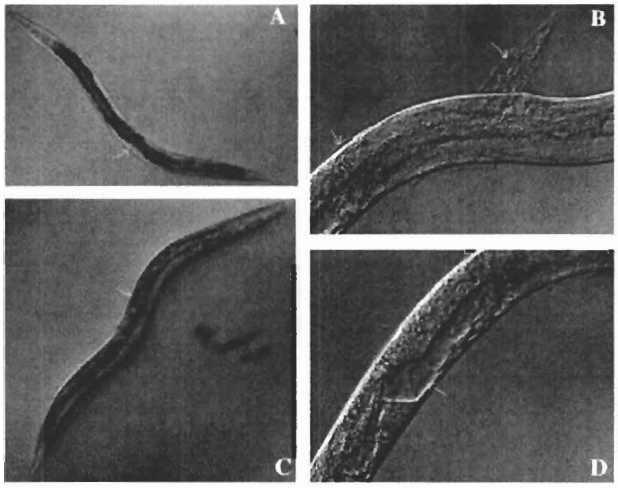


Figure 12. ABCE is an essential gene.

Double stranded RNAi targeting the first two exons of ABCE were injected into *C. elegans* N2 or *C. briggsae* G16 strain. Control worms were injected with injection buffer only. Around 20% laid eggs hatched and 5% of the laid eggs developed into sterile adults in *C. elegans*. RNAi phenotypes in *C. briggsae* are even more pronounced, with 12% laid eggs hatched and none of them develop into adults.

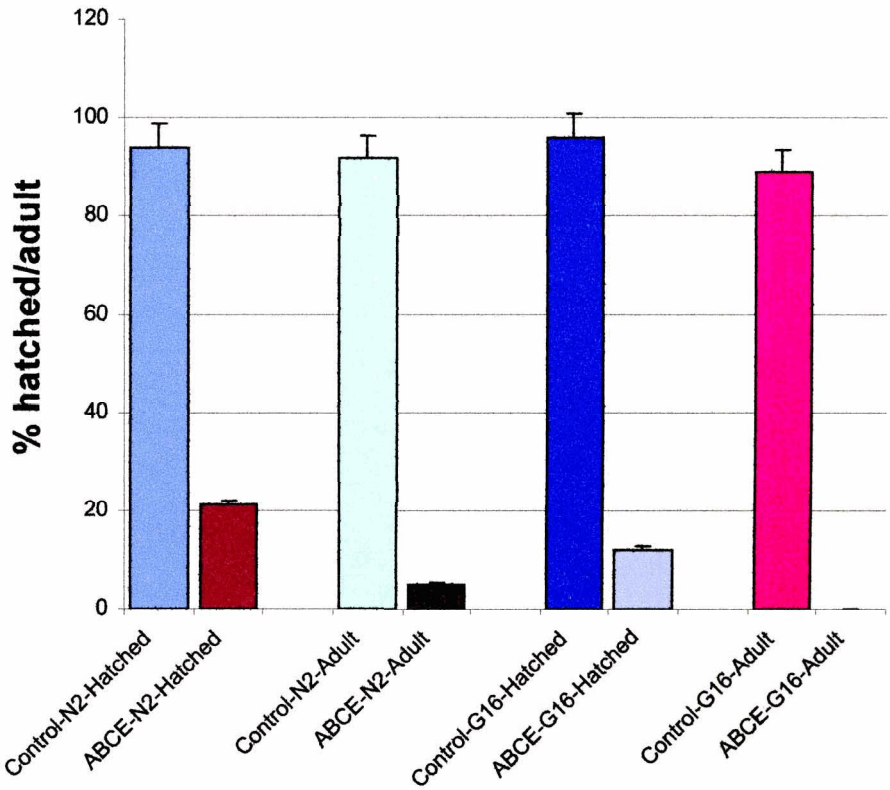
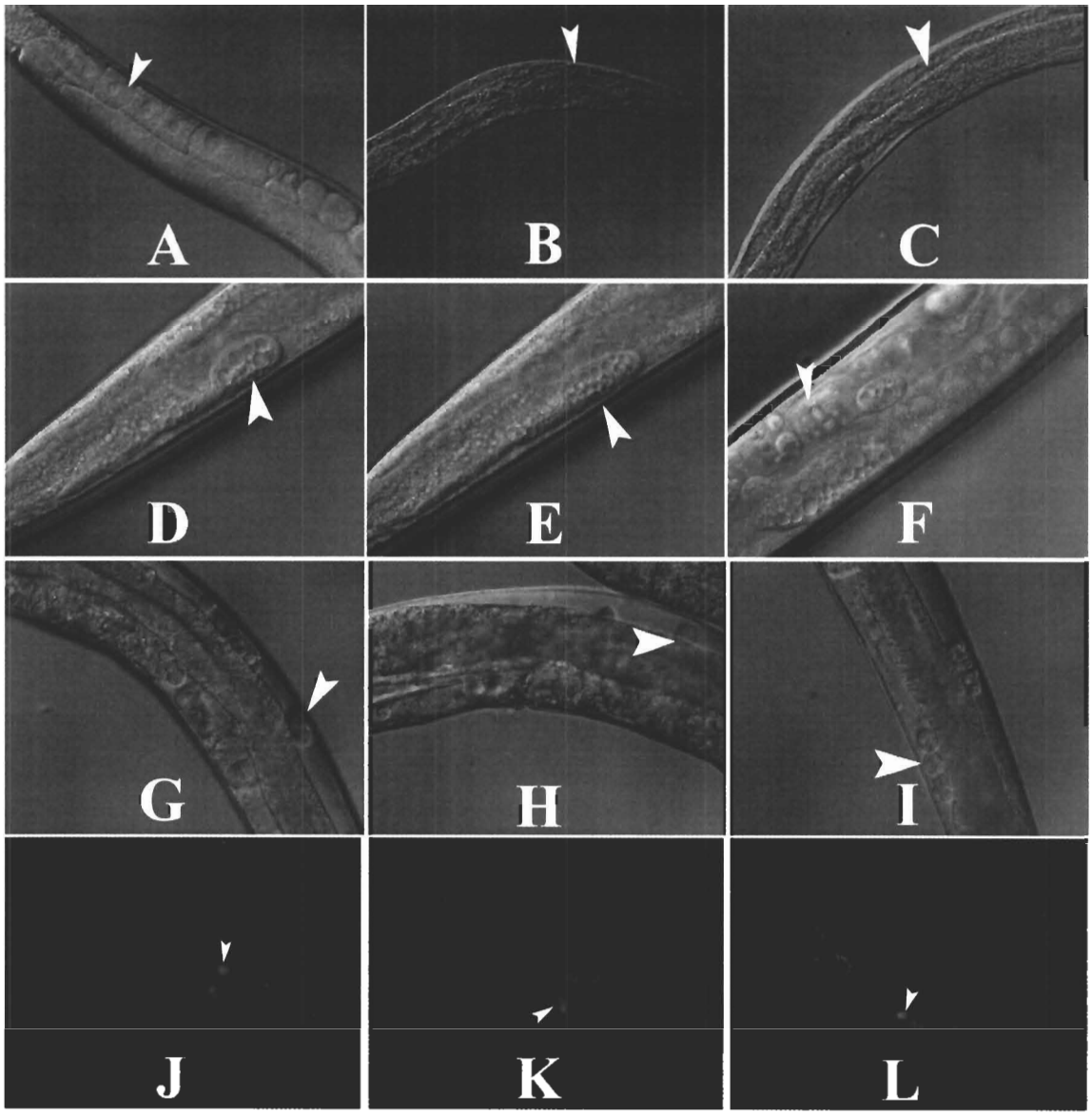


Figure 13. RNAi against ABCE causes defects in germline and nuclear protein trafficking

A-F: ABCE RNAi on *C. elegans* N2 strains. A: RNAi control (injection mixture only) with normal oocyte indicated by arrowhead. B: dying adult worms with body disintegrated. C: defective male germline. D and E: vacuolated hermaphrodite germline. F: Dead embryos with vacuolated nuclei inside. G-I: ABCE RNAi on *C. briggsae* G16. Germline full of liquid with few vacuolated nuclei. J-L: ABCE RNAi on *C. elegans* transgenic strain expressing GFP fused in frame to HIS-72. J: RNAi control with first two anterior intestine nuclei shown. They are "full" of GFP. K: ABCE RNAi caused black "hole" inside nuclei of first two anterior intestine nuclei. L: Similar "holes" in body wall muscle.



50 um

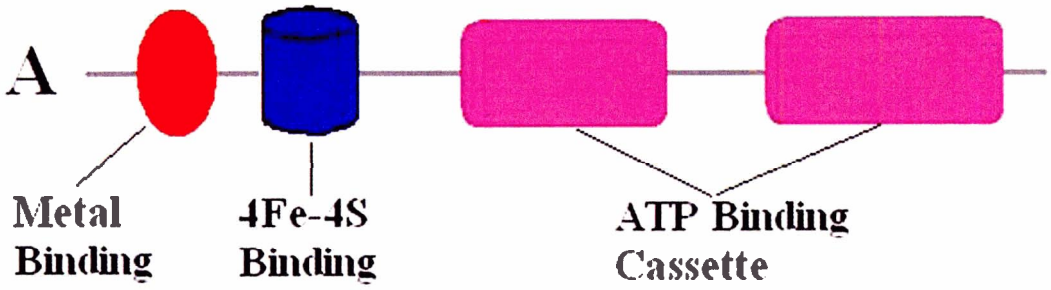
3.5 Discussion

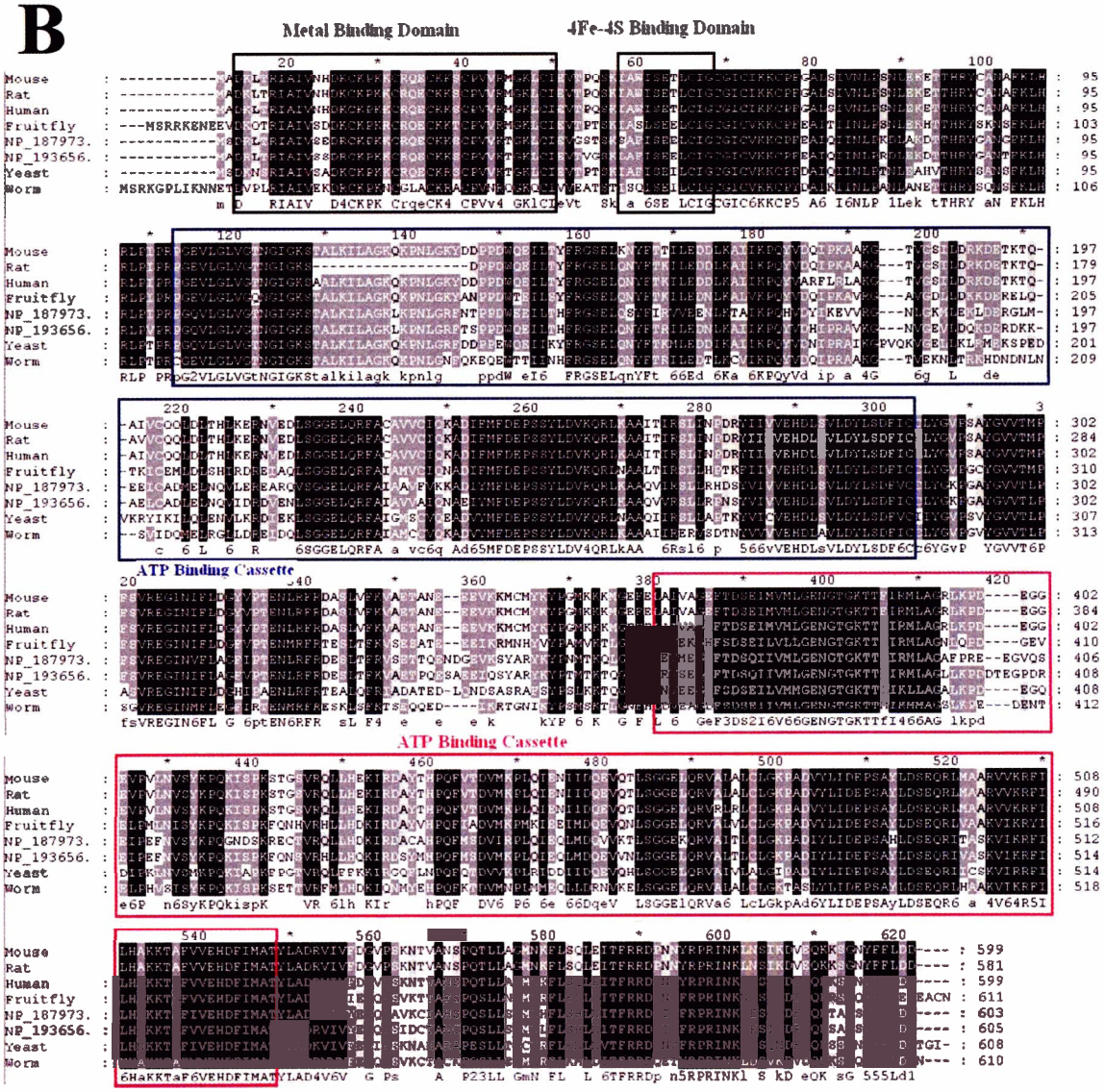
ABCE protein has been annotated as RNase L inhibitor in all sequenced eukaryotes because it has been demonstrated *in vitro* that this protein reversibly associates with and inhibits RNase L (Bisbal et al, 1995), although the exact mechanism of inhibition remains obscure. All eukaryotic species show the ubiquitous presence and high degree conservation of ABCEs; however, RNase L is present only in mammals. This implies that ABCEs may function as an RNase L inhibitor in mammals, but it may also have other functions in non-mammalian organisms.

As an initial step to investigate other potential roles of ABCE, I performed both *in silico* and functional experiments to investigate what roles ABCE might be involved in. Scanning of *C. elegans* ABCE protein sequence with Pfam yields four domains: possible metal-binding domain, 4Fe-4S binding domain as well as two ABC domains from N-terminal to C-terminal (Figure 14). Function of the metal binding domain remains to be determined. Proteins containing 4Fe-4S binding domain include bacterial ferredoxins, various dehydrogenases, and various reductases. It is speculated that 4Fe-4S binding domain may mediate interaction with DNA or RNA due to the two conserved positively charged lysines within the domain (Gabaldon et al, 2004). Two ABC domains are well conserved across all species. However, absence of transmembrane domains makes ABCE unlikely as transmembrane transporter, which is common for most ABC transporter proteins.

Figure 14. ABCE protein is well conserved across species

(A) Schematic representation of domain organization for ABCE based on Pfam annotations. *C. elegans* ABCE protein contains 610 residues. Residue 17–48 (red) at N-terminus is annotated as a metal-binding domain, 59–68 as 4Fe–4S binding domain (blue), and 114–303 and 381–548 as two ABC domains (pink). (B) ABCE protein sequences from different species are aligned by CLUSTALX (see Materials and methods). Except for *A. thaliana*, there is only one ABCE in each species. The GeneDoc was used to edit the resulted alignment. The corresponding domains in (A) are boxed accordingly. Sequences are strongly conserved not only for annotated domains, but also for sequences outside these domains.


A
Metal Binding
4Fe-4S Binding
ATP Binding Cassette


B
Metal Binding Domain
4Fe-4S Binding Domain
ATP Binding Cassette
ATP Binding Cassette
Mouse
Rat
Human
Fruitfly
NP_187973.
NP_193656.
Yeast
Worm

Extremely high conservation and lack of paralogues (Table 3) suggests ABCE plays essential roles among all eukaryotic species. This was supported by RNAi data in *C. elegans*, i.e. embryonic lethality. The fusion PCR products of ABCE promoter with GFP or RFP are so toxic that their concentrations in injection mixture were diluted to 10 to 100 times lower (100 pg/μl) than that used for regular injection to get successful transgenic animals, i.e., 10ng/μl (data not shown), suggesting the essential functions of ABCE. Constitutive expression of ABCE in rice also implies the essential roles of the protein in plant (Du, et al, 2003).

Three components were found to interact with ABCE protein (Y39E4B.1), namely, RPL-4, NHR-91 and C07B5.3 (Li et al, 2004). Co-localization of ABCE with two out of the three proteins confirms the interaction between ABCE and RPL-4/ NHR-91. *Rpl-4* encodes a large ribosomal subunit L4 protein. Yeast orthologous RPL4 may be one of the first ribosomal proteins that bind to the 35S pre-rRNA molecule during ribosome biogenesis (Yeh and Lee 1998). The interaction of and co-localization of ABCE with rpl-4 suggests that ABCE is involved in protein translation control. This agrees with the observations that human ABCE interacts with initiation factor 2 (Tyzack et al, 2000), which is also a part of translation machinery. However, low expression of ABCE as opposed to that of RPL-4 by SAGE data suggests that ABCE works as a regulator rather than as a constitutive component of ribosomes. The exact mechanism of ABCE in protein translation regulation remains to be further investigated. *Nhr-91* is a member of Nuclear Receptor (NR) superfamily. The NR superfamily has undergone a dramatic expansion and diversification in *C. elegans* but this has not been observed in other phyla (Sluder et al, 1999). The completed *C. elegans* genome contains 284 confirmed or predicted NR

genes, over 5-fold more than the number found in the human or *Drosophila melanogaster* genomes (Maglich et al, 2001). Among these 284 NR genes, 15 are conserved among the metazoan. *Nhr*-91 is one of these 15 conserved NR genes (Gissendanner et al, 2004), suggesting its significant roles in gene expression regulation. Interaction and co-expression of ABCE and NHR-91 (Figure 10) suggests that the two genes are involved in the same conserved pathway of gene expression regulation in *C. elegans*.

All NRs share a common structure with specific, conserved domains that function in DNA binding, ligand binding, and transcriptional activation (Gronemeyer and Laudet, 1995). The exact mechanisms of NHR-91 in transcriptional regulation remain to be determined, but because of its good conservation across species, it is reasonable to speculate that NHR-91may function as a transcription factor in a conserved pathway (Birchall et al, 1995). Given the interaction of ABCE between both cytoplasmic RPL-4 and nuclear NHR-91, ABCE may function in transcription regulation and communications between gene transcription and translation, possibly mediating feedback regulation between gene transcription and translation. This hypothesis is supported by PSORT (Gabaldon et al, 2004) prediction of worm ABCE: 37% possibility in nucleus and 20% in cytoplasm, suggesting it to be a shuttle protein between cytoplasm and nucleus. Consistent with the observation, yeast ABCE interacts with mRNA exporter protein (Ito et al, 2001), suggesting ABCE may be involved in mRNA export. Evidence from conserved co-regulation suggest ABCE is involved in biogenesis of ribosome. For example, *S. cerevisiae* and *C. elegans* ABCE are both co-expressed with a number of proteins involved in the processing of rRNA such as the nucleolar protein SIK1 (NOP56) from yeast that is involved in rRNA methylation (van Noort et al, 2003), implying

ABCE's role in ribosome biogenesis (protein translation). Knowledge from genomic context data and homology also predicts the possible function of ABCE in both translation and transcription. Among the 44 orthologous groups, which have an identical phylogenetic distribution with ABCE, most of them are either involved in translation or ribosome biogenesis (33, 60 %), in transcription (7, 16%) and in DNA replication, recombination and repair (3, 7%) (Gabaldon et al, 2004). Recent investigation of mRNP movement from the chromatin to the nuclear pore showed that there is an active component involved in the nuclear motion of these mRNP particles, and mRNP mobility was ATP dependent (Calapez et al, 2002). Together with the expression data, it is possible that ABCE is involved in transport of mRNA from nucleus to cytoplasm.

RNAi with *eri-1* (mg366) mutant against *nhr-91* gene gave significantly slow growing worms as opposed to wild type ones with N2 or rrf-3 strain. Expression of both ABCE and *nhr-91* is seen both in hypoderm and vulvae. RNAi with *nhr-91* on *eri-1* mutants also causes defects in vulvae and molting. The external surface of *C. elegans* is formed by hypodermis, which establishes the basic body form of the animal, acts in nutrient storage, secretes the cuticle and takes up apoptotic cell bodies by phagocytosis. Together, the two genes possibly function synergistically in vulval development and molting pathway. RNAi by injection gave higher penetrance of phenotypes than that by feeding, probably due to obvious neuronal distribution of ABCE based on GFP expression data, implying that the *eri-1* mutant is still resistant to RNAi by feeding, even it show improved susceptibility compared to N2 strain. Phenotypes captured using *eri-1* but not N2 suggest functional association of NHR-91 with nerve system. The underlying mechanisms of the association need to be further investigated. RNAi by injection also

demonstrated that ABCE is an essential gene. Interestingly, defective germline was frequently observed in both male and hermaphrodite, but not in those of feeding progeny. It seems ABCE plays significant roles in maintaining integrity of nuclear membrane in both germline and embryo. Changed nuclear distribution of HIS-72 suggest ABCE may be involved in importing nuclear protein from cytoplasm, which is in agreement with the prediction that ABCE is present in both cytoplasm and nucleus.

No obvious orthologues were identified for C07B5.3 in other phyla except in nematode (data not shown), indicating its unique roles in nematodes. It showed weak expressions in adult gut, which is different from what was reported previously (Birchall et al, 1995). Based on these data, interaction of ABCE and C07B5.3 is possibly a false positive due to lack of overlapping expression data between the two genes.

Like ABCE, ABCF proteins containing two ABC domains but no transmembrane domains have recently been proposed to be involved in translation initiation and elongation (Kerr 2004). Investigation of functional correlation of the structurally related ABCE and F proteins will provide new insights about the roles of both ABC subfamilies in eukaryotic species.

I generated full-length protein fusion of ABCE (including 2.7 kb of its upstream sequence) to RFP. The fusion construct was too toxic to get successful transgenic lines, implying overproduction of ABCE causes lethality during worm development.

CHAPTER 4 EXPRESSION ANALYSIS OF ABC TRANSPORTERS IN *C. ELEGANS* [1]

[1]This chapter has been published in *Journal of Molecular Biology*, 2004 under the co-authorship of Lily Fang, Jonathan A Sheps, Victor Ling and David L Baillie

4.1 Abstract

Sixty ABC transporter genes have previously been identified in the *C. elegans* genome and classified into 8 groups. As an initial step towards understanding how these putative ABC genes work in worms, I generated promoter-Fluorescent Protein fusions for the entire family to address when and where these genes are turned on *in vivo.* Both Green Fluorescent Protein (GFP) and DsRed Protein (RFP) were used as reporters in the transgenic assay. Promoter driven GFP and RFP expression patterns are highly reproducible for most ABC genes. Observable expression is more frequently seen in fusions to genes in subfamilies B, C, D and E than those in subfamilies A and G. As described in the previous chapter, sixteen worm ABC genes are found in tandem duplications, forming two four-gene clusters and two four-gene-clusters. Fifteen out of the 16 duplicated gene promoters drove different or overlapping expression patterns, suggesting active functions for these duplicated genes. The duplicated ABC genes may work more efficiently than its single ancestor do by partitioning the roles of the latter. Furthermore, the results suggest that an internal promoter can cause differential expression of genes within an operon. Finally, I found that coding sequences could function as a regulatory region for a neighboring gene.

4.2 Introduction

As described in the previous chapter, ABC transporters constitute one of the largest protein families in both prokaryotes and eukaryotes. Given the large size of the family and a variety of roles in which ABC genes are involved, it would be desirable to choose a model system to systematically investigate the function of the family. *C. elegans* is an excellent model animal for this purpose. It has a short life cycle (around 3 days at room temperature) and small size (about 1 mm in length), which makes it amenable to genetic and molecular analyses. Its transparent body makes it feasible to develop an expression marker like GFP to keep track of gene activities *in vivo*. The whole genome sequencing has been finished (*C. elegans* genome sequencing consortium, 1998). The completely mapped cell lineage (Sulston, et al., 1983) allows investigation of gene functions in anatomical detail.

Sixty ABC transporters have been identified in *C. elegans* with members grouped into each of the eight established subfamilies (Figure 16, Table 5). Only a few ABC transporters have been functionally characterized in *C. elegans*. For example, the expression of two P-glycoprotein (*pgp*) genes (subfamily ABCB) in transgenic *C. elegans* is confined to intestinal cells (Lincke, et al., 1993). *Pgp-1* (K08E7.9) and *pgp-3* (ZK455.7) are responsible for resistance to chloroquine and colchicine (Broeks, et al., 1995). ABCC subfamily member *mrp-1* (F57C12.5) and *pgp-1* are both involved in heavy metal resistance (Broeks, et al., 1996). The ABCA subfamily gene *ced-7* (C48B4.4) encodes a protein involved in cell corpse engulfment (Wu, et al., 1998). Little investigation has been done regarding functions of the remaining members.

Much of the genomes of flies and worms consist of duplicated genes (Rubin, et al., 2000). It is becoming common to find computational investigations into the evolution of genes or gene families with the increasing number of genome sequences available (Robertson, 2000; Harrison, et al., 2001; Copley, et al., 2002). However, in many of these studies genes in the same family are assumed to be functionally redundant, or similar in function, or else "dead", that is to say, pseudogenes. Conservation of coding sequences sometimes might be misleading if the regulatory regions have evolved differentially. Functional genomic methods, such as microarray, serial analysis of gene expression (SAGE) and RNA interference (RNAi), provide tremendous amounts of functional data. However, these methods cannot give detailed functional information, such as tissue-specific distribution, which is essential to understand the different roles of recently duplicated paralogues. The high-throughput transgenic assay with promoter::GFP fusions provides a sensitive means to address functions of gene family members, especially those of recently duplicated genes within a gene family. In particular, this method provides an *in vivo* expression profile both spatially and temporally. ABC transporters in *C. elegans* provide an excellent system to address these questions because there are 16 tandemly positioned ABC genes in two-gene or four-gene clusters on different chromosomes (Figure 16). This will also help answer whether these duplicated ABC genes are functionally redundant or complementary to each other. The presence of tandemly duplicated genes is not unique to the worm ABC genes. For instance, there are five-gene ABCA clusters in both human and mouse (Annilo, et al., 2003). As a preliminary investigation to address the roles of ABC transporter family in *C. elegans* I have performed transgenic expression analysis of all ABC transporters in *C. elegans*, using promoter-driven GFP or DsRed Fluorescent Protein (RFP) reporters. This method plays

an irreplaceable part alongside other experimental tools like RNAi, SAGE or Microarray. In particular, it can provide an *in vivo* expression profile both spatially and temporally. These results provide an overview of the expression patterns for the whole ABC transporter family and provide some insights into the roles of tandemly duplicated genes.

4.3 Materials and methods

4.3.1 Strains

All strains were maintained and cultured using standard techniques.

4.3.2 Construction of promoter::marker fusion and microinjection

These were done basically as described in chapter 2. Microscopy was also performed as described in chapter 2.

4.3.3 Drug sensitivity assay

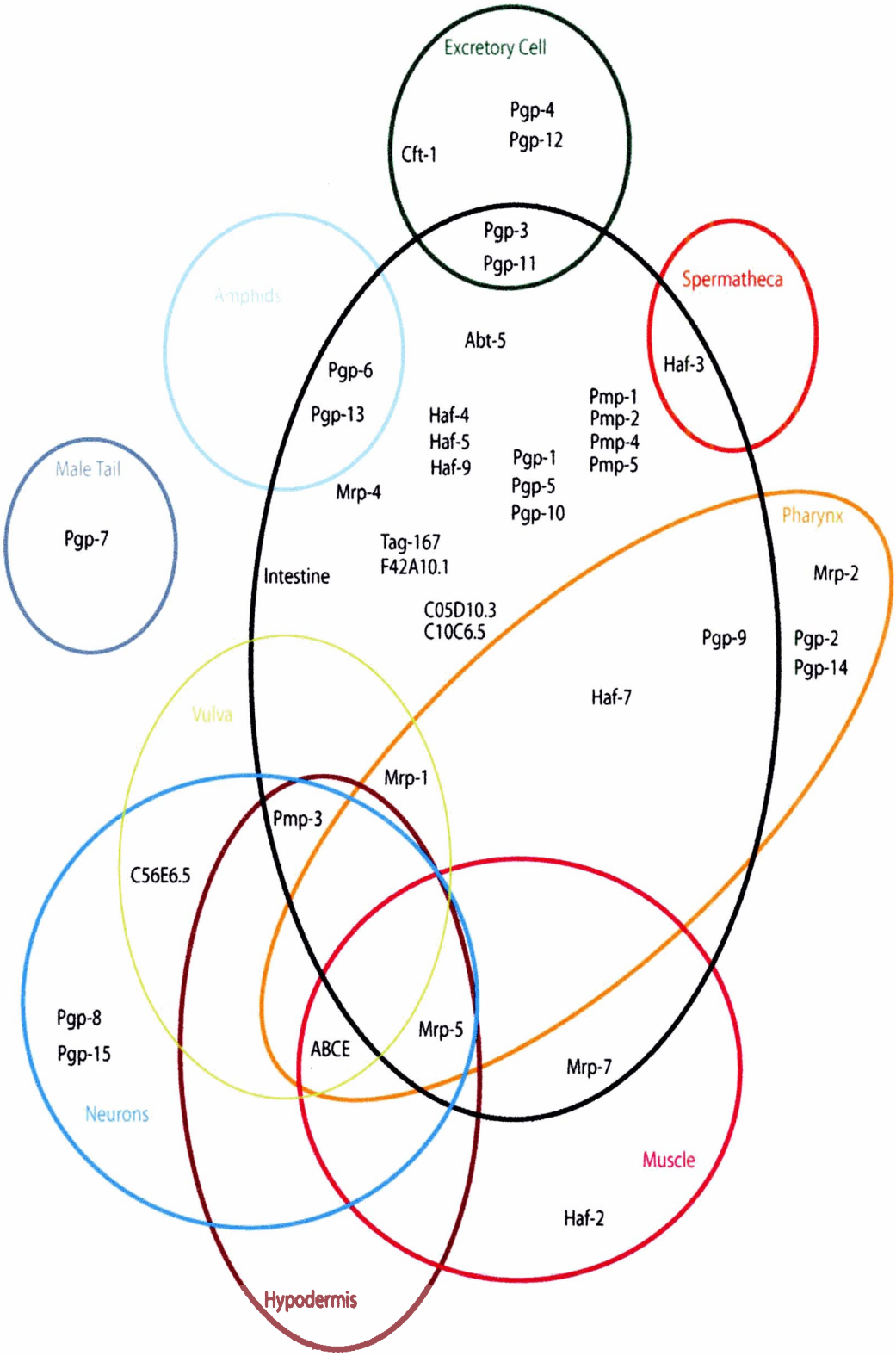
The assay was performed as described (Broeks et al, 1995). Chloroquine (C6228, Sigma) and colchicine (C9754, Sigma) were added to the NGM medium before pouring the plates at 10 and 1mM respectively. Each test was performed with three triplicates. Ten young adult N2 worms were placed on the plates and allowed to lay eggs for 3 hours. The parents were removed and eggs counted. After three days, viability was calculated as the percentage of adults out of the laid eggs. The scores were normalized to those on control plates (no drugs).

4.4 Results

All the promoters were derived from N2 Bristol genomic DNA by PCR. They consist of intergenic regions upstream of target gene start codons and in some cases, include coding sequence from adjacent upstream genes. Several factors are taken into account to make efficiently functioning expression constructs (see Materials and Methods). I have generated GFP transgenic strains for all of the 60 ABC genes in *C. elegans*. GFP constructs successfully produced visible expression patterns in 39 of these transgenic strains. RFP fusions were constructed for 19 ABC genes, 11 of which generated visible expression patterns. Almost all the RFP transgenic strains can faithfully reproduce the expression patterns that were observed from GFP assay (appendix 2), indicating the reporter itself do not make difference for promoter driven expression patterns. For the majority of genes their expression patterns are first reported here. The average intergenic space upstream of worm ABC transporter genes is 3118 bp and the average size of the putative promoter constructs was 1575bp. The largest promoter sequence used was 3422bp for F22E10.1; the smallest promoter construct was only 250bp for Y49E10.9 (See Appendix 2 for details). Up to five attempts to obtain successful expression were made for generating transgenic line for each ABC transporter gene. Transgene expression was frequently seen in the intestine, pharynx and excretory cell (Figure 15). Many of the genes in *C. elegans* are found in pairs, adjacent to their sister genes. The expression data for the whole ABC gene family provide functional insights into duplicated genes, especially for those tandem repeat ones.

Figure 15. Venn diagram showing the expression of 39 ABC transporters in different tissues of *C. elegans*

These data are mostly deduced from the expression of promoter-GFP (or RFP) fusion proteins. Not all genes shown are expressed in all, or the same, developmental stages. See Appendix 2 data for full details on the timing and relative strengths of expression observed. This figure is adapted from *Journal of Molecular Biology*, 2004 (by Zhao et al, 2004b)


Excretory Cell
Pgp-4
Pgp-12
Cft-1
Pgp-3
Pgp-11
Amphids
Spermatheca
Abt-5
Haf-3
Pgp-6
Pgp-13
Haf-4
Haf-5
Haf-9
Pmp-1
Pmp-2
Pmp-4
Pmp-5
Pgp-1
Pgp-5
Pgp-10
Mrp-4
Male Tail
Pgp-7
Pharynx
Tag-167
F42A10.1
Intestine
Mrp-2
C05D10.3
C10C6.5
Pgp-9
Pgp-2
Pgp-14
Haf-7
Vulva
Mrp-1
Pmp-3
C56E6.5
Pgp-8
Pgp-15
ABCE
Mrp-5
Mrp-7
Neurons
Muscle
Haf-2
Hypodermis

4.4.1 Tandemly duplicated ABC genes tend to be differentially expressed

It is noteworthy that sixteen worm ABC genes are tandemly arranged, forming two four-gene clusters and four two-gene clusters (Figure 16). Twelve of 16 ABCs on the X chromosome are present in tandem arrangement. Fifteen out of these sixteen ABC genes gave observable expression in the promoter driven GFP/RFP assays (see Table 5). Interestingly, all ABCs within the four-gene clusters showed differential tissue expression patterns while those within two-gene clusters gave similar or overlapping expression patterns, except for C56E6.1, which is arranged in a head to head orientation with C56E6.5, and yielded no observable expression. Among the 4 ABC genes clustered on cosmid clone F22E10 (having between 65-75% amino acid sequence identity), all are expressed in different tissues (Figure 17, Table 5). *Pgp-12* (F22E10.1) showed strong expression in the excretory cell in all stages; *pgp-13* (F22E10.2) yielded weak expression in the adult posterior intestine and amphid; *pgp-14* (F22E10.3) was expressed from the anterior-most pharynx until halfway into the first pharyngeal bulb while *pgp-15* (F22E10.4) is expressed in head and tail neurons in both adults and embryos. Another four-gene cluster includes *pgp-9* (C05A9.1), *pgp-6* (T21E8.1), *pgp-7* (T21E8.2) and *pgp-8* (T21E8.3). They are all expressed in different tissues and/or stages (see Table 5). The expression patterns for *mrp-2* (F57C12.4) and *mrp-1*(F57C12.5) overlap, while stage specific expression is observed for *pmp-1*(C44B7.8) and *pmp-2*(C44B7.9) though in the same tissues (appendix 2). Most of the duplicated ABCs are members of the ABCB subfamily (10 out of 16), including all ABCs within four-gene clusters (F22E10.1-4, C05A9.1 & T21E8.1-3) and two of the duplicated ABCs transcribed in the same

orientation, *pgp-3*(ZK455.7) and *pgp-4*(F42E11.1). The observed expression patterns, in combination with EST data, suggest that none of these tandemly duplicated ABC genes are pseudogenes. In other words, they all are functional *in vivo*. ABCB genes found in tandem duplications are most often seen expressed in the worm's pharynx, gut or excretory cell. It has previously been described that some members of the ABC subfamily are involved in drug resistance (Broeks, et al., 1995;1996). These duplicated ABC genes might provide the worm better protection against xenobiotics. Few of the duplicated worm ABCs show a visible RNAi phenotype (Kamath, et al., 2003), suggesting that most ABCs are not essential for normal development. One exception is C56E6.1. It has few ESTs and gave no observable expression, but yields an RNAi phenotype (larval arrest). Its detailed function has yet to be determined.

Table 5. Expression patterns and similarity between tandemly duplicated ABC genes

Cluster	Gene Name	Expression Pattern	E-value
1	C56E6.1	No observable expression	5.3e-14
1	C56E6.5	Larva & adult head and tail neuron and vulvae	5.3e-14
2	C44B7.8	Strong larva & weak adult intestine	3.3e-252
2	C44B7.9	Strong larva & adult intestine	3.3e-252
3	F57C12.4	Adult pharynx	0
3	F57C12.5	Adult pharynx, intestine, vulva	0
4	T21E8.1	Larva & adult intestine, embryo	0
4	T21E8.2	Adult male tail rays	0
4	T21E8.3	Adult head neuron	0
4	C05A9.1	Larva & adult anterior intestine, embryo	0
5	F42E11.1	Larva excretory cell	0
5	ZK455.7	Larva & adult excretory cell & intestine	0
6	F22E10.1	Larva & adult excretory cell, embryo	0
6	F22E10.2	Adult posterior intestine and amphid	0
6	F22E10.3	Larva & adult pharynx	0
6	F22E10.4	Adult head & tail neurons, embryo	0

Figure 16. Schematic representation of 16 tandemly duplicated ABC genes on chromosomes II and X of the *C. elegans* genome.

These form six clusters, which are illustrated using images drawn from Wormbase release WS130 (http://wormbase.org/). The two ABCH subfamily genes identified within cosmid clone C56E6 are the only ones found in a head-to-head arrangement. All other clusters contain at least two members in a head-to-tail orientation. Pmp-1 and -2 are members of subfamily ABCD, Mrp-1 and -2 are members of subfamily ABCC, while all Pgp genes are members of the ABCB subfamily. This figure is adapted from *Journal of Molecular Biology*, (by Zhao et al, 2004b)

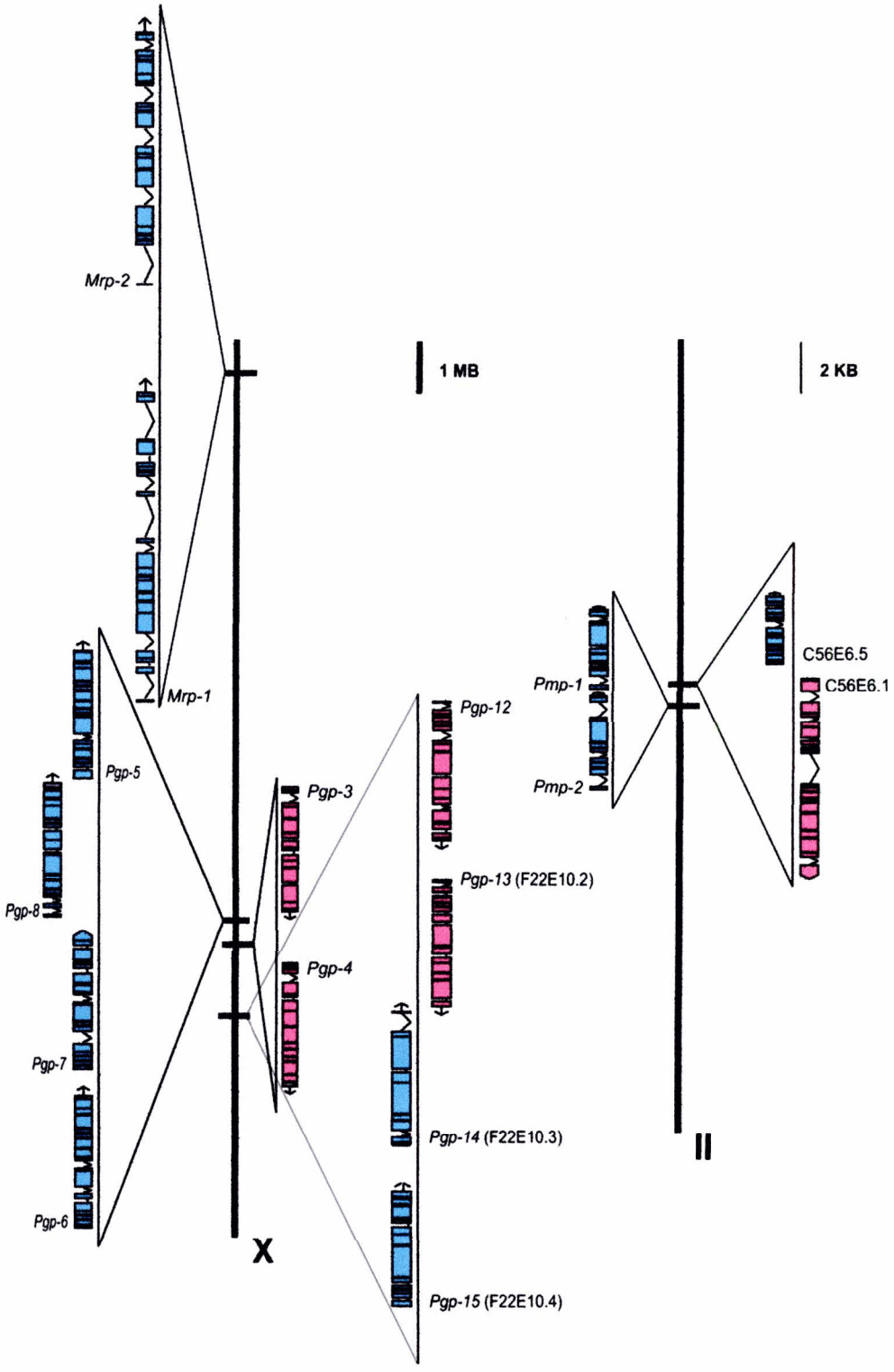

Mrp-2
Mrp-1
1 MB
2 KB
Pgp-5
Pgp-8
Pgp-7
Pgp-6
Pgp-3
Pgp-4
Pgp-12
Pgp-13 (F22E10.2)
Pgp-14 (F22E10.3)
Pgp-15 (F22E10.4)
X
Pmp-1
Pmp-2
C56E6.5
C56E6.1
II

Figure 17. Photomicrographs of GFP or RFP expression driven by promoters derived from the four ABC genes in a four gene-cluster.

Promoter–GFP fusion constructs were built as described (Hobert, 2002). Pgp-12 (F22E10.1) showed strong expression in excretory cell in all stages (only adult shown); Pgp-13 (F22E10.2) yielded expression in the adult posterior intestine and amphids; Pgp-14 (F22E10.3) gave expression in the anterior pharynx until half way along the first bulb while Pgp-15 (F22E10.4) was seen in adult head and tail neurons and in the embryo. This figure is adapted from *Journal of Molecular Biology*, 2004 (by Zhao et al, 2004b)

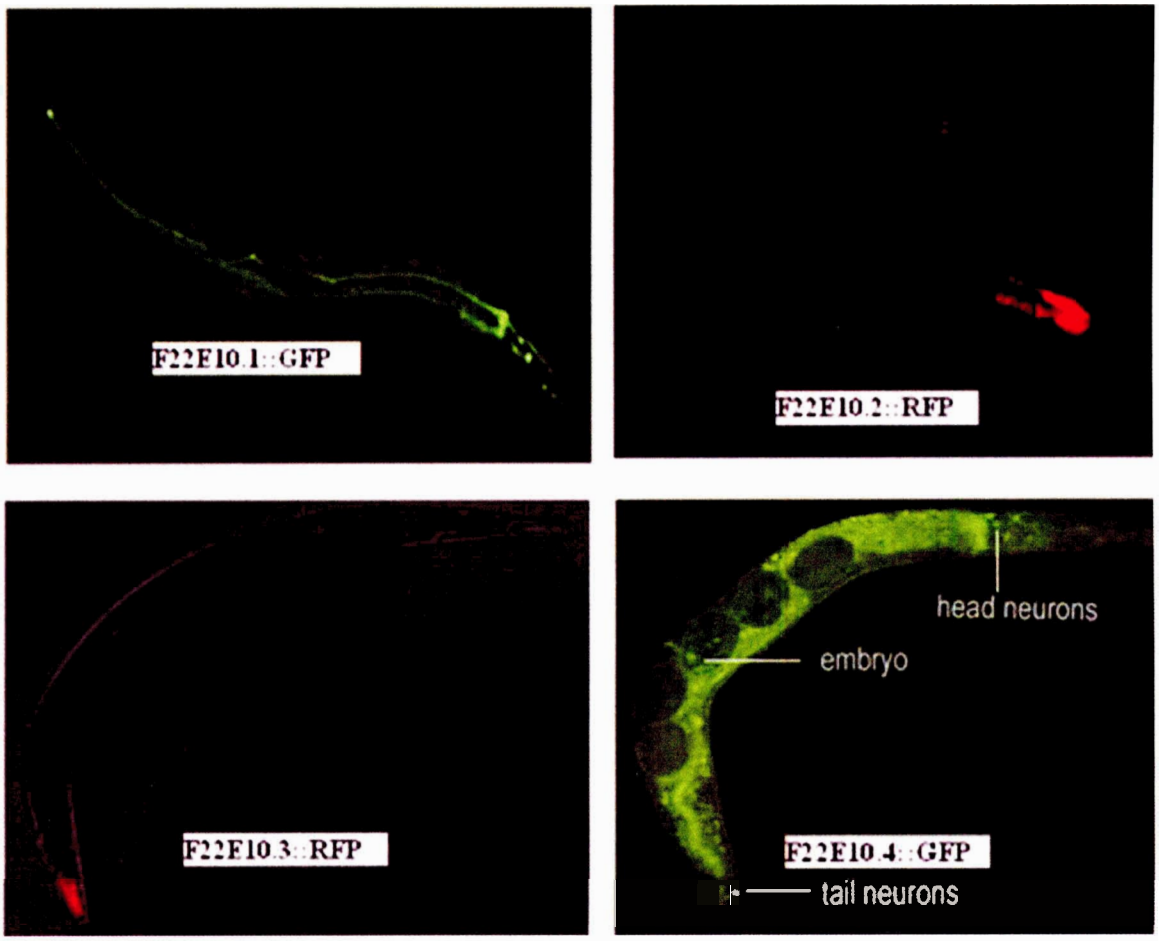
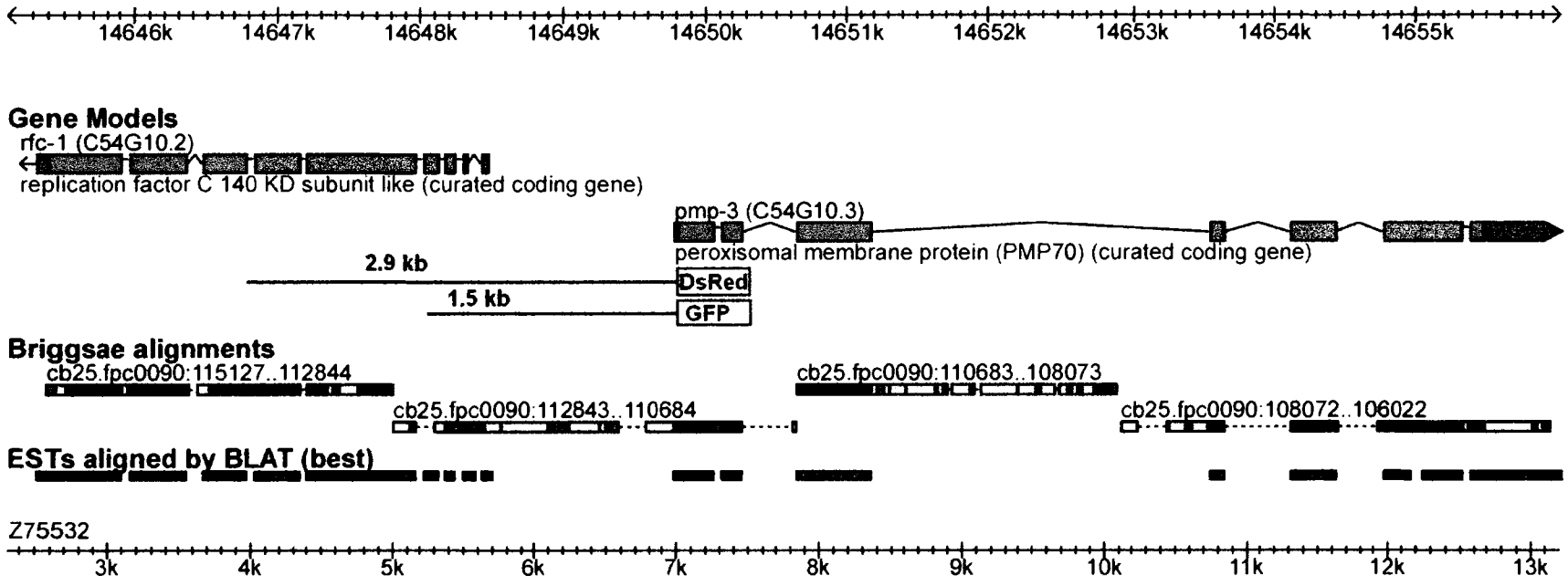


Figure 18. Coding sequences of neighbouring genes contain regulatory elements
The picture was generated using the Genome Browse tool on the Wormbase web site (release WS130, available at http://www.wormbase.org/). The top and bottle line is a sequence scale in kilobase pairs. Pink bars represent exons; gray bars 5' or 3' UTR and black lines introns. Small discontinuous bars denote cDNAs (from Yuji Kohara's collection) aligned to the corresponding genomic region. Small discontinuous blue/gray bars represent *C. briggase*'s genomic region aligned to corresponding *C. elegans* region. The putative promoters used for DsRed and GFP fusion are indicated. The promoter sequence used for DsRed fusion was extended into the coding sequence of its upstream gene, C54G10.2 as indicated by the scaled black line. The extended promoters were found to effectively drive strong RFP expression in gut and neurons in either the larval or adult stages while intergenic region gave no observable expression.


14646k
14647k
14648k
14649k
14650k
14651k
14652k
14653k
14654k
14655k
Gene Models
rfc-1 (C54G10.2)
replication factor C 140 KD subunit like (curated coding gene)
pmp-3 (C54G10.3)
peroxisomal membrane protein (PMP70) (curated coding gene)
2.9 kb
DsRed
1.5 kb
GFP
Briggsae alignments
cb25.fpc0090:115127..112844
cb25.fpc0090:110683..108073
cb25.fpc0090:112843..110684
cb25.fpc0090:108072..106022
ESTs aligned by BLAT (best)
Z75532
3k
4k
5k
6k
7k
8k
9k
10k
11k
12k
13k

4.4.2 Coding sequences of neighboring genes contains regulatory elements

It is common in *C. elegans* that intergenic spaces are too small to contain a reasonable promoter. It has not been determined how large the promoter really is for each gene. In many cases, it is possible that a given intergenic region is not enough to drive reporter expression, while the coding sequence of the adjacent gene might contain regulatory sequences for the downstream gene. For example, two ABC genes, F42A10.1 and *pmp-3* (C54G10.3) both have many ESTs and SAGE tags (McKay, et al., 2003) and have been successfully amplified by the ORFeome project (Reboul, et al., 2001), but failed to give observable GFP expression when integenic sequences alone were used as putative promoters. I reasoned that the intergenic region might not be the real promoter, or might not be the complete promoter for these two genes. The intergenic sequences I used for GFP fusions were initially 858bp and 1519bp for F42A10.1 and *pmp-3*, respectively. I then included 2928 and 2914bp, respectively, of intergenic sequences, including partial (opposite strand) coding sequences from the neighbouring genes, as the putative promoters for F42A10.1 (data not shown) and *pmp-3* (Figure 18). Both extended promoters were found to effectively drive strong RFP expression in gut and neurons in either the larval or adult stages. This suggests that transcriptional elements in *C. elegans* could be found buried within coding sequences.

4.4.3 ABCs on chromosomes V and X produce more visible expression patterns than those on chromosomes III and IV

I have generated transgenic strains for 60 worm ABC genes. A similar expression assay has been reported for three of them, i.e., *pgp-1*, *mrp-5* (F14F4.3) and *mrp-*

1(F57C12.5) (Broeks, et al., 1995; Mounsey, et al., 2002). In total, 39 out of 60 (65%) ABC genes gave observable expression patterns. However, successful expression was more often seen for ABCs on chromosomes V and X as opposed to those on chromosomes III and IV. Only one of nine ABCs on chromosome V did not show observable expression, and neither did two of the 17 ABCs on chromosome X. However, only four of 12 ABCs on chromosome III and three of 8 ABCs on chromosome IV yielded observable expression. The success rate for transgenic expression also varied among subfamilies. For instance, six out of seven ABCA genes and seven out of nine ABCG members did not give any observable expression. However, 22 out of 24 ABCB members did show expression. Interestingly, there are no local duplications within the ABCA subfamily whereas 10 tandemly duplicated genes are found in the ABCB subfamily. It has been shown, for a subset of *C. elegans* genes, that transgenic expression has a low success rate for evolutionarily recently duplicated genes (Mounsey, et al., 2002). However, according to the criteria used in that paper, 15 of the 24 ABCB genes in *C. elegans* are the products of recent duplications, whereas only four of seven ABCA and one out of nine ABCG genes can be described as recently duplicated (data not shown). This observation reinforces the impression, already explored in our previous paper (Sheps, et al., 2004) that the evolutionary dynamics and functional inter-relationships of the ABC transporter family are rather atypical of multi-gene families in general.

4.4.4 Effect of gene organization on expression

Among the 60 ABC genes, 30 of them are transcribed in diverging orientation from a common promoter region shared with the adjacent upstream gene, which I refer to

as a Head-to-Head orientation (Table 6). For these 30 genes, the average intergenic size is 4063bp; and the average segment used as a promoter in the expression constructs was 1579bp. Almost half of these genes (16 out of 30) did not give any observable GFP expression. The remaining 30 ABC genes are in a Head-to-Tail orientation; their average intergenic size is 2173bp; the average promoter construct derived from them is 1572bp; and only 5 of these genes failed to yield observable expression. It is not clear why the layout of the source genes in the genome has such an influence on the efficiency of the isolated promoters, but I speculate these results indicate the importance of *cis*-element orientation and position in the control of expression.

4.4.5 Most single ABC gene mutants show little or modest resistance to chloroquine or colchicine

As shown in table 7, there are in total 18 strains carrying single, 3 strains double and one strain triple ABC gene mutation in *C. elegans*. Most of the single gene mutant strains are generated by *C. elegans* Knockout Consortium. However, no one of them shows obvious phenotypes under normal growth conditions except for the *pgp-7* mutation, which is lethal. I found that the lethality is caused by mutations closely linked to *pgp-7* based on the following observations: first, a *pgp-6, 7* double mutant looks like wild type; second, cosmid T21E8 which covers the full length genomic DNA of pgp-7 cannot rescue the lethality. So there must be some mutations within the proximal essential genes that cannot be removed by out-crossing. *pgp-3* but not *pgp-1* deletion mutant is sensitive to both chloroquine and colchicine (Broeks et al, 1995). So I perform ed the similar drug screen using *pgp-1* or *pgp-3* single gene mutants in order to set up the screen

for the remaining pgp mutants. Unfortunately, I could not repeat the results reported by Broeks et al. Only modest increased sensitivity (17%) against chloroquine was detected for the *pgp-3* mutant strain compared to N2. It is yet to be determined what caused the discrepancy. Similar screens were performed using the strain that carries mutation in pgp-2, 4, 10, 12, 13 or 15. Again, only pgp-10 and 12 mutant strains show slightly higher sensitivity (16%) to chloroquine than that of N2 strain. No one was found showing increased sensitivity to colchicine.

Table 6. Effect of gene organization on expression

	Number of genes	Average intergenic size	Average promoter size	Number of expressed genes
H-H[a]	30	4063	1579	14
H-T[b]	30	2173	1572	25

a: Head-to-Head organization

b: Head-to-Tail organization

Table 7. List of ABC mutants and their phenotypes

1: from Plasterk lab; 2: phenotypes under normal lab conditions; 3: drug resistance against colchicines and chloroquine; 4,5 and 6: excretory cell (exc), colchicines (col) and chloroquine (chlo) respectively; 7: mutant showed small brood size and became wild type after out-crossing; 8: not determined; 9: mutant is lethal but cannot be rescued by the cosmid covering the full length of genomic sequence; 10: data from Wu et al, 1998.

Gene Name	Phenotype[2]	Localization	Drug resistance[3]	BC #
pgp-1[1]	WT	gut	No	BC5880
pgp-2	WT	gut	No	BC6176
pgp-3[1]	WT	exc[4], gut	col[5]/chlo[6]	BC5879
pgp-4[7]	WT	exc	chlo	BC6474
pgp-5	WT	gut	ND[8]	BC6444
pgp-7	Lethal[9]	male tail	ND	BC6450
pgp-10	WT	gut	chlo	BC6452
pgp-12	WT	exc	chlo	BC6476
pgp-13	WT	gut	No	BC6406
pgp-15	WT	neurons	No	BC6447
haf-2	WT	pharynx	ND	BC6176
haf-3	WT	gut,spermatheca	ND	BC6461
haf-5	WT	gut	ND	BC6366
haf-7	WT	pharynx,gut	ND	BC6175
haf-8	WT	No	ND	BC5925
haf-9	WT	gut	ND	BC 6174
mrp-1[1]	WT	gut,pharynx	ND	BC5881
ced-7	WT	pharyngeal valve[10]	ND	BC6164
pgp-1,3	WT	gut,exc	col/chlo	BC5878
pgp-1,3+mrp-1	WT	gut,exc	ND	BC5882
pgp-6,7	WT	gut,male tail	ND	BC6176
haf-2,8	WT	pharynx	ND	BC6095

4.5 Discussion

Much of the genomes of both flies and worms consists of duplicated genes. Tandemly or locally duplicated genes are more often seen in the worm than in the fly genome (Rubin, et al., 2000). It is generally accepted that duplicated paralogs are under little selection pressure and will most usually end up "dead" as pseudogenes. *C. elegans* ABC genes constitute an excellent system in which to test this hypothesis since sixteen tandemly duplicated ABC genes are found in the genome. One way to address this is to examine whether these duplicated genes are active *in vivo* or not by a promoter driven reporter assay. I found that fifteen out of 16 tandem ABC genes could effectively drive GFP or RFP expression *in vivo*, which strongly argues against the foregoing hypothesis.

Duplicated ABCs are most often seen in subfamily B. Ten out of the 16 tandemly duplicated ABCs are members of subfamily B. Two genes in this subfamily, *pgp-1* and *pgp-3*, have been demonstrated to be responsible for drug resistance (Broeks, et al., 1995). It is tempting to speculate that many other B subfamily members are involved in the similar functions. The tissue-specific expression patterns of these duplicated ABC transporters suggest that the individual genes have evolved to perform differential functions after the expansion of the cluster. The classical model predicts that the most common fate for the duplicate pair should be the formation of a pseudogene at one of the duplicate loci (Ohno, 1999; Watterson 1983). This seems not the case for duplicated ABC genes. However, our observation is in agreement with a recent hypothesis, which predicts that the usual mechanism of duplicate gene preservation is the partitioning of ancestral functions rather than the evolution of new functions (Force, et al., 1999). The duplicated ABC genes may divide the roles of its ancestor by partitioning the expression

patterns of its ancestor. Comparing the genomes of *C. elegans* and *C. briggsae* shows that protein and regulatory evolution is uncoupled in paralogs, suggesting that selective pressure on gene expression and protein evolution is different among duplicated paralogs (Castillo-Davis et al, 2004). In addition, duplicated genes exhibit a dramatic acceleration of both regulatory and protein evolution compared to orthologs, suggesting increased directional selection and/or relaxed selection on both gene expression patterns and protein function in duplicate genes. The presence of multiple, similar, paralogs of drug resistance genes may provide more effective protection against xenobiotics by spatially or temporally differential expression as is the case for 8 members of the two four-gene clusters. Differential expression of 5 duplicated ABC members was also observed in the mouse (Annilo, et al., 2003).

Overlapping or similar expression patterns were also observed for genes found in two-gene clusters. These members may not be functionally redundant, or else the gene duplication may be a mechanism to increase expression levels. One ABC transporter that did not provide GFP expression and showed few matching Yuji Kohara's ESTs was C56E6.1, a member of subfamily H. This gene, like many poorly expressed ABC genes in the assay system, is arranged in a head to head orientation with the other H subfamily member, C56E6.5. In other words, these two ABCs share a regulatory region but are transcribed in opposite directions. However, C56E6.1 is not a pesudogene because it gave an observable RNAi phenotype (Kamath, et al., 2003) and its cDNA has been successfully amplified (Reboul, et al., 2001). Thus, it is possible that this gene is expressed at a low level, or else in the germline in which tissue the transgene is expected to be silenced (Kelly, et al., 1997). Most members of subfamilies A and G failed to give

observable expression. Consistent with this observation, most of these genes had no or few Yuji's EST and SAGE tags (McKay, et al, 2003). The only two G subfamily members, C05D10.3 and C10C6.5 that gave observable GFP expression are exactly the same two for which SAGE tags have been assigned. In general it appears that genes in head-to-tail orientation tend to more easily be expressed in my assay compared to those in a head-to-head orientation. This may be partially explained that turning on of the one of the two head-to-head genes possibly prevent the activity of the other due to their "shared" regulatory region. It has also been reported that co-expression of neighboring genes in *C. elegans* is mostly due to the genes forming an operon or being recently duplicated genes (Lercher, et al., 2003). Genes arranged in the same orientation may be more easily co-expressed than those in the reverse arrangement.

I propose that not only can intronic regions of one gene be used as regulatory sequences for another gene, but also that the coding sequences of one gene can be used as a regulatory sequence for a neighboring gene. For example, the 1.5kb *pmp-3*(C54G10.3) promoter cannot drive observable GFP expression whereas a 2.9kb promoter, extending into the upstream coding region, C54G10.2(*rfc-1*) can effectively drive strong expression in intestine, vulva, hypoderm and neurons. Both *pmp-3* and *rfc-1*are well established gene models based on abundant Yuji Kohara's ESTs and good conservation between *C. elegans* and *C. briggsae*. A similar situation was found for *pgp-14* for which a longer promoter gave a more extensive GFP expression pattern. The included upstream introns are much smaller than the included exons, so it is plausible that the coding exons are also functioning as regulatory regions in this case. This may also explain why the worm genome is more compact than either the insect or vertebrate genome. Some worm genes

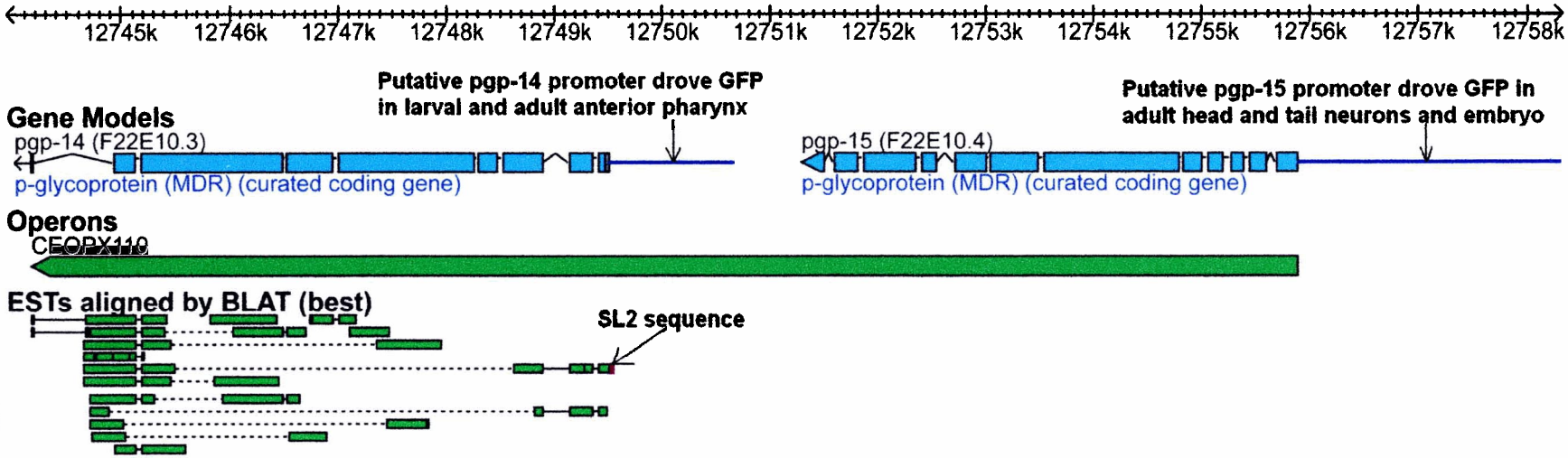
are located within the intron of another gene on the reverse strand. These intronic genes may also be under control of surrounding coding sequences.

Two ABCB members, *pgp-14* and *pgp-15* are annotated in WormBase as forming a two-gene operon (Figure 19). However, many *pgp-14* ESTs have been found and this gene can be successfully amplified by RT-PCR, while the latter has few ESTs and cannot be amplified by RT-PCR (Reboul, et al., 2001) Our transgenic data shows that a 1kb~1.7kb *pgp-14* promoter can drive strong reporter expression in the anterior pharynx in both larval and adult stages, while a 3.1 kb *pgp-15* promoter yielded weak expression in embryos and a few neurons in adults (Figure 17). The presence of the operon is confirmed both by microarray data (Blumenthal, et al., 2002) and the presence of an SL2 sequence in the 5' EST from Yuji Kohara's cDNA yk1279h09.5 (GenBank Accession BJ121724). I propose that transcripts from genes in operons may be generated both by co-transcription with upstream genes and independent transcription from internal promoters. In this case, the internal promoter for *pgp-14* is larger than 1kb, and can effectively drive strong reporter expression in pharynx. The lack of ESTs and RT-PCR products for *pgp-14* indicates that the transgenic assay is more sensitive than the strict molecular methods. It will be very interesting to examine whether other genes contained in operons with large intergenic spaces can practice dual transcription.

Mutation of single pgp gene did not significantly increase the sensitivity to both chloroquine and colchicine suggests that the two drugs are not substrates for these pgp proteins, or at least, these pgp proteins bind chloroquine and colchicines at low affinity. It is possible that some pgp proteins may be functionally redundant.

Figure 19. Presence of an SL2 sequence in a cDNA does not necessarily make the gene part of an operon.

The picture was generated using the Genome Browse tool on the Wormbase web site (release WS130, available at http://www.wormbase.org/). The top line is a sequence scale in kilobase pairs. Blue bars represent exons and black lines introns. Blue lines upstream of each gene represent putative promoters for *pgp-14* and *pgp-15* (not to scale). The continuous green bar indicates the predicted operon region. Small discontinuous green bars denote cDNAs (from Yuji Kohara's collection) aligned to the corresponding genomic region. An SL2 sequence (pink) in one cDNA is indicated by arrow. *pgp-14* is positioned downstream of *pgp-15*. The former has multiple ESTs while the latter has few ESTs. The 1.0 kb *pgp-14* promoter (intergenic sequence) drove strong GFP expression in the anterior pharynx in both larval and adult stages, while the 3.1 kb *pgp-15* promoter yielded weak expression in adult head and tail neurons as well as in embryos. The results suggest that *pgp-14* may not be part of an operon or can be transcribed either as part of an operon or independently. This figure is adapted from *Journal of Molecular Biology*, 2004 (by Zhao et al, 2004b)


12745k
12746k
12747k
12748k
12749k
12750k
12751k
12752k
12753k
12754k
12755k
12756k
12757k
12758k
Gene Models
Putative pgp-14 promoter drove GFP in larval and adult anterior pharynx
Putative pgp-15 promoter drove GFP in adult head and tail neurons and embryo
pgp-14 (F22E10.3)
p-glycoprotein (MDR) (curated coding gene)
pgp-15 (F22E10.4)
p-glycoprotein (MDR) (curated coding gene)
Operons
CEOPX110
ESTs aligned by BLAT (best)
SL2 sequence

CHAPTER 5 TRANSCRIPTIONAL CONTROL OF *PGP-12*

5.1 Abstract

Pgp-12 is a duplicated full-size P-glycoprotein that shows exclusive expression in excretory cell. *C. elegans* excretory cell is an equivalent of human kidney, which is involved in detoxification and osmolarity regulation. Investigation of transcriptional control of *pgp-12* may provide an alternative way to understand the mechanism that underlies the multiple drug resistance (MDR) developed during chemotherapy. Biological tube morphogenesis is central to the development of various metazoan structures, including vertebrate respiratory, excretory, and circulatory systems. The *C. elegans*' excretory cell provides an excellent model for investigating how these tubes are generated. I undertook a series of truncations of the *pgp-12* promoter to identify critical *cis*-regulatory regions necessary for its expression. I isolated a 30 base pair (bp) region necessary, in an orientation and distance dependent manner, for ectopic expression in excretory cell. Centred in the 30 bp region is Ex-1 a 10 bp segment (ccatacatta) that is completely conserved between *C. elegans* and *C. briggsae* (a related nematode separated by a 100 million year in evolutionary history). Site-directed mutageneis of Ex-1 completely abolished excretory cell specific expression. I used Ex-1 as bait in a one-hybrid screen and isolated a *trans*-acting factor - C26C6.5 that contains a putative zinc finger DNA binding domain. C26C6.5 is also expressed in the excretory cell and bind Ex-1 *in vitro*. RNAi analysis shows this gene is a *trans*-activator for excretory cell specific expression, and is involved in the morphogenesis of the excretory cell. In summary, I identified a novel *cis*-element and its binding *trans*- activator for excretory cell specific expression. This result may help unveil the mechanisms that underlie the phenotypes caused by P-glycoproteins.

5.2 Introduction

With the increasing number of genomic sequences available for a variety of organisms the next challenge is to make sense of non-coding sequences, which might play regulatory roles in gene expressions. Many of these sequences are important because they dictate the distribution and timing of different gene products during the development of organisms, especially multiple cellular ones with various types of differentiated cells. Investigations into how genes are expressed temporally and spatially will provide insights into fundamental mechanisms by which cells differentiate and function. Gene expression can be controlled at multiple steps, including pre-transcription control such as chromatin structure, transcriptional control (when, where and how much are mRNAs to be made), and translational control as well as post-translational control. There is increasing evidence showing that expression of the ABC transporters is highly regulated, particular at the level of transcription, suggesting a future target for modulation of the MDR phenotype (Scotto, et al., 2003). For example, p53 was found to repress the transcription of MDR1 gene (Thottassery, et al., 1997). The putative AP-1 binding sites have been reported to be involved in repression of murine pgp gene (Ikeguchi et al, 1991) but activation of human homolog (Teeter et al, 1991). MDR1 gene is also regulated by the ras/raf signalling pathway (Kim, et al., 1997; Yang, et al., 2001).

Analysis of expression patterns of *C. elegans'* genes has been greatly facilitated by new reporter genes such as green fluorescent protein (GFP) (Chalfie et al., 1994). These enable visualization of gene expression in real time *in vivo*. The usefulness of GFP as a reporter has been further facilitated by the creation of a convenient series of modular vectors (Miller et al., 1999). Fusion of a promoter containing region with a cassette

encoding GFP provides a convenient way to identify, dissect, and analyze the promoter, for example, identifying short *cis*-elements or motifs that work as binding sites for tissue specific transcription factors. Investigation of tissue specific *cis*-elements began over a dozen years ago in *C. elegans* (Spieth et al., 1988) but progress has been limited because of the complexity and labor intensity of the analysis. *Cis*- and *trans*-regulation of tissue specific gene expression is known only for a few tissues such as intestine (Gilleard et al.,1999), excretory duct cell (Wang et al., 2002) and muscle (Guhathakurta et al., 2002). However, *cis* and *trans* control of excretory cell expression are largely unknown.

The excretory cell, also called the excretory canal cell, is the largest mononucleate cell in *C. elegans*. This H-shaped cell lies just behind the isthmus of the pharynx on the ventral side and extends four long "canals"- two running anteriorly and two running posteriorly in the lateral hypodermis. The excretory cell functions in part as a kidney, excreting saline fluid via the duct and pore in order to maintain the animal's salt balance (osmoregulation) and probably to remove metabolites (Nelson and Riddle, 1984; Buechner et al., 1999). The birth of the excretory cell provides an easily identified landmark midway through embryonic development (Sulston et al., 1983).

The morphogenesis of biological tubes is central to the development of a wide variety of metazoan structures, such as vertebrate respiratory, excretory, and circulatory systems. *C. elegans'* excretory cell provides an excellent model for investigating mechanisms under which these tubes are generated. Current research on the excretory cell focuses on its morphogenesis by the characterization of mutants that show morphological defects. For example, disruption of the POU homeobox gene, *ceh-6*, leads to inappropriate canal structures in *C. elegans* (Bürglin et al., 2001). Similarly, loss of

function of *unc-53*, which contains putative SH3- and actin-binding sites, also blocks the progression of canal development (Stringham et al., 2002). Mutations in a mucin encoding gene, *let-653*, result in an extremely large canal lumen, swollen large enough to be visible under a low-power dissecting microscope (Jones and Baillie, 1995). Several other mutants have also been isolated, which show enlarged canal lumens termed cysts (Buechner et al., 1999).

pgp-12 is a homolog of human P-glycoprotein (pgp) genes, which are involved in developing multi-drug resistance during cancer chemotherapy (Sheps et al., 2004). Using an excretory cell specific *pgp-12* promoter containing region, Ce-pgp-12, I isolated a 10 bp *cis*-element called Ex-1 (excretory cell), which is critical for driving tissue specific expression in the excretory cell. A putative transcription factor that binds to the element has also been identified. Expression and functional analyses demonstrated the regulatory roles of the isolated transcriptional factor in the control of tissue specific expression in the excretory cell. I also demonstrated that Ex-1 binding factor is involved in morphogenesis of excretory cell. Thus, my data provided both a novel *cis*-element and its binding *trans*-activator necessary for tissue specific expression in the excretory cell of *C. elegans*. Investigation on transcription control of *pgp-12* may provide insights into drug resistance caused by pgp proteins.

5.3 Materials and methods

5.3.1 Strains

All strains were maintained and cultured using standard techniques. N2 (Bristol); KR3532, dpy-5(e907); GR1373 *eri-1(mg366)* IV; VC26 *pgp-12(gk19)* X; BC06405, dpy-

5(e907), sEX970[dpy-5(+) + rCesF22E10.1-260-GFP+pCes361]; BC06392, dpy-5(e907), sEX958[dpy-5(+) + rCes F22E10.1::RFP + pCeh361]; BC10210, dpy-5(e907), sIs10089[dpy-5(+) + rCesF22E10.1-GFP+pCes361]; BC06288, dpy-5(e907), sEX906[dpy-5(+) + rCesCbF22E10.1-GFP+pCes361]; BC06293, dpy-5(e907), sEX911[dpy-5(+) + rCesF22E10.1-228-GFP+pCes361];

BC10036, dpy-5(e907), sEX870[dpy-5(+) + rCesW04C9.1-GFP+pCes361]; BC06295, dpy-5(e907), sEX913[dpy-5(+) + rCesZK470.5-GFP+pCes361]; BC06417, dpy-5(e907), sEX981[dpy-5(+) + rCesC18C4.2-RFP+pCes361]; BC12778, dpy-5(e907), sEX12778[dpy-5(+) + rCesC26C6.5-GFP+pCes361]; BC12781, dpy-5(e907), sEX12781[dpy-5(+) + rCesR05D11.3-GFP+pCes361].

5.3.2 Comparison of *pgp-12* promoter to its *C. briggsae* orthologous region

Promoter sequences of *pgp-12*, Ce-pgp-12 and its orthologous sequence in *C. briggsae* Cb-pgp-12 were derived from Wormbase. Pair-wise alignments between 500 bp Ce-pgp-12 and Cb-pgp-12 were done with NCBI online pair-wise BLASTN program: BLAST 2 SEQUENCES with default parameters but with "Filter" off. Alignment was also done for the two sequences using DOTTER (Sonnhammer et al, 1995) with default parameters.

5.3.3 Site directed mutagenesis

All the promoter::GFP constructs were built basically as described (Hobert O, 2002). The nested primers used in promoter truncation were designed manually or by online program: Primer 3 (Rozen and Skaletsky 2000). Site-directed deletions were

performed by similar techniques as those for fusion PCR: the two primer sets were designed to amplify fragments upstream and downstream of the sequence to be deleted or mutated. The reverse primer used to amplify the upstream region contains an overhang, which is reversibly complementary to the forward primer used to amplify the downstream region. The two fragments were stitched together by PCR using the forward primer to amplify the upstream fragment and reverse primers to amplify the downstream fragment. The resulting PCR products were sequenced to confirm the expected mutations.

5.3.4 DNA transformation and microscopy

DNA transformation and microscopy was performed as described in last chapter . The DNA concentrations for GFP construct and *dpy-5* rescuing plasmid in injection mixture are 5 and 100 ng/μl. Pictures were taken for 20 worms each for at least 3 independent lines per strain using the same exposure time (1.05 second). GFP intensity was quantified based on GFP signal from individual worm.

5.3.5 Construction of integrated line for F22E10.1::GFP transgene

Chromosomal integration of the transgene was performed as described (McKay S.J et al, 2003). The transgenic worms for pgp-12::GFP were exposed to low dose X-ray irradiation (1500R) to induce double-stranded breaks in chromosomes. The resulting F_2 worms were screened for integrants.

5.3.6 Yeast one-hybrid screening

Yeast one-hybrid screen was performed according to the manufacturer's specifications (MATCHMAKER One-Hybrid System; Clontech). Briefly, three tandem copies of the Ex-1 element were synthesized as two reverse complementary strands with *EcoR*I and *Xba*I overhangs at the ends. Annealing of the two reverse complementary single-stranded DNAs yielded an *EcoR*I and *Xba*I adaptor. The resulting adaptor was cloned upstream of the HIS3 in the reporter plasmid pHISi digested with EcoRI and XbaI. The reporter plasmid was then integrated into yeast strain YM4271 to generate the reporter strain. Background growth of this reporter strain was ablated by titration with 30 mM 3-aminotrizole (3-AT) (Sigma). The resulting reporter strain was transformed with worm cDNA library, a gift from Dr. Robert Barstead using LiOAc strategy (Gietz RD *et al.*, 1995). Plasmids were extracted from the yeast colonies growing in His-Leu- agar plates with 30mM 3-AT using Zymoprep yeast plasmid miniprep kit (Cat No: D2001). The resulting plasmids from these clones were then transformed into DH5α. competent cells. Primers were designed based on flanking sequences of XhoI site from plasmid pACT2 (GenBank Accession #: AF134168) to amplify and sequence the inserts. The cDNA inserts were sequenced and then BLASTed against *C. elegans* genomic DNA through Wormbase. To confirm yeast one-hybrid screening results, the resulting positive clones were transformed back into the Ex-1 integrated YM4721 strain. The yeast colonies from each transformation were then streaked onto His-Leu- agar plates with 30 mM 3-AT. The growth of the streaked yeast colonies was compared to control (YM7271 with Ex-1 reporter integrated within) after 3 days incubation at 30°C.

5.3.7 Gel shift assay

Primers were designed so that C26C6.5 cDNA sequence isolated from one-hybrid screen was amplified as template for making protein using Promega TNT Quick Coupled Transcription/Translation Systems as described in manufacture's specifications. The PCR product encodes 216-681 out of 681 amino acids which contains the putative zinc-finger domain. The resulting protein was quantified by both Pierce Micro BCA™ Protein Assay Kit (Cat No: 23227) with Test Tube Procedure and SDS-polyacrylamide gel analysis. The three tandem Ex-1 was labelled by PIERCE Biotin 3' End DNA Labelling Kit (Cat No:89818). The labelling efficiency was determined as described. The labelled Ex-1 was annealed with its reverse complementary strand to form double stranded oligoes. C26C6.5 cDNA encoding 122 amino acids that covers the predicted zinc finger domain was used as template to produce protein using TNT Quick Couple Transcription/Translation System from Promega. Electrophoresis mobility shift assay was performed using PIERCE LightShift™ Chemiluminescent EMSA Kit (Cat No: 20148X) following manufacture's specifications.

5.3.8 RNA interference (RNAi) by feeding

RNAi by feeding was done as described previously (Fire et al, 1998). RNAi bacteria were kindly provided by Dr. Ahringer and streaked onto NGM plates with ampicillin, tetracycline and IPTG overnight at room temperature. Ten L4 worms were seeded onto the plates and incubated at room temperature for another 24 hours. Three worms were transferred onto the 3 replica plates and scored for RNAi phenotype every 24

hours. The whole process was repeated for the second generation on the same bacteria to re-score phenotypes.

5.3.9 RNAi by injection

PCR primers were designed so that 615 bp PCR product covers the second exon of C26C6.5 with T7 or T3 overhang on either side. It was used as template to make single stranded RNAs using Ambion MEGAscript T3/T7 High Yield Transcription Kit (Cat. No. 1338) according to manufacture's specifications. The resulting RNAs were combined together and purified with Invitrogen Micro-to-Mini Total RNA Purification System (Cat. No. 12183-018). The purified RNAs were annealed by incubation at 68 °C for 10 minutes and 37 °C 30 seconds and quantified on 1% agrose gel. The final injection mixture contains 20 ng/μl double stranded RNA, 4 mM KPO4, 0.6 mM K Citrate and 0.4% PEG8000. L4 or young adult were used for microinjection. The injected worms were transferred to fresh plates after six hours and transferred again after another three consecutive 24 hours. The plates with eggs for the first six hours were discarded. Score the phenotypes on the remaining plates.

5.3.10 Mapping of potential C26C6.5 mutant

Two candidate lethal mutants, *let-535*(KR1692) and *let-538* (KR1598) were mapped against C26C6.5. Mid-larval arrested DpyUnc let-535 worms and sterile adult DpyUncs let-538 worms were picked and used for single worm PCR. The resulting PCR product were sequenced and compared against C26C6.5 genomic DNA.

5.3.11 Screen for suppressors for *pgp-12* expression

In order to screen for potential suppressor that block the expression of *pgp-12*, an EMS mutagenesis on sIs10089, an integrated *pgp-12* expressing strain, was performed. 8 µl Ethyl methane sulfonate (EMS) (Sigma, M-0880) was added into 2.992 ml M9 buffer to give 25 mM EMS. Synchronized L4 worms from two large plates were washed into 15 mL Falcon tube using sterile M9 buffer. Worms were washed three more times in 5 ml M9 and re-suspend into 2.5 ml M9. 2.5 ml 25 mM EMS was added into the tube to give 12.5 mM final concentration and mix well. Worms were lefted in room temperature for 4 hours. Worms were washed three times with M9 to remove EMS. After each wash, the worms were collected by centrifugation at 1000g at 4°C for 2 minutes. Mutagenized worms allowed to lay eggs on fresh plates for 1 hour. 50 healthy worms were piched onto 10 large plates and allow them to lay eggs for 24 hours. P0 worms were picked off and the progeny were allowed to grow for 2-3 days. Grown progeny were screened for those that lose GFP expression or show significant decrease in *pgp-12* derived expression.

5.4 Results

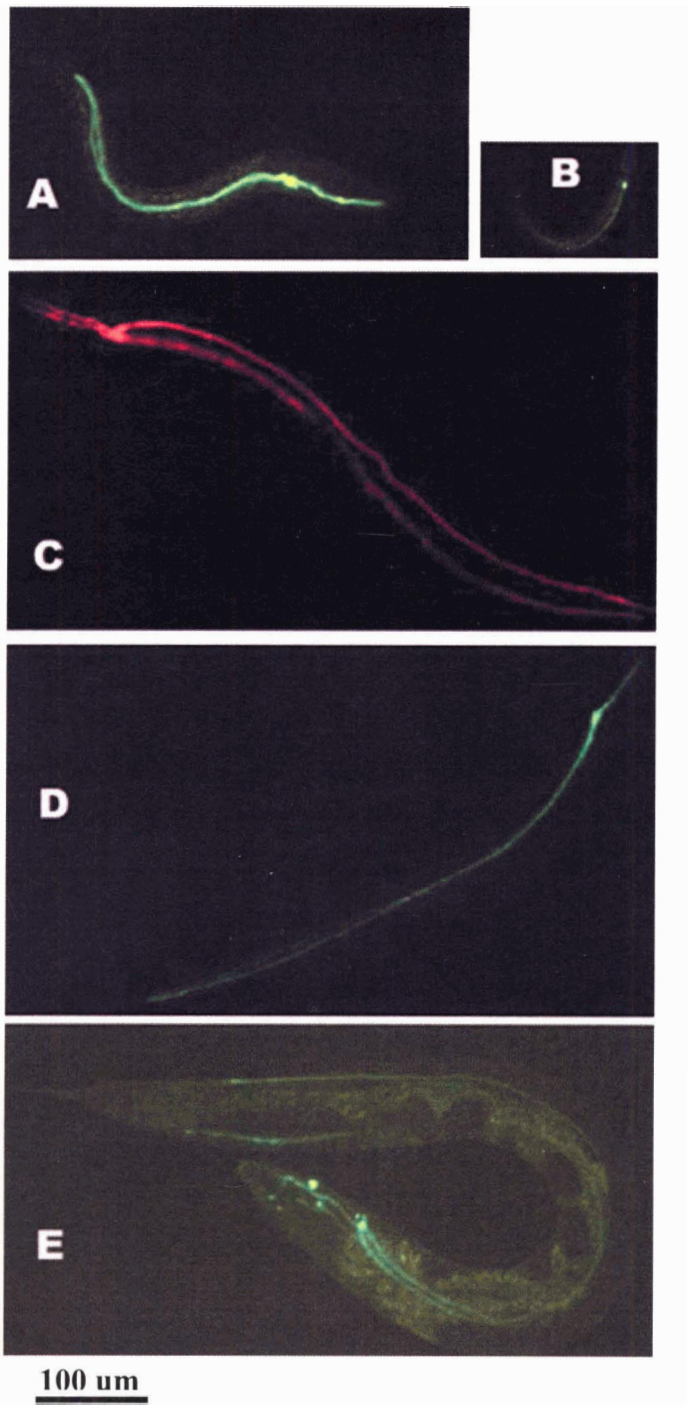
5.4.1 A 3.4 kb *pgp-12* promoter drives specific reporter expression in the excretory cell at all developmental stages

In order to map the *cis*-elements responsible for tissue specific expression in the excretory cell, I used a putative *pgp-12*(F22E10.1) promoter-containing region, Ce-pgp-12. It includes a 3410 bp sequence upstream of the start codon (ATG) of *pgp-12* and drives strong and exclusive expression of reporters, i.e. GFP or DsRed (Clontech) in *C. elegans'* excretory cell at all developmental stages (Figure 20). DsRed was used to

confirm that expression patterns were the result of the specific promoter and not the specific reporter. In order to facilitate the identification of potential *cis*-elements that may be conserved between *C. elegans* and *C. briggsae*, the *C. briggsae* orthologous region of Ce-pgp-12, Cb-pgp-12, was used to form a Cb-pgp12::GFP fusion construct, which was injected into *C. elegans*. The 1160 bp Cb-pgp-12 drives the same expression patterns as those of Ce-pgp-12. Promoters that give strong and exclusive expression in excretory cells provide a convenient basis for mapping critical causative *cis*-elements and subsequently their *trans*-activators.

Figure 20. Putative *pgp-12* promoter and its *C. briggsae* ortholog, Cb-pgp-12 drives strong and exclusive reporter expression at all stages.

A and B show Ce-pgp-12::GFP expression in L4 and L2 larvae. C and D show adult Ce-pgp-12::DsRed and Ce-pgp-12::GFP respectively. E shows Cb-pgp-12::GFP expression in *C. elegans*.


A
B
C
D
E
100 um

5.4.2 A 10 bp motif termed Ex-1 is critical for driving excretory cell specific expression

Starting with Ce-pgp-12, I performed a series of truncations of the 3410 bp promoter to determine definitive regions responsible for excretory cell specific expression. Deletion of more than 3 kb furthest upstream from the translation start codon did not change the expression patterns. The remaining 286 bp region drove comparable expression patterns to those using the 3410 bp region (Figure 21A). However, truncating another 48 bp region resulted in a 238 bp region that yielded only modest expression in larvae and adults and no expression in embryos. This indicated that some of the excretory cell specific *cis*-elements and the embryonic specific elements are within that 48 bp region. Deletion of another 10 bp sequence caused complete loss of GFP expression in both larvae and adult. This clearly indicated that the deleted 10 bp sequence, `ccatacatta`, termed Ex-1, is critical for tissue expression in the postembryonic excretory cell of *C. elegans*. A search of TransFac (Wingender et al., 2001) using Ex-1 as query did not give any hits for *C. elegans*. In order to confirm the importance of Ex-1 in excretory cell specific expression, a site-directed mutagenesis was performed on the 475 bp most downstream region in Ce-pgp-12 (475-Ce-pgp-12). The 475-Ce-pgp-12 region yielded comparable expression patterns to that of the 3.4 kb Ce-pgp-12. Removal of the Ex-1 element from the 475-Ce-pgp-12 also caused complete loss of GFP expression in the excretory cell (Figure 21A). These data confirm the importance of Ex-1 in tissue specific expression in *C. elegans'* excretory cell.

Alignment of Ce-pgp-12 and Cb-pgp-12 using NCBI BLAST 2 SEQUENCES showed perfect conservation of Ex-1 between the two species (Figure. 21B). The Ex-1 flanking sequences, Ex-L (10 bp to the left of Ex-1), Ex-R (10 bp to the right of Ex-1), and Ex-R1 (another 10 bp to the right) show approximately 50% identity between *C. elegans* and *C. briggsae*. Individual site-directed mutagenesis was done on Ex-L, Ex-R and Ex-R1 (Figure. 21A) in the 475-Ce-pgp-12. None of the sequence disruptions completely abolished GFP expression, indicating that they were not as important as Ex-1 for tissue specific expression in *C. elegans'* excretory cell.

Alignment of 500-Ce-pgp-12 and Cb-pgp-12 by DOTTER (Sonnhammer et al., 1995) revealed a large conserved region upstream of Ex-1 (Figure. 21C). The truncation analysis showed that the 48 bp immediately upstream of Ex-1 was required for efficient GFP expression in the excretory cell. However, the conserved regions upstream of the 48 bp sequence (circled regions in Figure. 21C) appear not to be necessary for excretory cell specific expression.

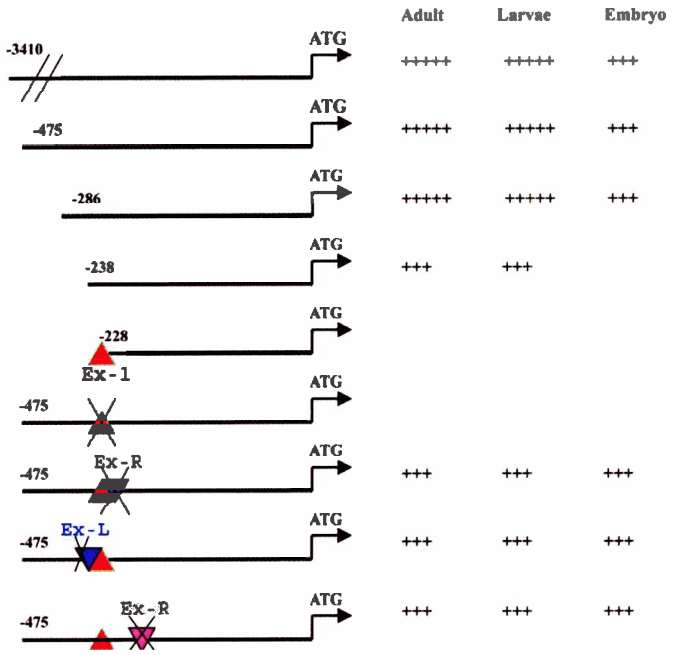
Figure 21. Mapping of *cis*-elements that are essential for excretory cell specific expression.

A: a 3410 bp excretory cell specific *pgp-12* promoter-containing region was truncated using a series of nested PCR primers. Deletion of a 10 bp motif, termed Ex-1 (red triangle), caused complete loss of GFP expression at all developmental stages. The crosses denote deletion of specific elements. Site directed deletion of Ex-1 within 475-Ce-pgp-12 also eliminated GFP expression at all stages. Deletion of Ex-1 flanking sequences, Ex-L (purple triangle), Ex-R (green triangle) or Ex-R1 (pink triangle) decreased but did not cause complete loss of GFP expression at any stage.

B: Alignment of 536 bp Ce-pgp-12 and its 493 bp *C. briggsae* orthologous region using NCBI BLAST 2 SEQUENCES with "Filter" off, word size 7, Expect 30, and gap extension 1. The 10 bp element, Ex-1 (red), Ex-L (green) , Ex-R(blue) and Ex-R1 (pink) are highlighted. Only Ex-1 is perfectly conserved between *C. elegans* and *C. briggsae*.

C: Alignment of the two sequences used in B using DOTTER (Sonnhammer et al, 1995) with default parameters. Ce-pgp-12 was shown in horizontal axis and Cb-pgp-12 in vertical axis. The conserved Ex-1 site was indicated by arrow. The two big conserved segments upstream of Ex-1 were highlighted by circles and not necessary for excretory cell specific expression.

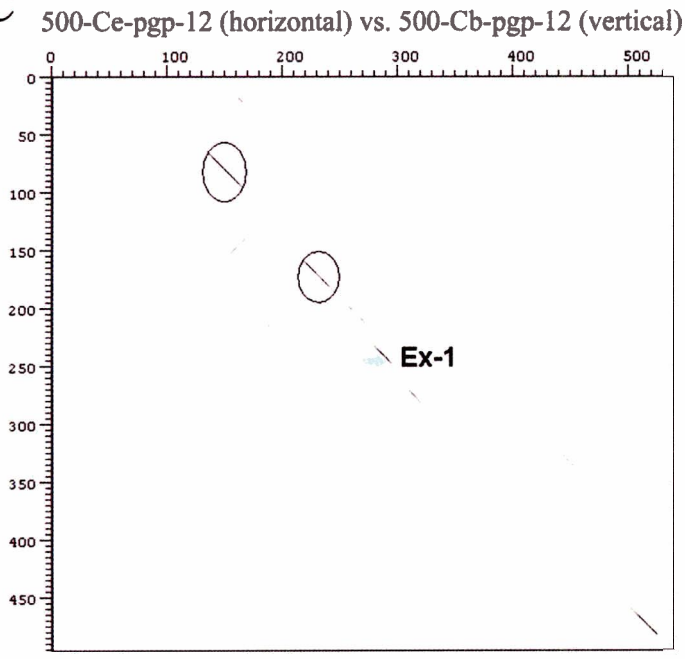
A


Adult
Larvae
Embryo
-3410
ATG
+++++
+++++
+++
-475
ATG
+++++
+++++
+++
-286
ATG
+++++
+++++
+++
-238
ATG
+++
+++
-228
ATG
Ex-1
-475
ATG
-475
Ex-R
ATG
+++
+++
+++
-475
Ex-L
ATG
+++
+++
+++
-475
Ex-R
ATG
+++
+++
+++

B


Ex-L
Ex-1
Ex-R
Ex-R1
CE caaaggatttccatacattattcccg-cg-------ttttggcaccca
CB aaagtgatttccatacattattaccgacgatttttttttctggcaccca

C


500-Ce-pgp-12 (horizontal) vs. 500-Cb-pgp-12 (vertical)
0
100
200
300
400
500
0
50
100
150
200
250
300
350
400
450
Ex-1

5.4.3 Ex-1 plus its flanking 10 bp on each side (Ex-L and Ex-R) is the minimum sequence that is sufficient for ectopic expression in the excretory cell in an orientation and distance dependent manner

Given the importance of Ex-1 in driving tissue specific expression in the excretory cell, I tried to determine whether Ex-1 could change the expression patterns of a non-excretory cell specific promoter, i.e., ectopic expression. A 286 bp intestine specific promoter, Ce-haf-4 yields moderate GFP expression in the intestine (Figure 22 and 23). I chose Ce-haf-4 for two reasons: first, Ex-1 would be roughly the same distance from the start codon as that from its endogenous position if it is put at the 5' end of the Ce-haf-4 (286 vs. 238); second, Ce-haf-4 contains a functional promoter that drives moderate GFP expression exclusively in intestine. So Ce-haf-4 will provide the necessary elements for basal transcription if Ex-1 is put at its 5' end. A single Ex-1 at the 5' end did not cause any changes in expression patterns of Ce-haf-4 (Figure. 22). But when I joined Ex-L-Ex-1-Ex-R to the 5' end of Ce-haf-4, I got moderate GFP expression in both the intestine and the excretory cell. Therefore Ex-1 is necessary but not sufficient for excretory cell specific expression. It has to work together with Ex-L and Ex-R to drive expression in the excretory cell. Interestingly, the reversed 30 bp sequence (Ex-L-Ex-1-Ex-R) inserted on the 5' end of Ce-haf-4 did not cause any changes in the expression patterns, indicating that Ex-1 and its flanking sequences work in an orientation dependent way. The Ex-L-Ex-1-Ex-R sequence was also tested on the 5' end of two other promoters: a 1724 bp pharynx

specific promoter region Ce-pgp-14 and an 1193 bp intestine specific promoter region Ce-pmp-1. No changes were observed in expression patterns of the two promoters (Figure. 22), implying that Ex-1 and its flanking sequences function only within a certain distance from the start codon.

Figure 22. Ex-1 with its two flanking regions can cause ectopic expression in a distance and orientation dependent manner.

Expression patterns in adults are shown. A single Ex-1 at the 5' end of Ce-haf-4 did not change the expression patterns of the intestine specific promoter. Ex-1 at the 5' end of the Ce-haf-4 is roughly the same distance from the start codon (286 bp) as its endogenous position that is 228 bp. Ex-L-Ex-1-Ex-R is sufficient to cause expression pattern changes, i.e., from intestine to both intestine and excretory cell. Expression patterns were not altered when the three elements were inserted in reverse orientation at the 5' end of Ce-hal-4. Similarly, the altered expression patterns were not observed when the three elements were put at the 5' end of larger promoters: the 1724 bp pharynx specific promoter containing region, Ce-pgp-14 or the 1193 bp intestine specific promoter containing region, Ce-pmp-1. Red triangle: Ex-1, green triangle: Ex-R, purple triangle: Ex-L.

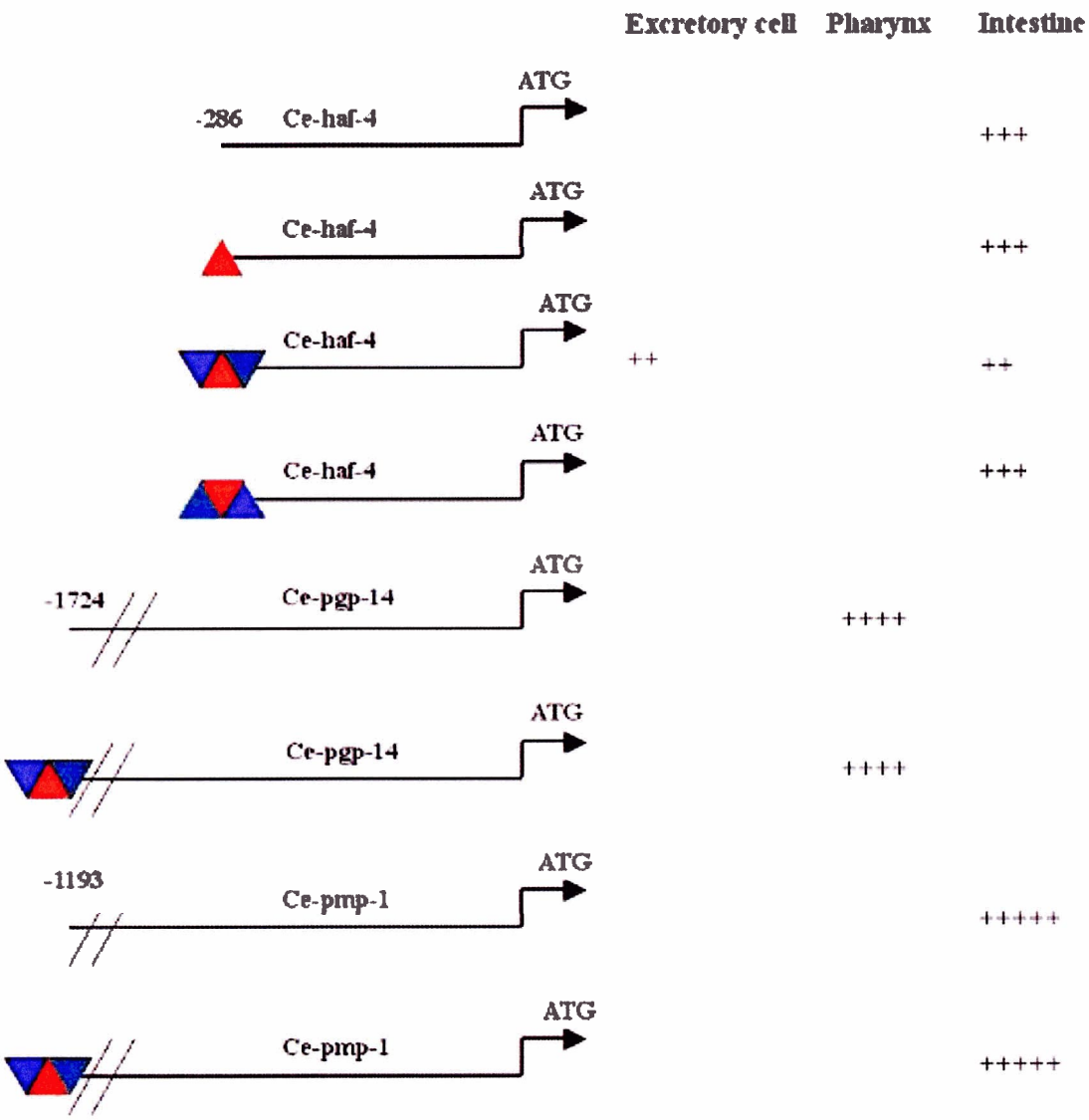
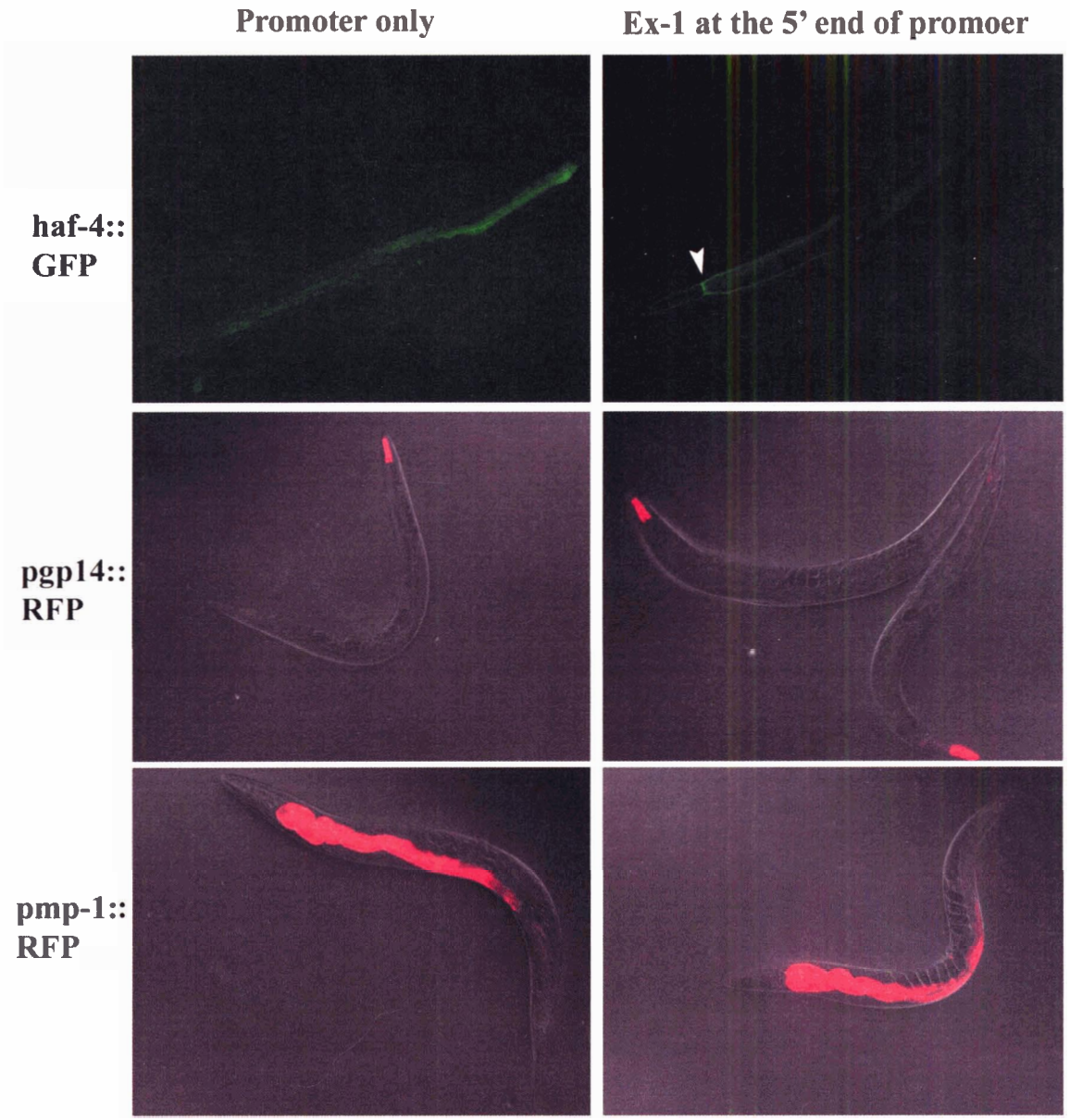

Excretory cell
Pharynx
Intestine
ATG
-286
Ce-haf-4
+++
ATG
Ce-haf-4
+++
ATG
Ce-haf-4
++
++
ATG
Ce-haf-4
+++
ATG
-1724
Ce-pgp-14
++++
ATG
Ce-pgp-14
++++
ATG
-1193
Ce-pmp-1
+++++
ATG
Ce-pmp-1
+++++

Figure 23. Ex-1 caused ectopic expression in a distance dependent manner.

A 286 bp intestine specific promoter, Ce-haf-4 yielded moderate GFP expression in intestine. Ex-L-Ex-1-Ex-R joined at the 5' end of Ce-haf-4 caused ectopic expression in the excretory cell (indicated by arrow). The Ex-L-Ex-1-Ex-R sequence did not cause ectopic expression when fused at the 5' end of a 1724 bp pharynx specific promoter Ce-pgp-14 and an 1193 bp intestine specific promoter Ce-pmp-1.


Promoter only
Ex-1 at the 5' end of promoer
haf-4::
GFP
pgp14::
RFP
pmp-1::
RFP

5.4.4 Transcriptional control of excretory cell specific expression may be mediated by multiple elements upstream of Ex-1

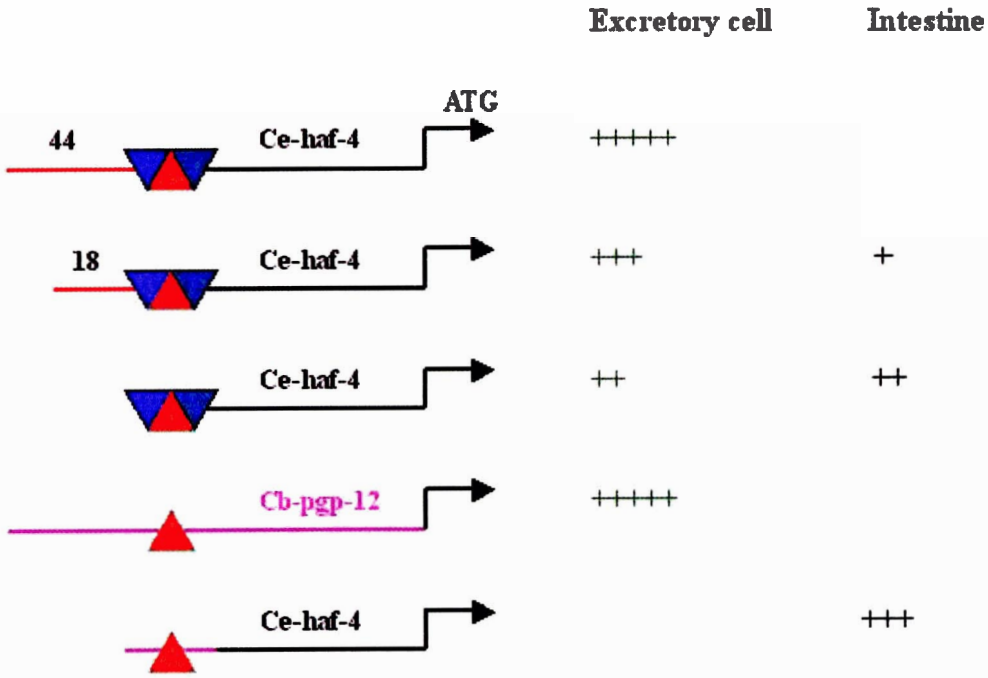
As shown in Figure. 18, Ex-L-Ex-1-Ex-R plus its upstream 44 bp (red) region on the 5' end of Ce-haf-4 gave strong GFP expression only in the excretory cell, comparable to that of 475-Ce-pgp-12 (Figure. 21A). In addition, expression in the intestine was suppressed by the grafted sequences. However, Ex-L-Ex-1-Ex-R plus its upstream 18 bp region on the 5' end of Ce-haf-4 yielded expression in both intestine and excretory cell. Decreasing the size of the grafted sequences led to reduced GFP signal in the excretory cell, but increased GFP signal in the intestine, indicating that there is *cis*-element upstream of Ex-L that plays a role in the regulation of expression in the excretory cell. Suppression of intestine specific expression by the grafted sequences suggests that excretory cell specific transcription factors may compete with intestine specific ones for basal transcriptional machinery. The minimum sequences required for ectopic expression in the excretory cell are Ex-L, Ex-1 and Ex-R in their endogenous order.

Cb-pgp-12 contains perfect Ex-1 but imperfect Ex-L and Ex-R sequence (Figure. 21B). The orthologous *C. briggsae* region, including sequences 44 bp upstream of Ex-1 to ATG (Figure. 24) injected into *C. elegans* drove comparable expression patterns to those of *C. elegans'* orthologous region (Figure. 24). However, Ex-1 plus its flanking 10 bp on each side within Cb-pgp-12 did not change the expression patterns of Ce-haf-4 (Figure. 24). Similar patterns were observed when I used a fragment containing the 44 bp region

upstream of the Ex-1, Ex-1 and 10 bp on Ex-1's right side, suggesting *C. briggsae* uses a different combination of Ex-1 and other elements for excretory cell specific expression.

Figure 24. Transcriptional control of excretory cell specific expression may be mediated by multiple elements upstream of Ex-1 in *C. elegans*.

A combination of Ex-L-Ex-1-Ex-R with its upstream 44 bp (red) inserted at the 5' end of Ce-haf-4 gave comparable expression patterns to those of 475-Ce-pgp-12 while expression in the intestine was suppressed (undetectible). GFP signal reduced in excretory cells but increased in the intestine with decreasing sizes of the grafted sequences. The minimum sequences required for observable excretory cell expression are Ex-L-Ex-1-Ex-R. Cb-pgp-12 contains only Ex-1 but not Ex-L or Ex-R (see text). A truncated Cb-pgp-12, containing 44 bp sequence upstream of Ex-1 to ATG (pink), drove comparable GFP expression in the excretory cell to that of Ce-pgp-12. Cb-pgp-12 Ex-1 plus its two 10 bp flanking sequence inserted on the 5' end of Ce-haf-4 did not change the expression patterns. Red triangle: Ex-1; purple triangle: Ex-L; green triangle: Ex-R


Excretory cell
Intestine
ATG
44
Ce-haf-4
+++++
18
Ce-haf-4
+++
+
Ce-haf-4
++
++
Cb-pgp-12
+++++
Ce-haf-4
+++

5.4.5 A putative transcription factor that binds Ex-1 was isolated by one-hybrid screening

Two positive clones were isolated from a one-hybrid screen using Ex-1 as a bait. The resulting cDNA sequences were aligned using BLASTN against *C. elegans* genomic DNA. One of the inserts contains an in-frame cDNA sequence for C26C6.5. The other one matches a cDNA sequence of R05D11.3. C26C6.5 shows similarity to Deacetylase Complex Protein (Wade et al., 1999). A putative glucocorticoid receptor-like zinc finger DNA-binding domain (residues 432-501) was identified within the C26C6.5 by scanning against the InterPro database (Mulder et al., 2003), suggesting it is a nuclear DNA-binding protein. PSORT II also predicts the nuclear localization of C26C6.5 (73.9%, data not shown). It is annotated as two isoforms in wormbase (WS133), C26C6.5a and C26C6.5b (Figure 25A). C26C6.5a has a bigger untranslated regions (UTR), especially 3'-UTR than those of C26C6.5b (Figure 25A). The latter encodes three more amino acids (9 nucleotides immediately after the fifth exon). The first 644 bp from translation start (ATG) is absent in the cDNA sequence isolated from the one-hybrid screen that matches C26C6.5a. The striking feature for the C26C6.5a cDNA is presence of two identical 595 bp repeats (spaced by one nucleotide) in the same orientation within the long 3' UTR region (red bars in Figure 25A). The two repeat sequences do not seem to encode any open reading frame (ORF) due to frequent stop codons found by six-frame translation (data not shown). Sequencing of full-length Yuji Kohara cDNA, yk1379-h11 confirmed the presence of two tandem repeats. Investigation the roles of these repeats is urgent.

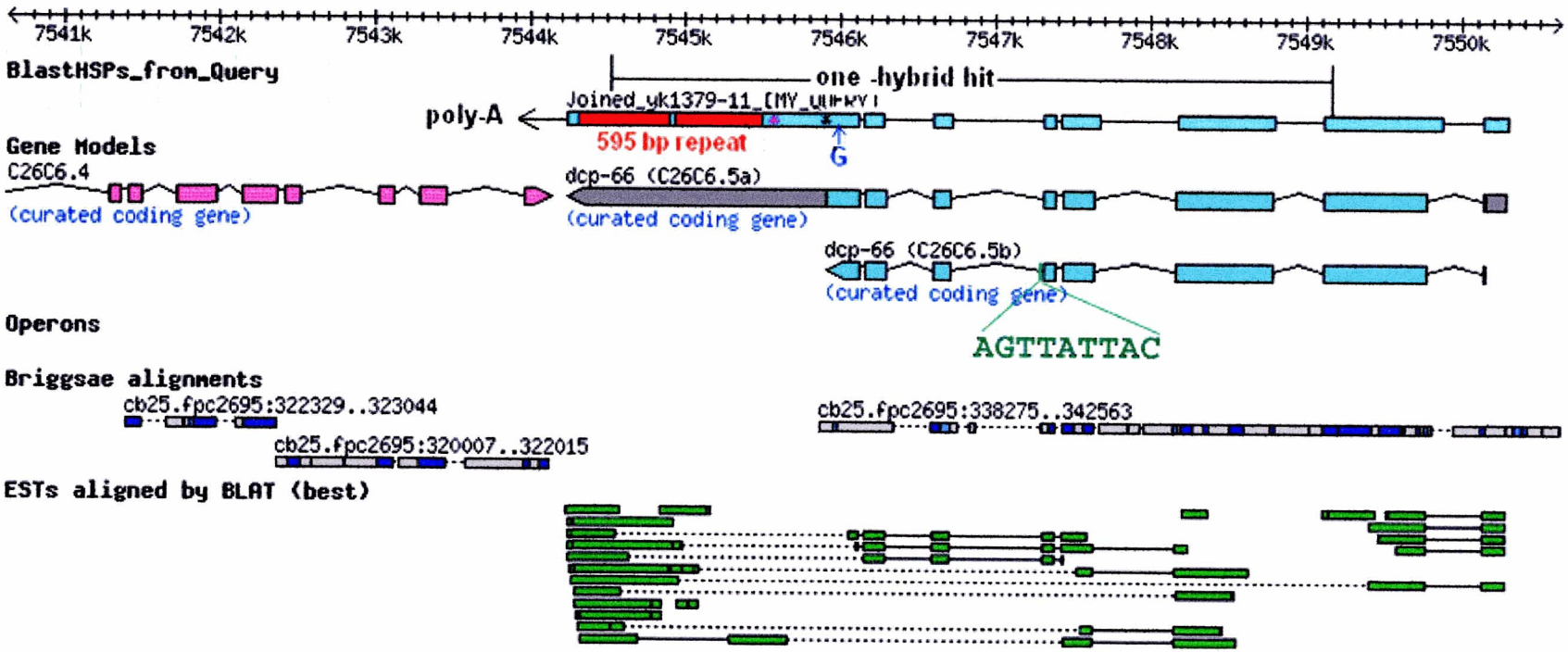
Attempts have been made to map two potential candidate lethal mutants, let-535 and let-538 against the C26C6.5 gene. However, none of them turned out to have mutation in the coding region of C26C6.5. Co-incidence of a single nucleotide insertion of G in position 1995 from ATG in both cases revealed a genomic sequencing error in *C. elegans* genome. The unambiguous presence of G was confirmed by sequencing of both genomic DNA (N2 and let-535, let-538) (Figure 25B,C) and cDNA (yk1379-h11 and one-hybrid hit) (data not shown). The insertion of the G extends the ORF further downstream for another 100 amino acids, making the C26C6.5a encode 781 amino acids. Scan of C26C6.5 protein sequence using COILS algorithm yields multiple runs of coiled coils (http://www.ch.embnet.org/software/COILS_form.html). The coiled coils can be found in proteins involved in transcription, in cell-cell and viral-cell fusion events as well as in maintaining structural identity of cells (Wolf et al, 1997). Alignment of C26C6.5a amino sequences with its homologs from other organisms revealed two conserved fragments. The one close to amino terminus seems well-conserved only among vertebrates, while the one near carboxyl terminus shows high conservation across all species and contain a putative zinc finger domain (Figure 26).

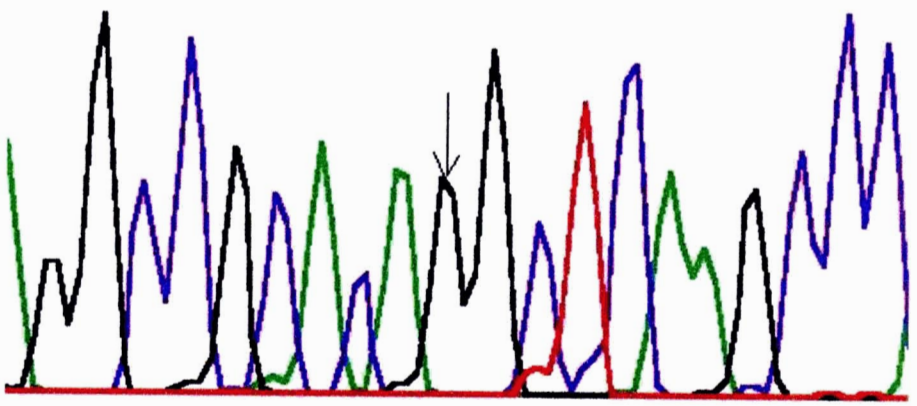
No substantial DNA binding domain was found within R05D11.3 amino sequence. It encodes a nuclear transport factor which is predicted to interact with nucleoporins and to promote transport of RAN-1 (+GDP) into the nucleus. A fusion construct of R05D11.3 promoter containing region with a GFP cassette gave expression only in the intestine but not in the excretory cell (data not shown). This suggests it is possibly a false positive cDNA clone picked up during one-hybrid screening.

Figure 25. Genomic sequence of C26C6.5 contains two 595 bp repeats within the 3' UTR

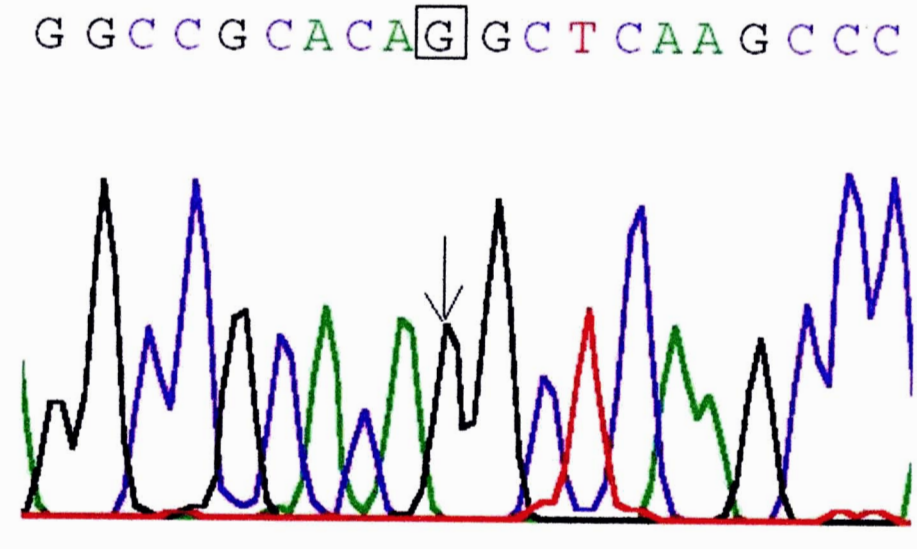
A. Gene modes for C26C6.5 were derived from Genome Browser (WS132). C26C6.5 encodes two alternative splicing variants, C26C6.5a and C26C6.5b. The former contains big un-translated region (UTR), especially 3'-UTR. The later encodes 3 more amino acids immediately following the 5th exon as indicated by 9 nucleotides (green). The first gene model is produced based on sequencing result of full-length Yuji Kohera EST, yk1379-h11. The sequence coverage of the one-hybrid hit is indicated. The 644 bp at the 5' end and 224 bp from 3' end are absent as compared to yuji EST. The blue bar represents exons and back line introns. Red bar denotes two identical 595 bp repeats (spaced by one nucleotide) in the same orientation within the long 3' UTR region (grey in 3'-UTR of C26C6.5a). Sequencing of both N2 genomic region and Yuji cDNA confirmed the presence of the two tandem repeats. A genomic sequencing error in *C. elegans* genome was identified by sequencing of C26C6.5 regions in both wild type N2 strain and two candidate mutant strains, let-535 (KR1692) and let-538 (KR1598). A single nucleotide, G was missed by genome sequencing project, which is 1995 bp from ATG (cDNA). The insertion of the G extends the ORF further downstream for another 100 amino acids, making the C26C6.5a encode 781 amino acids. Stop codons with or without the G were indicated by red and black star respectively. Scale in kilobases is shown on the top of the panel. The discontinuous blue and grey bars represent *C. briggsae* alignment with high and medium similarity respectively. The interrupted green bars indicate the Yuji Kohara ESTs aligned to corresponding exons.

B and C: sequencing chromographs from KR1692 (B) and KR1598 (C). The unambiguous presence of G was shown.


7541k
7542k
7543k
7544k
7545k
7546k
7547k
7548k
7549k
7550k
BlastHSPs_from_Query
one -hybrid hit
poly-A
595 bp repeat
G
Gene Models
C26C6.4
(curated coding gene)
dcp-66 (C26C6.5a)
(curated coding gene)
dcp-66 (C26C6.5b)
(curated coding gene)
AGTTATTAC
Operons
Briggsae alignments
cb25.fpc2695:322329..323044
cb25.fpc2695:320007..322015
cb25.fpc2695:338275..342563
ESTs aligned by BLAT (best)

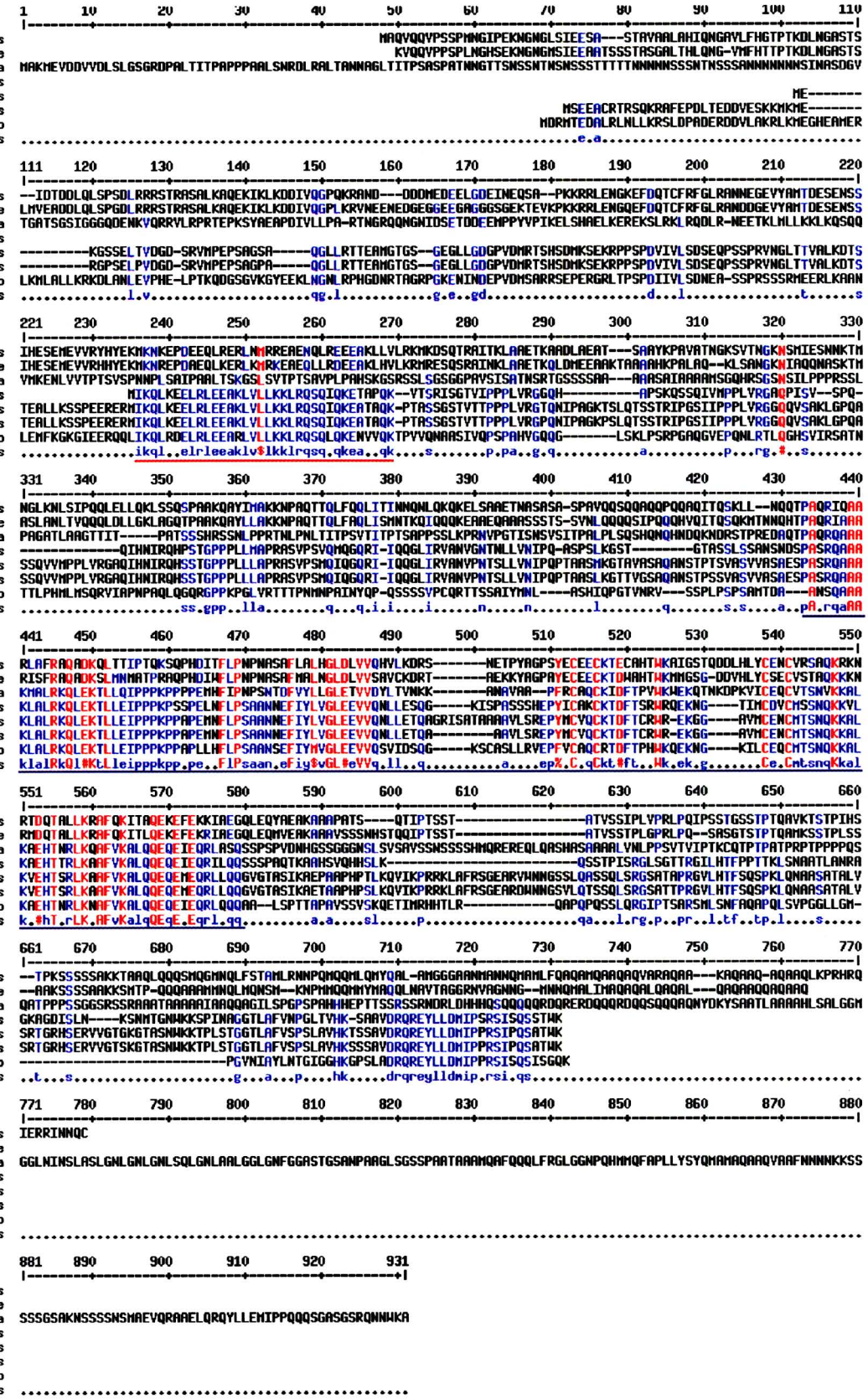


B



C

Figure 26. Alignment of C26C6.5 with its homologs from other organisms
Protein sequences for C26C6.5a is used to search non-redundant database using PSI-BLAST at NCBI. The resulting hits with highest score were used for multiple alignment. Sequences used in the alignment include *C. elegans* (gi|12705727), *C. briggsae* (gi|39582481), *Drosophila* (gi|23093698), *Mus musculus* (gi|23398610), *Rattus norvegicus* (gi|34877681) and *Homo sapiens* (gi|21218438). Alignment was performed using the MultAlin program (www.toulouse.inra.fr/multalin.html). Red and blue residues represent high and low identity, respectively. The two well-conserved regions were underlined in red and blue. The first region seems only conserved in vertebrates. The second region contains zinc finger domain. Symbols: !, I or V; $, L or M; %, F or Y; #, N, D, Q, E, B, or Z.


C.elegans
C.briggsae
Drosophila
Xenopus
Mus
Rattus
Homo
Consensus

5.4.6 C26C6.5 binds Ex-1 element *in vitro*

In order to independently confirm the binding of C26C6.5 protein with Ex-1, I performed electrophoresis mobility shift assay (EMSA) using biotin labeled three tandem Ex-1 sequences as probe (see materials and methods). As shown in figure 27, the 122 amino acids from C26C6.5 protein that encodes predicted zinc finger domain binds the Ex-1 element in vitro. Since the C26C6.5 protein binds the Ex-1 *in vitro*, the question arises as to whether it bind Ex-1 in vivo in *C. elegans*.

5.4.7 C26C6.5 is expressed in the excretory cell in vivo

In order to further confirm the binding of C26C6.5a to Ex-1, I generated a promoter::GFP transgenic line for C26C6.5a. The 2525 bp C26C6.5a putative promoter containing region (Ce-C26C6.5) drove GFP expression in the excretory cell and pharynx in both larvae and adults. Weak expression was observed in adult pharynx, vulvae and posterior neurons as well as in embryos (Figure 28). Stronger expression of C26C6.5 in excretory cell of larvae than that in adult was observed. Excretory cell expression of C26C6.5a was only observed in late embryo. Full length genomic DNA of C26C6.5 plus its putative 2.5 kb promoter (deriver from wormbase 128) was fused with DsRed. However, no observable expression was seen in the transgenic lines with the construct (data not shown). This is possibly due to a frame shift caused by missing of single guanine at 1995 from *C. elegans* genome sequencing (Figure 25). Overlapping expression patterns of Ce-C26C6.5a and Ce-pgp-12 suggest that Ex-1 works as a substrate for

C26C6.5a. The C26C6.5a protein possibly together with other transcription factors, bind to corresponding elements to initiate pgp-12 expression in the excretory cell. To examine the effects of C26C6.5a on the expression of pgp-12, I performed RNAi against C26C6.5a by injection using *pgp-12* expressing strain.

Figure 27. C26C6.5 binds Ex-1 from pgp-12 promoter sequence *in vitro*
EMSA was performed as described in Materials and Methods.

A: Free labeled Ex-1; B: Labeled Ex-1 plus reticulate lysate, shifted bands is indicated by arrowhead; C, D and E: Competitive binding with 200X, 50X and 10X molar excess unlabelled Ex-1 respectively. The thick dark bands at the bottom are free-labeled Ex-1s in 5 fmol each.

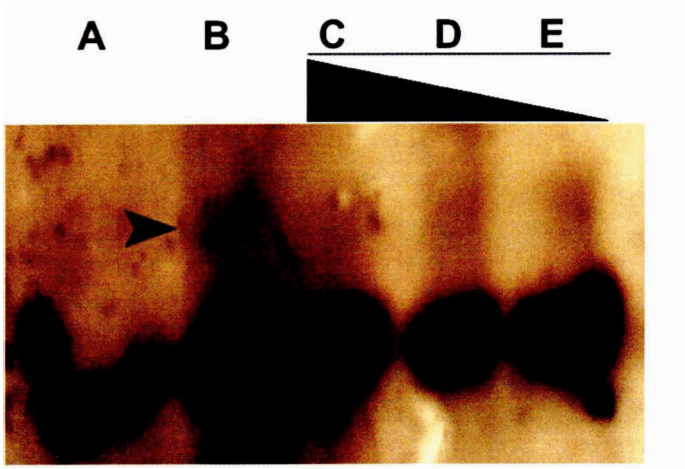


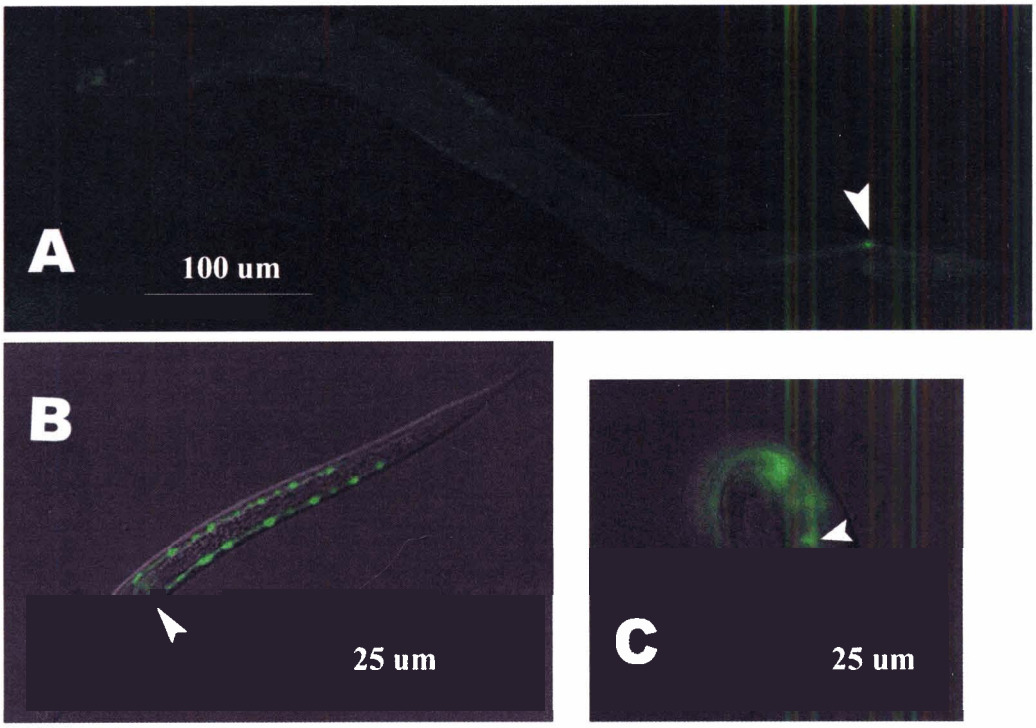
Figure 28. Expression pattern of C26C6.5 overlaps with that of pgp-12

A: In adults, weak expression of C26C6.5 was observed in the excretory cell, pharynx bulbs, posterior neurons, and vulvae.

B: In larvae, C26C6.5 was strongly expressed in the excretory cell.

C: Embryonic expression of C26C6.5 was observed in the excretory cell.

Note: cell body is indicated by arrowheads



5.4.8 RNAi "knockdown" against C26C6.5 specifically attenuated pgp-12 expression in the excretory cell

In order to test the effects of C26C6.5 on excretory cell specific expression driven by *pgp-12* promoter in vivo, I performed RNAi "knockdown" against C26C6.5 on the stable strain expressing GFP in the excretory cell by both feeding and injection. An integrated Ce-pgp-12::GFP transgenic strain, BC10210, generated with Ce-pgp-12::GFP transgenic line, was fed on bacteria carrying the C26C6.5 RNAi vector. GFP expression in BC10210 was significantly attenuated or lost after RNAi treatment (20% penetrance as opposed to 0% with an empty vector control). RNAi by injection was also performed but with much higher penetrance (85% F1 progeny lost GFP expression as opposed to 0% in control (injection mixture only)) (Figure 29 A-F). The integrated strain, BC10210, gained modest expression in pharynx that was not seen in unstable strains. It is noteworthy that C26C6.5 specifically suppressed expression only in the excretory cell, but not those in pharynx, indicating its roles in specific up-regulation of *pgp-12* expression in the excretory cell. Taken together, I conclude that binding of the C26C6.5 to Ex-1 is both specific and essential for tissue specific expression of *pgp-12* in the excretory cell. I have tried to screen for genetic suppressors that inhibit the excretory cell expression by *pgp-12* using ethyl ethanesulfonate (EMS) mutagenesis. However, it turned out to be ineffective in that integrated BC10210 strain does not give 100% penetrance for GFP expression, so all the picked F1 and F2 no-expressing worms end up with false positive. Nevertheless, a couple of strains that gain expression in spermathecas were

achieved (data not shown). The molecular mechanism underling the gain of expression remains to be determined.

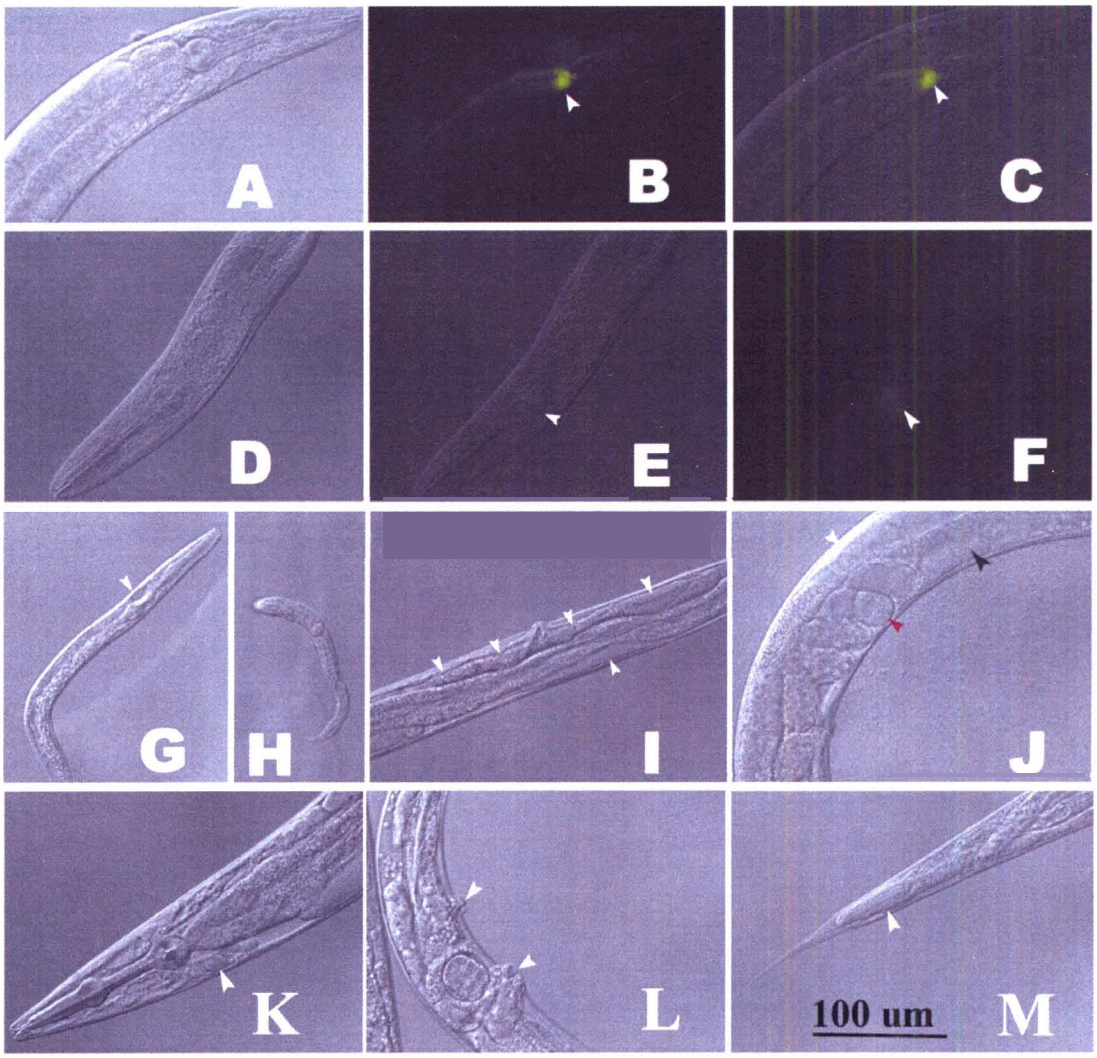
5.4.9 RNAi by injection against C26C6.5 causes defects in multiple tissues in addition to the excretory cell

In addition to specific suppression of excretory cell driven by *pgp-12* promoter, substantial embryonic lethality and larval arrest were observed as described previously (Simmer et al, 2003). Only ~20% of RNAi progeny can develop into adult, but most of them are sick and show obvious defects in excretory cell, germline and vulvae (Figure. 29G-M). These adults are sterile with a few dead embryos in uterus, indicating C26C6.5 is an essential gene. For example, a big cavity around excretory cell nucleus was observed which is believed to be degraded excretory cell body in some larval worms (Figure.29G). Some hatched worms arrested at different larval stages (Figure. 29H). Obvious under-developed germline was frequently seen (Figure. 29 I-J). Instead of the cell body and nucleus, which are normally beside the second pharyngeal bulb, I observed cystic swelling (liquid accumulation) around the excretory cell body in both RNAi progeny (Figure. 29K) and its parents (data not shown), which reinforces the roles of C26C6.5 in the morphogenesis as well as physiology of the excretory cell. Multiple vulvae were also seen in some F1 worms (Figure. 29L) while some other worms are vulvaless with embryo hatched inside parental worms (Figure. 29M). Given that C26C6.5 expresses in the excretory cell from larvae to adult and causes defects in the excretory cell by RNAi, I conclude that C26C6.5 is a novel trans-activator that up-regulates gene expression in and is involved in morphogenesis of the excretory cell. RNAi data

demonstrate that C26C6.5 is an essential gene that controls development of multiple tissues especially germline and vulva in addition to excretory cell.

Figure 29. RNAi against C26C6.5a suppress excretory cell specific expression and caused defects in multiple tissues

A, B and C: RNAi control (injection buffer only) by injection was performed on stable line expressing pgp-12::GFP (see text). A,B and C show Nomarski, GFP and superimposed pictures respectively; D (Nomarski), E(GFP) and F(superimposed): C26C6.5 RNAi causes specific loss of expression in excretory cell but not pharynx as indicated by arrowhead; G-M: RNAi against C26C6.5 (except for J) on N2 worms. G: degraded L3 excretory cell body was shown (by arrowhead); H: arrested L2 larvae. I and J: germline defects by C26C6.5 RNAi and control. The arrowheads indicated underdeveloped germline (I). The normal germline from control RNAi (J) with meiotic germ cell, oocyte and embryo indicated by white, red and black arrowheads. K: fluid accumulated around excretory cell body. L: multiple vulvae. M: Vulva-less parents with embryos hatched inside.


A
B
C
D
E
F
G
H
I
J
K
L
100 um
M

5.5 Discussion

5.5.1 *pgp-12* is expressed exclusively in the excretory cell at all developmental stages but is not involved in its morphogenesis

Eukaryotic gene transcription is regulated by a large variety of protein factors, including general transcription factors, specialized transcriptional factors such as activators and repressors responsible for temporal and/or spatial specific expressions, proteins involved in chromatin modifications and remodelling, and those which organize the genomic DNA inside nucleus (Orphanides et al., 1996; Blackwood and Kadonaga, 1998; Muller and Leutz, 2001; de Laat and Grosveld, 2003). Multicellular nematodes are composed of a variety of cell types that result from differential expression of various genes. The differential gene expression is mediated by various *cis*-elements that can recruit corresponding DNA-binding proteins such as transcription factors. Identification of those *cis*-elements as well their binding proteins will significantly improve the understanding of cell differentiation and tissue functions in multicellular organisms. It is well established that overcoming the ability of cancer cells to develop simultaneous resistance to multiple drugs, i.e., multidrug resistance (MDR), is the biggest challenge in cancer chemotherapy. MDR is found concurrent with increased human *pgp-1* mRNA level that results from gene amplification and/or increased gene transcription (Shen DW e al, 1986). Progress has been made on the transcription control of human *pgp-1* gene in the last decades. However, knowledge of the transcriptional regulation of pgp genes is far from being complete. Pgp gene in *C. elegans* is expanded over years and different paralogous members are deployed at various locations and/or stages, presumably to

provide more efficient protection against xenobiotcs (Zhao et al., 2004a). *pgp-12* is a paralogous pgp gene that shows exclusive expression in excretory cell. The *pgp-12* promoter, Ce-pgp-12, drove exclusive GFP expression in *C. elegans'* excretory cell at all developmental stages. Ce-pgp-12's orthologous region in *C. briggsae,* Cb-pgp-12, was also capable of driving GFP expression in *C. elegans'* excretory cell and vice versa, indicating the conserved regulation pathways for the pgp-12 and its ortholog expressions between the two species. Therefore, the *pgp-12* promoter and its *C. briggsae* orthologous sequence make a convenient system for mapping *cis*-elements responsible for tissue specific expression in the excretory cell. *pgp-12* is a member of the ABCB subfamily of the ATP Binding Cassette genes in *C. elegans* (Sheps et al., 2004). Loss of *pgp-12* function does not give any phenotypes under normal laboratory conditions, indicating *pgp-12* is not involved in morphogenesis but possibly in excretory cell physiology. Investigation of the transcription control of genes like *pgp-12* will provide insights into the mechanisms why genes are expressed in the excretory cell. This may also facilitate identification of novel transcription factors whose loss may result in hypomorphic phenotypes that could be missed by classic genetic screens. The excretory cell, a functional equivalent of vertebrate kidney is an "H" shaped single tubular cell extending from *C. elegans'* head to tail. The cell is an excellent model for investigating morphogenesis of biological tubes. Investigations into transcriptional control of excretory cell specific gene expression will facilitate the understanding of how these tubular organs are generated in metazoa. Also it may provide insights into the mechanisms for kidney development and its related diseases. Since *pgp-12* is a member of the ABCB subfamily, expression of *pgp-12* in the excretory cell suggests its roles in drug resistance based on

the characterized phenotypes of its mammalian homologues (Sheps, et al, 2004; Childs et al., 1994).

5.5.2 A 10 bp element, Ex-1, conserved between *C. elegans* and *C. briggsae* is critical for excretory cell specific expression

I isolated an excretory cell specific 10 bp *cis*-element, Ex-1, deletion of which caused complete loss of expression in the excretory cell at all developmental stages. However, Ex-1 alone was not sufficient to drive ectopic GFP expression, suggesting Ex-1 works synergistically with other *cis*-elements to drive tissue specific expression. Ex-L-Ex-1-Ex-R, in its endogenous orientation, was the minimum sequence capable of generating ectopic GFP expression in the excretory cell (Figure 22), suggesting that Ex-L and Ex-R facilitate tissue specific expression in the excretory cell. Gene expression, and in particular, transcription in eukaryotic cells is an important process that is regulated in complex ways, through intricate systems of mutual interactions between transcription factors, whose effects (activation/repression) are mediated via DNA binding sites present in their target genes. Within a multi-cellular organism, each cell type or tissue, at a specific developmental stage, has its own characteristic gene expression profile that is defined, at least in part, by the presence of a specific combination of transcription factors. It is possible that Ex-L and Ex-R are bound by distinct transcription factors that work cooperatively with C26C6.5. In agreement with this hypothesis, it was reported that the binding sites in the human *pgp-1* gene promoter could be bound by transcription factors that may act through competitive or cooperative interactions (Labialle S et al, 2002).

Aligning Ce-pgp-12 and Cb-pgp-12 gave perfect conservation of Ex-1 but not of Ex-L or Ex-R (Fig. 21B), suggesting that Ex-L and Ex-R are undergoing active evolution. I performed site-directed deletion mutagenesis of Ex-L and Ex-R in 475-Ce-pgp-12 (Figure 21A). However, unlike Ex-1, neither deletion could completely abolish excretory cell specific expression though both attenuated GFP intensity. Ex-R1 is more conserved than either Ex-L or Ex-R between *C. elegans* and *C. briggsae* (Figure 21B). Deletion of Ex-R1 gave a similar phenotype to those of deletions of Ex-L and Ex-R. Together, these data indicate that only Ex-1 is the dictating element for excretory cell specific expression. I found that, Ex-L-Ex-1-Ex-R, and the 44 bp region immediately upstream of Ex-L attached to the 5' end of Ce-haf-4 yielded comparable GFP expression to that of 475-Ce-pgp-12 in contrast to moderate GFP expression with Ex-L-Ex-1- Ex-R only (Figure 22, 24). In addition, the normal intestine expression of Ce-haf-4 was suppressed, this suppression was not observed for the hybrid promoter without the 44 bp sequence, suggesting suggests that other *cis*-elements, within the 44 bp conserved region, also contribute to *pgp-12* expression in the excretory cell. Further work needs to be done to identify relevant elements within the region as well as binding factors that facilitate expression in the excretory cell while inhibit expression in the intestine. There are two possible explanations for loss of expression in intestine by the hybrid promoter: first, Ex-L, Ex-1 and Ex-R plus the 44 bp upstream region compete with intestine specific elements within Ce-half-4 for basal transcriptional machinery, and the *trans*-acting factors binding the former sequences showed stronger affinity to basal transcription factors than those for the latter; second, the 44 bp sequence may work as a repressor for the intestine specific promoter. It will be interesting to experimentally explore the two possibilities. Moderate expression was observed in both excretory cell and intestine when

only Ex-L, Ex-1 and Ex-R were fused at the 5' end of intestine specific Ce-haf-4, implying that the two tissue specific elements are comparable to each other in strength, thus the tissue specific transcription factors for both tissues bind their corresponding elements and share the basal transcriptional machinery to initiate their own transcription. It is noteworthy that much of the upstream conserved region (Figure 22C circled region) seems to have little effect on excretory cell expression. This implies that *C. elegans* and *C. briggsae* have not diverged enough to facilitate identification of all *cis*-elements. Genomic sequencing of more related species should decrease such noise within these alignments and thus facilitate efficient identification of conserved *cis*-elements.

5.5.3 Divergence of excretory cell specific *cis*-elements between *C. elegans* and *C. briggsae*

Ex-1 plus its two 10 bp flanking sequences in Cb-pgp-12 were not able to cause ectopic GFP expression when they were fused on the 5' end of 286-Ce-haf-4 (Figure 18). This is consistent with the alignment of Ce-pgp-12 and Cb-pgp-12 in which only Ex-1 was perfectly conserved. However, Cb-pgp-12 was able to drive GFP expression in C. elegans excretory cell. This suggests that *C. elegans* and *C. briggsae* might use different combinations of *cis*-elements to drive similar expression patterns, or the excretory specific elements within the two promoters were re-positioned but still able to drive similar expressions. Further divergence of these *cis*-elements may lead to differential expression profiles and eventually different function of the orthologous genes. There is accumulating evidence that evolutionary changes in regulation sequences were involved in generating major evolutionary innovations and transitions. For example, it has been

shown that evolutionary innovation can be achieved by altered expression of a transcription factor without co-evolution of its target genes (Wang and Chamberlin, 2004). The different effects of *C. elegans* and *C. briggsae* elements on ectopic GFP expression might reflect active evolution divergence between the two genes.

5.5.4 Ex-1 independent expression in excretory cell

The Ex-1 and its *trans*-acting factor define excretory cell specific expression, yet I found Ex-1 independent excretory cell expression. For example, the promoters of another two excretory cell genes, ZK470.5 and C18C4.2, do not contain any apparent Ex-1 element. This suggests that it is possible that different combinations of *cis*-elements and *trans*-factors result in the same expressional output or multiple independent regulation pathways direct the similar expression pattern. In other words, co-expressed genes are not necessarily co-regulated. Further work needs to be done to investigate the underlying mechanisms for the apparent co-expression of the genes. Identification of the Ex-1 and its *trans*-activator represents only one class of regulatory mechanism that governs excretory cell specific expression. However, this protocol can be used as an entry point for investigating the transcriptional regulation that controls other tissue specific gene expression and eventually help establish regulatory network that is responsible for specific expression pattern.

5.5.5 C26C6.5 binds Ex-1 and works as an activator

C26C6.5 encodes a *trans*-activator for *pgp-12* expression based on the following observations: C26C6.5 protein contains a zinc finger DNA binding domain (predicted by

InterPro Mulder et al., 2003); RNAi against the gene significantly attenuated GFP expression driven by Ce-pgp-12; C26C6.5 binds Ex-1 *in vitro* and it expresses in the excretory cell in both larvae and adult. It will be interesting to determine whether the C26C6.5 protein works co-operatively with other transcriptional factors to drive endogenous and ectopic GFP expression in the excretory cell. Presence of coiled coils within the protein sequences strongly suggests the interaction of C26C6.5 with other yet identified partners. Further work need to be done to isolate its interacting partners and identify potential binding proteins for Ex-L and Ex-L1. A C26C6.5 human homolog, showing 25.3% identity to C26C6.5a has a different form of zinc finger domain (Feng et al., 2002). Presence of a zinc finger domain within a protein suggests its roles in DNA binding. The human homolog of C26C6.5 was identified as a component of a gene silencing complex and involved in MeCP1 complex targeting. A C26C6.5 homolog in *Xenopus* had been identified as a component of the Mi-2 complex that couples DNA methylation to chromatin remodelling and histone deacetylation (Wade et al., 1999). Lack of the C26C6.5 homolog in S*accharomyces Cerevisiae* suggests that C26C6.5 is likely involved in multicellular processes rather than a ubiquitous silencer for eukaryotes. In contrast to the existing functional information for C26C6.5 homologs, I demonstrated that C26C6.5 functions as a *trans*-acting activator rather than a suppressor for excretory cell expression. It is yet to be determined whether C26C6.5 works as a repressor in some other cellular context in *C. elegans*. However, it is possible that the C26C6.5 protein plays dual roles as both suppressor and activator. Such dual roles are observed for suppressor protein p53. It is well established that p53 regulates cell growth and death decisions through its role as a transcriptional activator. However, p53 can also repress the transcription of a number of genes, including those involved in regulatory cascades

mediating cell proliferation and tumorigenesis (el-Deiry, WS, 1998). For example, p53 can repress transcription directly by binding to a novel head-to-tail (HT) site within the MDR1 promoter in an orientation-dependent manner (Johnson RA, et al 2001). Stronger expression of C26C6.5a in larvae than in adult excretory cells as well as embryonic expression implies that the C26C6.5a protein may be involved in the morphogenesis of the excretory cell. This is consistent with the observations that RNAi causes cystic swelling of the excretory cell body. C26C6.5 was also found in vulvae, pharynx and posterior neurons, suggesting its possible regulatory roles in these tissues. This is in agreement with the observations that RNAi against C26C6.5 resulted in slow growth and vulvae defects (Fraser et al., 2000; Simmer et al., 2003). Incomplete overlapping patterns of C26C6.5 and *pgp-12* suggest that C26C6.5 may have roles in multiple tissues in combination with other *trans*-factors present in specific cellular context.

CHAPTER 6 GENERAL DISCUSSION

This thesis has focused on functional analysis of ABC transporter genes in *C. elegans*. I concentrated on four aspects of the family: phylogenetic and comparative analyses of the whole ABC family between *C. elegans* and *C. briggsae*; characterization of the single member of the E subfamily, ABCE, due to its apparently improper annotation; expression analysis of the whole family, which leads to the finding that locally duplicated genes stay functionally active probably by subfunctionalization; and characterization of a transcription regulation pathway for excretory cell-specific expression using *pgp-12* promoter.

6.1 ABCE is an essential gene that functions more than as an RNase L inhibitor

ABC transporters constitute a big gene family that is well conserved in both prokaryotes and eukaryotes. Sixty ABC genes have been identified in the *C. elegans* genome which can be classified into eight subfamilies based on homology to those in humans (Figure 3) (Sheps, et al, 2004).

Most of the worm ABC genes have little functional information. Comparison of ABC transporter family in *C. elegans* and those in newly sequenced *C. briggsae*, a sister species that diverged from *C. elegans* more than 50 million years ago revealed extensive conservation, not only in gene number but also in genomic arrangement of the family. One particular ABC protein that caught my attention is the only member in the subfamily E, ABCE protein, because of the following observations: it is not a canonical ABC

transporter due to the abssence of transmembrane domain; it is well-conserved across species, showing 95% and 65% identity between its orthologs in *C. briggsae* and human respectively; no apparent paralog has been found except in *Arabidopsis* which encodes many more ABC proteins, i.e., 129 putative ABC transporters that other eukaryotic species do (Sanchez-Fernandez et al, 2001); ABCE is annotated as ribonuclease L (RNAse L) inhibitor, but such RNAse L is not present in *C. elegans*. Based on these observations, it does not make sense to annotate ABCE protein as RNAse L inhibitor. I propose that ABCE is an essential gene that functions as something else in addition to an RNAse L inhibitor in eukaryotes. RNAi by feeding with *eri-1* mutant does show embryonic lethality, but penetrance is around 20%, which makes me suspicious of the effectiveness of this type of RNAi given the extensive neuronal expression of the gene (data not shown). So, RNAi by injection was performed which is targeting first two exons of the ABCE gene with the least homology in the genome. The RNAi phenotypes penetrance is much higher as opposed to RNAi phenotypes by feeding, i.e., 100% vs 20%. RNAi progeny are arrested at different stages with a few developing into sick sterile adults, clearly indicating the ABCE is an essential gene. This is consistent with the fact that loss of ABCE in yeast caused lethality (Giaever G, et al, 2002). Death is expected with the knockout of the sole copy of ABCE in the worm genome. Interaction of ABCE with a ribosome protein, RPL-4 and a transcriptional factor, NHR-91 implies its role in both translation and transcription. Roles of ABCE protein in the translation initiation and ribosome biogenesis have recently been demonstrated in yeast (Dong et al, 2004). It is noteworthy that members of another non-transporter subfamily F may also be involved in protein translation. For example, mammalian ABC50 protein, a member of the subfamily F interacts with eIF2 and associates with ribosome (Tyzack et al, 2000). It is likely that

the two subfamilies, both having no transmembrane domains but two nucleotide binding ones, function in similar pathways. Interaction of ABCE with ribosomal protein and transcriptional factors implies that ABCE possibly works as a shuttle protein between cytoplasm and nuclei, given the subcellular localization of the ribosomal protein and transcriptional factors. This hypothesis is in agreement with PSORT prediction of worm ABCEs' subcellular localization, i.e., both cytoplasmic and nuclear. Yeast ABCE ortholog was also observed in both cytoplasm and nucleus. Presence in eukaryotes but not prokaryotes suggests that the ABCE may play roles unique to the eukaryotic organisms. The primary distinction between the prokaryotic and eukaryotic organisms is the compartmentalization of nucleus from its remaining cytoplasm in the latter, which makes gene transcription and protein translation separate from each other. Extra efforts have to be made to deliver compounds across the nuclear membrane, which might be an active process in many cases. Given the presence of ABCE in both cytoplasm and nuclei as well its ATP binding domain, I propose that ABCE is possibly involved in active substrate delivery across nuclear membrane, such as transporting nuclear protein from cytoplasm into nucleus or in opposite direction for mRNA. Preliminary results came from the RNAi against ABCE using transgenic strains expressing full-length HIS-72::GFP. It seems that the loss of ABCE affects the nuclear distribution of HIS-72, which might be a result of defective transport of HIS-72 into the nucleus. Interaction of yeast ABCE with nuclear pore complex protein Nup116, a protein involved in nucleocytoplasmic transport (Ito et al, 2001), also suggests the possible roles of ABCE in mRNA export. Co-localization of *C. elegans* ABCE and nuclear pore complex protein F53F10.5 (data not shown) also suggests the role of ABCE in substrate transport across nuclear membrane. Further work

needs to be done to determine whether ABCE could bind mRNA directly or indirectly and whether it is co-localized with mRNA binding and transporting components.

6.2 Paralogous ABC transporters may function by splitting the roles of its ancestor

Given the fact that ABC family is a big gene family with sixty members, most of which have little functional information, I start by asking when and where these ABC genes are expressed *in vivo*. Using fusion PCR technique, I generated the promoter::GFP and/or promoter::DsRed fusion constructs for the whole family and used them to produce transgenic animals. The strength of the assay is that expression patterns resulting from individual promoters could be examined *in vivo* and in real time due to the transparency of the worm body. Most ABC promoters can successfully drive observable GFP expression. As expected, many members of subfamily B and C are frequently seen in intestine, excretory cell, pharynx and neurons, consistent with its possible roles as transporters for a variety of substrates. Expression patterns of ABC genes constitute the first step to understand the function of these genes, especially those without any clue of its roles in *C. elegans*. Comparison of the ABC family between genomes reveals much gene duplication and loss in the family (Sheps et al, 2004). Species-specific expansion is found in different subfamilies. For example, *Drosophila* evolves a big subfamily G while nematode shows an expansion in subfamily B (Table 2). There are 16 ABC genes existing as two- or four-gene tandem clusters in the *C. elegans* genome, which are apparently duplicated copies (Figure 16). Investigation of the expression patterns of the whole family will be able to provide insights into why the tandem repeats have been retained in the nematode genomes which are diverged over 100 millions of years (Stein et al, 2003),

e.g., between *C. elegans* and *C. briggsae*. It will be intriguing to examine whether the species-specific expansion or loss of ABC genes in particular subfamily reflects biological changes. It is generally accepted that duplicated paralogs are under little selection pressure and most of them become pseudogenes with time. However, my expression data show that fifteen out of the sixteen duplicated ABC genes gave observable expression, suggesting most of these duplicated ABC genes are functionally active ones. The question is why worm maintains multiple copies of ABC genes if they play similar roles? All the duplicated ABC sequences do have the highest similarity to those of its adjacent paralogs, but the differential expression patterns among paralogs within the same cluster suggest that these duplicated ABC genes primarily diverge from each other by their regulatory regions rather than protein sequences. *C. elegans* contains twenty-four members of ABCB subfamily, fourteen of which are among these tandem repeats. It is frequently observed that members within the same cluster tend to be expressed in different tissues or at distinct stages. For example, F22E10.1, 2,3 and 4 are four adjacent pgp genes on the same cosmid that show expression in excretory cell (a likely equivalent of human kidney), intestine, pharynx and neurons respectively. The human genome has one known member of ABCB subfamily involved in drug resistance, i.e. *pgp-1*, which is expressed in adrenal, intestine, kidney, brain (Dean et al, 2001). Based on these observations, I propose that duplicated pgp genes may work synergistically by dividing the roles of their ancestor through evolving different regulatory regions. This is in agreement with a recent hypothesis that predicts that the usual mechanism of duplicate gene preservation is the partitioning of ancestral functions rather than the evolution of new functions (Force, et al., 1999). Tissue specific distributions controlled by individual regulatory region may provide more flexibility

compared to a compact regulatory sequence that controls multiple tissue expression. Given the toxic soil where nematodes worm dwell, presence of multiple paralogs of drug resistance genes may provide more effective protection against xenobiotics.

We requested knockout mutants from *C. elegans* Knockout Consortium for all ABC genes. Eighteen strains that contain a single ABC gene mutation are currently available. Most of these ABC are members of subfamily B. All of these mutants are wild-type looking under normal laboratory conditions, indicating that none of these mutants are essential for normal growth except for ced-7, which was reported to be involved in cell death pathway (Wu et al, 1998). ABC members in subfamily B have undergone frequent duplication as evidenced by half members of the subfamily (12 members) are tandemly duplicated genes. The genes within the same cluster tend to be differentially expressed. However, those not in the same cluster still show apparent overlapping expressions. Whether these overlapping patterns by different pgp genes reflect the functional redundancy remains to be determined. This may partially explain why mutation in single or double pgp genes does not give obvious phenotypes. It has been demonstrated that some of the ABC genes function under certain conditions, i.e., inducible expresison. For example, *Pgp-1* and *pgp-3* are involved in resistance to chloroquine and colchicine (Broeks, et al., 1995). ABCC subfamily members *mrp-1* and *pgp-1*are both involved in heavy metal resistance (Broeks, et al, 1996). Future work needs to be done to determine whether the paralogous ABC genes are functionally accumulative by combination of available mutant and RNAi.

6.3 Investigation of transcriptional regulation of ABC drug transporters is just a start

Much effort has been made to search for inhibitors for PGP-1 since it was first isolated by the virtue of its overexpression in multidrug resistance (MDR) tumor cells (Juliano and Ling, 1976). However, it turns out to be not very successful due to the complexity of MDR that was not initially anticipated. Regulation of MDR can take place at multiple levels. There is emerging evidence that expression of the ABC transporters is highly regulated, particularly at the level of transcription, suggesting a future target for modulation of the MDR phenotype (Scotto, 2003). Investigation of transcriptional of ABC transporters has been focusing on human MDR-1 gene (hMDR-1/PGP-1) and much progress has been made during last two decades. It is interesting to note that all of the human drug-related transporters lack an appropriately positioned TATA box, while several of their rodent homologues are TATA-dependent. Like most "TATAless" genes, the hMDR1 promoter contains both an inverted CCAAT box that interacts with trimeric transcription factor NF-Y (Hu, et al, 2000) and a GC-rich element that interacts with Sp1 and Sp3 (Cornwell and Smith, 1993a,b; Sundseth et al, 1997). Roles of p53 in the regulation of hMDR1 remain controversial due to contradicting results reported by different groups (Nguyen et al, 1994; Goldsmith et al, 1995). Direct or indirect evidence suggests the activation of hMDR-1 by AP-1 (Daschner et al, 1999; Labialle et al, 2002). It has been shown that heat shock up-regulates the transcription of hMDR1 (Chin et al, 1990). Several putative HSE sequences have been identified in its promoter, suggesting hMDR1 may also work as stress response gene. Rapid advances in the transcription field indicate that knowledge of the transcription regulation of hMDR1 is far from complete. Searching for hMDR1 specific transcriptional suppressor remains a big challenge.

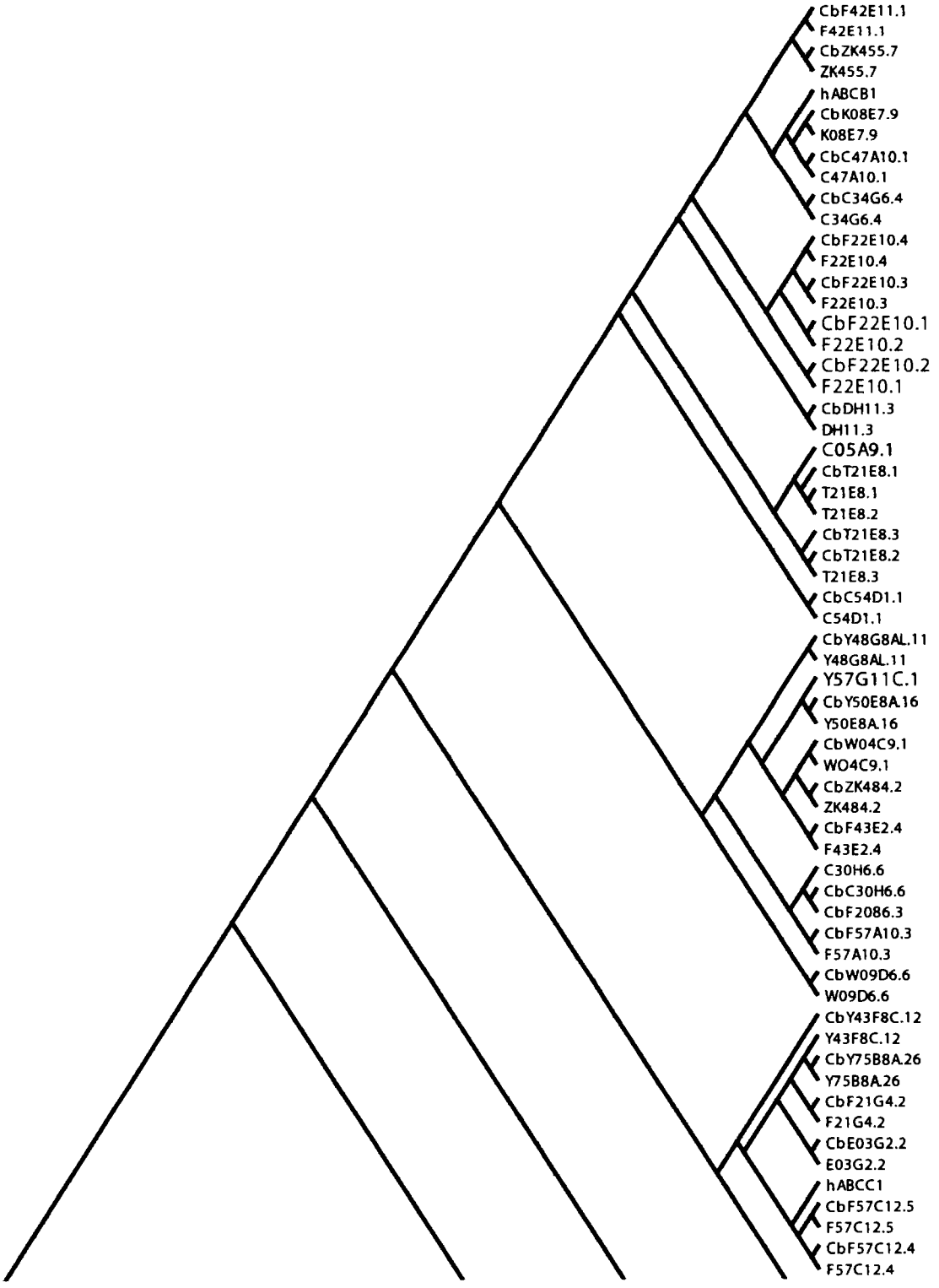
Regulation of other MDR genes lags far behind, which reinforces the necessity to improve the understanding of regulation on the hMDR1 related genes.

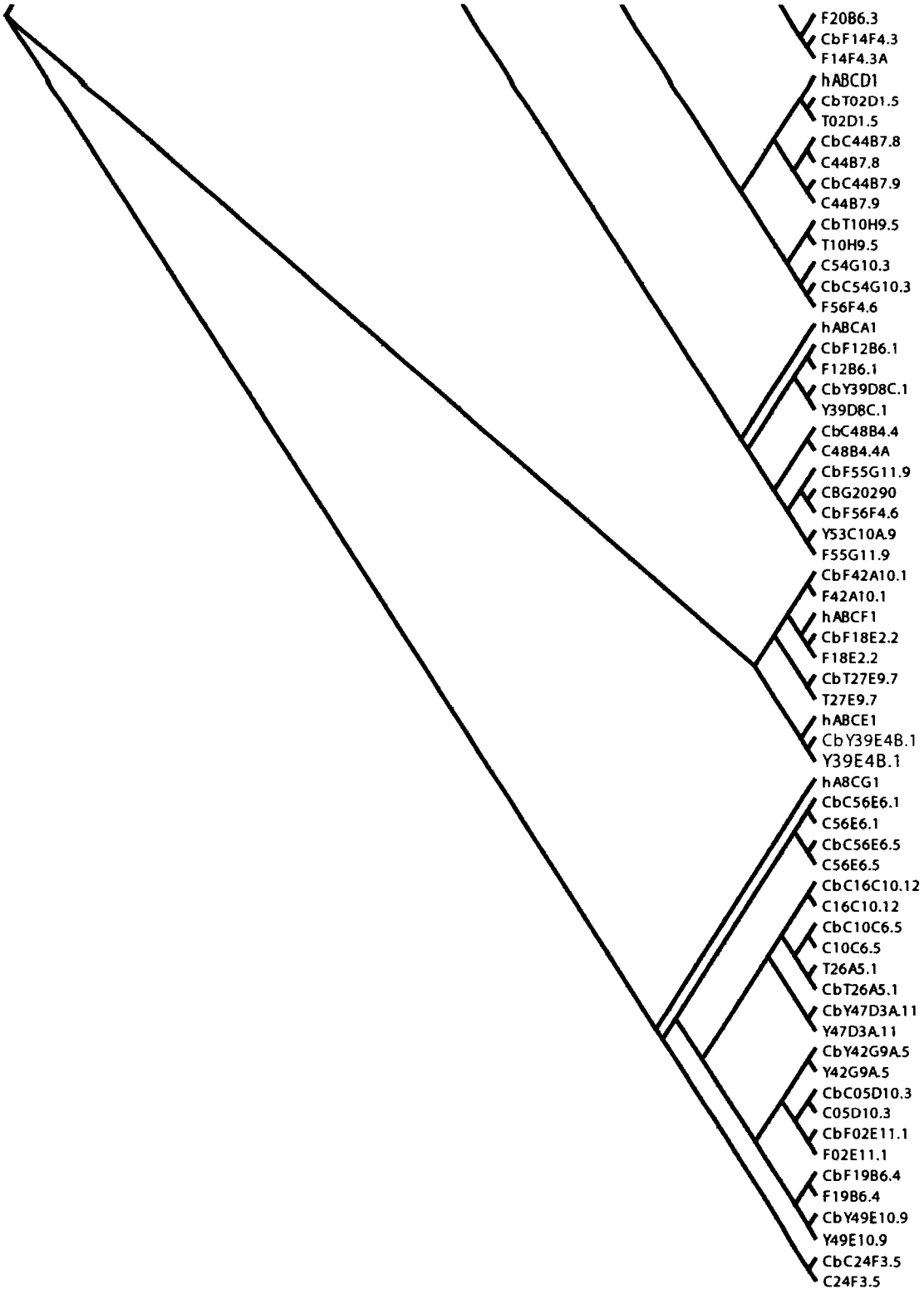
Human MDR related genes are significantly expanded in *C. elegans* (Sheps et al, 2004). *pgp-12* is one of these genes whose 286 bp promoter drives exclusive GFP expressions in excretory cell, a likely functional equivalent of human kidney. It is reasonable to speculate that *pgp-12* is a potential drug transporter because human MDR1 is also expressed in kidney (Dean et al, 2001), where excretion of a lot of toxins take place. Scanning of the 286 bp *pgp-12* promoter by MatInspector (Quandt et al, 1995) indicated the presence of both TATA and CAAT box. I have identified a 10 bp element, termed Ex-1 that is critical for excretory cell specific expression. Together with its flanking 10 bp on each side, they can make intestine specific promoter drive GFP expression in both excretory cell and intestine if they were fused on the 5' end the promoter, indicating the significance of the 30 bp in driving excretory cell specific expression. A putative transcription factor, C26C6.5 has been identified by yeast one-hybrid screen that specifically bind the Ex-1. Binding of C26C6.5 to Ex-1 was further confirmed by gel shift assay *in vitro*. Significance of C26C6.5 in regulation of *pgp-12* expression was demonstrated by *in vivo* assay, i.e., C26C6.5 RNAi suppressed the GFP expression driven by *pgp-12* promoter. Clearly, C26C6.5 works as up-regulator of *pgp-12* expression. However, C26C6.5 homolog has been identified as a component of the Mi-2 complex that couples DNA methylation to chromatin remodelling and histone deacetylation (Wade et al, 1999), implying C26C6.5 is a gene silencer. The paradox between annotation and my data could be partially explained by the multiple alignments among C26C6.5 protein sequence and its homologs in other species. It is obvious that

C26C6.6 shows conservation in DNA binding domain but not in the putative regulatory domains that are only conserved between vertebrate homologues (Figure 25). Conservation of DNA binding domain suggests it works as transcription factor. Failure to show alignment within the remaining sequences implies that C26C6.5 is not a functional equivalent of its vertebrate homologs, but is the *trans*-activator for *pgp-12* expression. However, the possibility that C26C6.5 works as both activator and silencer cannot be ruled out because it is expressed not only in the excretory cell, but also in pharynx, vulva and neurons. Further work needs to be done to characterize the function of this putative transcription factor. For example, is C26C6.5 involved in deacetylation of histones or connected with methylation of DNA as it is in *Xenopus* (Wade et al, 1999). Generation and characterization of null mutants will definitely pave the way to understand the functions of the gene. The excretory cell is a highly tubular cell that extends from head to tail, and investigation of excretory cell specific transcriptional control may improve the understanding of morphogenesis of biological tubes.

APPENDICES

Appendix 1. Most *C. elegans* ABC proteins cluster with their orthologs in *C. briggsae*. Human ABC proteins were included in each corresponding subfamily and prefixed with "h". *C. briggsae* orthologs were prefixed with "Cb" where applicable. Assignment of orthologs for F22E10.1 and F22E10.2 is based on synteny and not in agreement with tree data (highlighted in red color), indicating duplicated genes may not have the highest sequence similarity with its orthologs. This assignment was substantiated by GFP expression data. A single member of ABCE is highlighted in blue colour. Bootstrapped numbers are not shown due limited space.


CbF42E11.1
F42E11.1
CbZK455.7
ZK455.7
hABCB1
CbK08E7.9
K08E7.9
CbC47A10.1
C47A10.1
CbC34G6.4
C34G6.4
CbF22E10.4
F22E10.4
CbF22E10.3
F22E10.3
CbF22E10.1
F22E10.2
CbF22E10.2
F22E10.1
CbDH11.3
DH11.3
C05A9.1
CbT21E8.1
T21E8.1
T21E8.2
CbT21E8.3
CbT21E8.2
T21E8.3
CbC54D1.1
C54D1.1
CbY48G8AL.11
Y48G8AL.11
Y57G11C.1
CbY50E8A.16
Y50E8A.16
CbW04C9.1
WO4C9.1
CbZK484.2
ZK484.2
CbF43E2.4
F43E2.4
C30H6.6
CbC30H6.6
CbF2086.3
CbF57A10.3
F57A10.3
CbW09D6.6
W09D6.6
CbY43F8C.12
Y43F8C.12
CbY75B8A.26
Y75B8A.26
CbF21G4.2
F21G4.2
CbE03G2.2
E03G2.2
hABCC1
CbF57C12.5
F57C12.5
CbF57C12.4
F57C12.4


F20B6.3
CbF14F4.3
F14F4.3A
hABCD1
CbT02D1.5
T02D1.5
CbC44B7.8
C44B7.8
CbC44B7.9
C44B7.9
CbT10H9.5
T10H9.5
C54G10.3
CbC54G10.3
F56F4.6
hABCA1
CbF12B6.1
F12B6.1
CbY39D8C.1
Y39D8C.1
CbC48B4.4
C48B4.4A
CbF55G11.9
CBG20290
CbF56F4.6
Y53C10A.9
F55G11.9
CbF42A10.1
F42A10.1
hABCF1
CbF18E2.2
F18E2.2
CbT27E9.7
T27E9.7
hABCE1
CbY39E4B.1
Y39E4B.1
hA8CG1
CbC56E6.1
C56E6.1
CbC56E6.5
C56E6.5
CbC16C10.12
C16C10.12
CbC10C6.5
C10C6.5
T26A5.1
CbT26A5.1
CbY47D3A.11
Y47D3A.11
CbY42G9A.5
Y42G9A.5
CbC05D10.3
C05D10.3
CbF02E11.1
F02E11.1
CbF19B6.4
F19B6.4
CbY49E10.9
Y49E10.9
CbC24F3.5
C24F3.5

Appendix 2: Expression patterns of ABC transporters in *C. elegans*

Sub family	Gene Name	DsRed Expression patterns	GFP Expression Patterns	SIZE DsRed	SIZE GFP
A	F12B6.1	No observable expression, 3 isolates	No observable expression, 2 isolates	857	1060
A	C48B4.4	No observable expression, 2 isolates	No observable expression, 2 isolates	2063	2115
A	F55G11.9	N/A	No observable expression, 1 isolates		831
A	F56F4.6	N/A	No observable expression, 3 isolates		930
A	Y53C10A.9	N/A	Weak in posterior gut, 2 isolates		1619
A	Y39D8C.1	N/A	No observable expression, 2 isolates		2476
A	C24F3.5	N/A	No observable expression, 5 isolates		1282
B	F22E10.1	Excretory cell at all stages, 8 isolates	Excretory cell at all stages, 3 isolates	3422	3422
B	F22E10.2	Medium in adult larval posterior gut,amphid, 2 isolates	No observable expression, 2 isolates	1667	2727
B	F22E10.3	Strong in anterior and 1st bulb of pharynx, 3 isolates	Strong in anterior pharynx, 2 isolates	1724	1003
B	F22E10.4	Head & tail neurons, 2 isolates	Head & tail neurons, 2 isolates	2953	3163
B	T21E8.1	Larval, adult amphid, weak in adult gut, 2 isolates	Weak in adult anterior gut, 3 Isolates	2949	1934
B	T21E8.2	Not observed	Male tail	1089	1025
B	T21E8.3	Head neuron	Head neorons	689	689
B	C05A9.1	Weak in gut , 1 isolate	Weak in gut, 3 isolates	498	498
B	C54D1.1		Medium in gut, 3 isolates		1334
B	DH11.3	N/A	Excretory cell, gut, 2 isolates		1800
B	K08E7.9	N/A	Gut		2140
B	C47A10.1	N/A	Pharynx 1st, 2ed bulb, strong in gut, 3 isolates		1802
B	C34G6.4	N/A	Pharynx 1st and 2ed bulb		3170
B	ZK455.7	Weak in adult excretory cell, 2 isolates	Excretory cell, gut, 4 isolates	1051	1047
B	F42E11.1	Medium excretory cell, 1 isolates	Weak in excretory cell, 3 isolates	2145	2138
B	F43E2.4	N/A	Embryo, adult muscle		2511
B	W04C9.1	N/A	Medium in gut, 2 isolates		286
B	ZK484.2	N/A	Strong in gut, 4 isolates		2905
B	Y50E8A.26	N/A	Pharynx and gut, 2 isolates		1980

B	Y57G11C.1	N/A	No observable expression, 3 isolates		3018
B	Y48G8AL.11	No observable expression, 1 isolate	Fail to generate transgenic worm	2992	707
B	C30H6.6	N/A	No observable expression, 2 isolates		1564
B	F57A10.3	N/A	Gut, spermatheca		944
B	W09D6.6	N/A	Strong in adult gut		1674
C	F14F4.3	N/A	Pharynx, gut, neurons, muscle, hypodermis		2829
C	F20B6.3	N/A	No observable expression, 3 isolates		1991
C	F57C12.4	Strong in pharynx, 5 isolates	Strong in pharynx, 5 isolates	2924	1999
C	F57C12.5	Weak in adult gut,vulva and pharynx, 2 isolates	Pharynx, gut, vulva	2925	2925
C	E03G2.2	N/A	No observable expression, 2 isolates		2890
C	F21G4.2	N/A	Embryo, adult & larval gut, 3 isolates		1930
C	Y43F8C.12	N/A	Strong in anterior gut, muscle, larval intestine		2168
C	Y75B8A.26	No observable expression, 2 isolates	No observable expression, 4 isolates	2935	1999
C	C18C4.2	N/A	Excretory cell		1195
D	T10H9.5	N/A	Strong in embryo, larvae, adult gut		1652
D	C54G10.3	Adult, larval gut, vulva, hypoderm, neurons,1 isolate	No observable expression, 2 isolates	2914	1519
D	T02D1.5	N/A	Medium in adult and larvae gut		1361
D	C44B7.8	Strong in adult, larval gut, embryo, 2 isolates	Strong in larval but weak in adult gut, 1 isolate	1193	1856
D	C44B7.9	Strong in adult and larvae gut	Strong in adult and larvae gut	981	2027
E	Y39E4B.1	N/A	Hypoderm, pharynx, vulvae,muscle, nerve system	2964	2964
F	T27E9.7	N/A	No observable expression, 3 isolates		679
F	F42A10.1	Strong in larval and adult gut, 2 isolates	No observable expression, 2 isolates	2928	858
F	F18E2.2	N/A	Strong in embryo, weak in gut, 3 isolates		635
G	Y42G9A.6	No observable expression, 1 isolates	No observable expression, 1 isolates	2908	915
G	C05D10.3	N/A	Strong even in embryo, male gut		1483
G	F02E11.1	No observable expression, 2 isolates	No observable expression, 6 isolates	2905	1909

G	F19B6.4	No observable expression, 1 isolates	No observable expression, 2 isolates	2975	2188
G	Y49E10.9	No observable expression, 1 isolate	Fail to generate transgenic worm	719	250
G	Y47D3A.11	No observable expression, 1 isolates	No observable expression, 1 isolates	2949	976
G	C16C10.12	No observable expression, 1 isolates	No observable expression, 3 isolates	2918	572
G	C10C6.5	N/A	Larval & adult gut, 4 isolates		2445
G	T26A5.1	N/A	No observable expression, 5 isolates		1071
H	C56E6.1	No observable expression, 3 isolates	No observable expression, 3 isolates	2954	673
H	C56E6.5	N/A	Adult, larvae head, tail neuron, adult vulve	639	

REFERENCES

Allikmets, R. (2000) Further evidence for an association of ABCR alleles with age-related macular degeneration. The International ABCR Screening Consortium. *Am J Hum Genet.* **67**(2):487-91.

Annilo, T., Chen, Z. Q., Shulenin, S. and Dean, M. (2003) Evolutionary analysis of a cluster of ATP-binding cassette (ABC) genes. *Mamm. Genome* **14**: 7-20.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M. and Sonnhammer, E. L. L. (2002) The Pfam protein families database. *Nucleic Acids Research* **30**: 276-280.

Bauer B.E., Wolfger H, Kuchler K. (1999) Inventory and function of yeast ABC proteins: about sex, stress, pleiotropic drug and heavy metal resistance. *Biochim Biophys Acta* **1461**(2):217-36.

Birchall, P. S., Fishpool, R. M., Albertson, D. G. (1995) Expression patterns of predicted genes from the C. elegans genome sequence visualized by FISH in whole organisms. *Nature Genetics* **11** 314-20.

Bisbal, C., Martinand, C., Silhol, M., Lebleu, B., Salehzada, T. (1995) Cloning and characterization of a RNase L inhibitor. A new component of the interferon-regulated 2-5A pathway. *J. Biol.Chem.* **270**: 13308–13317.

Bisbal, C., Silhol, M., Laubenthal, H., Kaluza, T., Carnac, G., Milligan, L., Le Roy, F. and Salehzada, T. (2000) The 2'-5' oligoadenylate/RNase L/RNase L inhibitor pathway regulates both MyoD mRNA stability and muscle cell differentiation. *Mol. Cell. Biol.* **20**: 4959-4969.

Blackwood, E. M., Kadonaga, J. T. (1998) Going the distance: a current view of enhancer action. *Science.* **281**: 61–63.

Blumenthal, T., Evans, D., Link, C. D., Guffanti, A., Lawson, D., Thierry-MIieg, J., Thierry-Mieg, D., Chiu, W. L., Duke, K., Kiraly, M. and Kim, S. K. (2002) A global analysis of *Caenorhabditis elegans* operons. *Nature* **417**: 851-854.

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

Broeks, A., Gerrard, B., Allikmets, R., Dean, M. and Plasterk, R. H. A.. (1996) Homologues of the human multidrug resistance genes MRP and MDR contribute to heavy metal resistance in the soil nematode Caenorhabditis elegans. *EMBO J.* **15**: 6132-6143.

Broeks, A., Janssen, H. W. R. M., Calafat, J. and Plasterk, R. H. A. (1995) A P-glycoprotein protects Caenorhabditis elegans against natural toxins. *EMBO J.* **14**: 1858-1866.

Brown, P. O. and Botstein D. (1999) Exploring the new world of the genome with DNA microarrays. *Nat Genet* **21**: 33-37.

Buechner, M., Hall, D. H., Bhatt, H., Hedgecock, E. M. (1999) Cystic canal mutants in C. elegans are defective in the apical membrane domain of the renal (excretory) cell. *Dev. Biol.* **214**(1): 227-241.

Burge, C. B. and Karlin, S. (1997) Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology* **268**: 78-94.

Bürglin, T. R., Ruvkun, G. (2001) Regulation of ectodermal and excretory function by the C. elegans POU homeobox gene ceh-6. *Development* **128**: 779–790.

Calapez, A., Pereira, H. M., Calado, A., Braga, J., Rino, J., Carvalho, C., Tavanez, J. P., Wahle, E., Rosa, A. C., Carmo-Fonseca, M. (2002) The intranuclear mobility of messenger RNA binding proteins is ATP dependent and temperature sensitive. *J. Cell Biol.* **159:** 795–805.

Castillo-Davis CI, Hartl DL, Achaz G. (2004) cis-Regulatory and protein evolution in orthologous and duplicate genes. *Genome Res.* **14**(8):1530-6.

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

Chen H, Rossier C, Lalioti MD, Lynn A, Chakravarti A, Perrin G, Antonarakis SE.(1996) Cloning of the cDNA for a human homologue of the Drosophila white gene and mapping to hromosome 21q22.3.*Am J Hum Genet.***59**(1):66-75.

Cherry, J. M., Ball, C., Dolinski, K., Dwight, S., Harris, M., Matese, J. C., Sherlock, G., Binkley, G., Jin, H., Weng, S. and Botstein, D. (2002) Saccharomyces Genome Database.

Chervitz, S. A., Aravind, L., Sherlock, G., Ball, C. A., Koonin, E. V., Dwight, S. S., Harris, M. A., Dolinski, K., Mohr, S., Smith, T., Weng, S., Cherry, J. M. and Botstein, D. (1998) Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science* **282**: 2022-2028.

Childs, S. and Ling, V. (1994) The MDR superfamily of genes and its biological implications. In Important advances in oncology (eds. DeVita, V. T., Hellman, S. and Rosenberg, S. A.), pp. 21-36. J.B. Lippincott Co.

Chin, K. V., Tanaka, S., Darlington, G., Pastan, I., Gottesman, M. M. (1990) Heat shock and arsenite increase expression of the multidrug resistance (MDR1) gene in human renal carcinoma cells. *J Biol Chem.* **265**(1): 221-6.

Coghlam, A and Wolfe, K.H. (2002) Fourfold faster rate of genome rearrangement in nematodes than in Drosophila. *Genome Res.***12**(6):857-67.

Copley, R. R., Letunic, I. and Bork, P. (2002) Genome and protein evolution in eukaryotes. *Curr Opin Chem Biol* **6**: 39-45.

Cornwell, M. M., Smith, D. E. (1993) SP1 activates the MDR1 promoter through one of two distinct G-rich regions that modulate promoter activity. *J Biol Chem.* **268**(26): 19505-11.

Croop, J. M. (1998) Evolutionary relationships among ABC transporters. *Methods Enzymol* **292**: 101-116.

Cui, M and Han M. (2003). *Cis*-regulatory requirements for vulval cell-specific expression of the Caenorhabditis elegans fibroblast growth factor gene egl-17. *Dev Biol.* **257**(1):104-16.

Dandekar, T., Snel, B., Huynen, M. and Bork P. (1998) Conservation of gene order: A fingerprint of proteins that physically interact. *Trends in Biochemical Sciences* **23**: 324-328.

Daschner, P. J., Ciolino, H. P., Plouzek, C. A., Yeh, G. C. (1999) Increased AP-1 activity in drug resistant human breast cancer MCF-7 cells. *Breast Cancer Res Treat.* **53**(3): 229-40.

de Laat, W., Grosveld, F. (2003) Spatial organization of gene expression: the active chromatin hub. *Chromosome Res.* **11**: 447–459.

Dean, M. and Allikmets, R. (1995) Evolution of ATP-binding cassette transporter genes. *Current Opinion in Genetics and Development* **5**(6):779-85 .

Dean, M., Rzhetsky, A., Allikmets, R. (2001) The human ATP-binding cassette (ABC) transporter superfamily. *Genome Res.* **11**(7): 1156-66. Review.

Dean, M. (2002) The human ATP-binding cassette (ABC) transporter superfamily. Available through NCBI bookshelf.

Decottignies,A and Goffeau,A. (1997) Complete inventory of the yeast ABC proteins. *Nat Genet.* **15**(2):137-45.

Deleuze, J. F., Jacquemin, E., Dubuisson, C., Cresteil, D., Dumont, M., Erlinger, S., Bernard, O. and Hadchouel, M. (1996) Defect of multidrug-resistance 3 gene expression in a subtype of progressive familial intrahepatic cholestasis. *Hepatology* **23**: 904–908.

Dixon, P. H., Weerasekera, N., Linton, K.J., Donaldson, O., Chambers, J., Egginton, E., Weaver, J., Nelson-Piercy, C., De Swiet, M., Warnes, G., et al. (2000) Heterozygous MDR3 missense mutation associated with intrahepatic cholestasis of pregnancy:Evidence for a defect in protein trafficking. *Hum. Mol. Genet.* **9**: 1209–1217.

Dong, J., Lai, R., Nielsen, K., Fekete, C. A., Qiu, H. and Hinnebusch, A. G. (2004) The Essential ATP-binding Cassette Protein RLI1 Functions in Translation by Promoting Preinitiation Complex Assembly. *J. Biol. Chem.* **279**: 42157- 42168.

Douglas S, Zauner S, Fraunholz M, Beaton M, Penny S, Deng LT, Wu X, Reith M, Cavalier-Smith T, Maier UG: (2001). The highly reduced genome of an enslaved algal nucleus. *Nature* **410**: 1091-1096.

Du, X. L., Wang, D., Qian, X. Y., Jiang, L. Z., Chun, W., Li, K. G., Shen, G. A., Lin, C. F., Yang, J. S. (2003) cDNA cloning and expression analysis of the rice (Oryza sativa L.) RNase L inhibitor. *DNA Seq.* **14**(4) 295-301.

el-Deiry, W. S. (1998) Regulation of p53 downstream genes. *Semin. Cancer Biol.* **8**(5): 345-357.

Feng Q, Cao R, Xia L, Erdjument-Bromage H, Tempst P, Zhang Y (2002) Identification and Functional Characterization of the p66/p68 Components of the MeCP1 Complex. *Mol Cell Biol.* **22**(2): 536–546

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature.* **391**(6669): 806-11.

Floyd-Smith, G., Slattery, E., Lengyel, P. (1981) Interferon action: RNA cleavage pattern of a (2'-5') oligoadenylate-dependent endonuclease. *Science* **212**: 1030-1032.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. and Postlethwait, J. (1999) The preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531–1545.

Fraser, A. G., Kamath, R. S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M., Ahringer, J. (2000) Functional genomic analysis of C. elegans chromosome I by systematic RNA interference. *Nature* **408**(6810): 325-330.

Furuya, K. N., Bradley, G., Sun, D., Schuetz, E. G. and Schuetz, J. D. (1997) Identification of a new P-glycoprotein-like ATP-binding cassette transporter gene that is overexpressed during hepatocarcinogenesis. *Cancer Res.* **57**: 3708-3716.

Gabaldon, T., Huynen, M. A. (2004) Prediction of protein function and pathways in the genome era. *Cell Mol. Life Sci. 61*(7-8) 930-44.

Giaever G, *et al.* (2002) Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418**(6896):387-91.

Gietz RD, Schiestl RH, Willems AR and Woods RA. (1995) Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast.* **11**(4):355-60.

Gilleard, J. S., Shafi, Y., Barry, J. D., McGhee, J. D. (1999) elt-3: a C. elegans GATA factor expressed in the embryonic epidermis during morphogenesis. *Developmental Biology* **208**: 265-280.

Gissendanner, C. R., Crossgrove, K., Kraus, K. A., Maina, C. V., Sluder, A. E. (2004) Expression and function of conserved nuclear receptor genes in C. elegans. *Dev. Biol.* **266**(2) 399-416.

Goldsmith, M. E., Gudas, J. M., Schneider, E., Cowan, K. H. (1995) Wild type p53 stimulates expression from the human multidrug resistance promoter in a p53-negative cell line. *J Biol Chem.* **27**;270(4):1894-8.

Gronemeyer, H & Laudet, V (1995) Transcription factors 3: nuclear receptors. *Protein Profile.* **2**: 1173– 1308.

Guhathakurta, D., Schriefer, L. A., Hresko, M. C., Waterston, R. H., Stormo, G. D. (2002) Identifying muscle regulatory elements and genes in the nematode C. elegans. *Pac Symp Biocomput.* 425-436.

Harrison, P. M., Echols, N. and Gerstein, M. B. (2001) Digging for dead genes: An analysis of the characteristics of the pseudogene population in the Caenorhabditis elegans genome. *Nucleic Acids Res earch* **29**: 818-830.

Hassel, B. A., Zhou, A., Sotomayor, C., Maran, A., Silverman, R. H. (1993) A dominant negative mutant of 2-5A-dependent RNase suppresses antiproliferative and antiviral effects of interferon. *EMBO J.* **12** : 3297-3304.

Higgins, C. F. and Sharp, P. M. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73**: 237-244.

Higgins, C.F. (1995) The ABC of channel regulation. *Cell.* **8**;82(5):693-6.

Higgins, C.F. 1992. ABC transporters: from microorganisms to man. *Annu. Rev. Cell Biol.* **8**: 67-113.

Hobert, O. (2002 Apr) PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic C. elegans. *Biotechniques* **32**(4):728-30.

Hu, Z., Jin, S., Scotto, K. W. (2000) Transcriptional activation of the MDR1 gene by UV irradiation. Role of NF-Y and Sp1. *J Biol Chem.* **275**(4): 2979-85.

Hyde, S. C., Emsley, P., Hartshorn, M. J., Mimmack, M. M., Gileadi, U., Pearce, S. R., Gallagher, M. P., Gill, D. R., Hubbard, R. E. and Higgins, C. F. (1990) Structural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport. *Nature* **346**: 362-365.

Ikeguchi, M., Teeter, L. D., Eckersberg, T., Ganapathi, R., Kuo, M. T. (1991) Structural and functional analyses of the promoter of the murine multidrug resistance gene mdr3/mdr1a reveal a negative element containing the AP-1 binding site. *DNA Cell Biol.* **10**(9): 639-49.

Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M and Sakaki Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome.*Proc Natl Acad Sci* U S A. **98**(8):4569-74.

Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. and Gibson, T. J. (1988) Multiple sequence alignment with Clustal X. *Trends in Biochemical Sciences* **23**: 403-405.

Jedlitschky, G., Burchell, B. and Keppler, D. (2000) The multidrug resistance protein 5 functions as an ATP-dependent export pump for cyclic nucleotides. *J. Biol. Chem.* **275**: 30069-30074.

Johnson, R. A., Ince, T. A. and Scotto, K. W. (2001) Transcriptional repression by p53 through direct binding to a novel DNA element. *J Biol Chem.* **276**(29): 27716-20.

Jones, S. J., Baillie, D. L.(1995) Characterization of the let-653 gene in C. elegans. *Mol. Gen. Genet.* **248**: 719–726.

Juliano, R. L. and Ling, V. (1976) A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. *Biochem. Biophys. Acta.* **455**: 152-162.

Kamath, R. S., Fraser, A. G., Dong, Y., Poplin, G., Durbin, R., M., Kingpin, A., Le Boot, N., Moreno, S., Shoran, M., Welshman, D. P., Zipper, L. P., Harbinger, J. (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi, *Nature* **421**(6920): 231-7.

Kamath RS, Martinez-Campos M, Zipperlen P, Fraser AG, Ahringer J. (2001) Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans*. *Genome Biol.***2**(1):RESEARCH0002.

Kaminski, W. E., Orso, E., Diederich, W., Klucken, J., Drobnik, W. and Schmitz, G. (2000) Identification of a novel human sterol-sensitive ATP-binding cassette transporter (ABCA7). *Biochem. Biophys. Res. Commn.* **273**: 532-538.

Kaminski, W. E., Piehler, A., Pullman, K., Porsch-Ozcurumez, M., Duong, C., Bared, G. M., Buchler, C. and Schmitz, G. (2001) Complete coding sequence, promoter region, and genomic structure of the human ABCA2 gene and evidence for sterol-dependent regulation in macrophages. *Biochemical and Biophysical Research Commnications* **281**: 249-258.

Kathleen. W., Scotto, K. W. (2003) Transcriptional regulation of ABC drug transporters. *Oncogene*. **22**: 7496-7511.

Kelly, W. G., Xu, S. Q., Montgomery, M. K. and Fire, A. (1997) Distinct requirements for somatic and germline expression of a generally expressed Caernorhabditis elegans gene. *Genetics* **146**: 227-238.

Kennedy, B.P, Aamodt E,J, Allen, F.L, Chung, M.A, Heschl, M,F, McGhee J.D. (1993). The gut esterase gene (ges-1) from the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae. J. Mol. Biol.* **229**:890-908.

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

Kerr, I. D. (2004) Sequence analysis of twin ATP binding cassette proteins involved in translational control, antibiotic resistance, and ribonuclease L inhibition. *Biochem. Biophys. Res. Commun.* **27**: 315(1) 166-73.

Kerr, I. M., Brown, R. E. (1978) pppA2'p5'A2'p5'A: an inhibitor of protein synthesis synthesized with an enzyme fraction from interferon-treated cells. *Proc. Natl. Acad. Sci.* USA. **75** : 256-260.

Kim, S. H., Hu, W. Y., Kang, C. D., Lim, Y. S., Kim, D. W., Chung, B. S. (1997) Involvement of heat shock factor in regulating transcriptional activation of MDR1 gene in multidrug-resistant cells. *Cancer Lett.* **115**(1): 9-14.

Kirouac, M., Sternberg, P. W. (2003) *cis*-Regulatory control of three cell fate specific genes in vulval organogenesis of Caenorhabditis elegans and C. briggsae. *Dev Biol* **257**: 85–103.

Kramer, J.M, Johnson, J.J. (1993) Analysis of mutations in the sqt-1 and rol-6 collagen genes of Caenorhabditis elegans. *Genetics*. **135**(4):1035-45

Labialle, S., Gayet, L., Marthinet, E., Rigal, D. and Baggetto, L. G. (2002) Transcriptional regulators of the human multidrug resistance 1 gene: recent views. *Biochem Pharmacol*. **64**(5-6): 943-8.

Lamesch, P., Milstein, S., Hao, T., Rosenberg, J., Li, N., Sequerra, R., Bozak, S., Doucette-Stamm, L., Vandenhaute, J., Hill, D & Vidal, M. (2004) *C. elegans* ORFeome version 3.1: increasing the coverage of ORFeome resources with improved gene repredictions. *Genome Res*. **14**(10B):2064-9.

Le Roy, F., Laskowska, A., Silhol, M., Salehzada, T. and Bisbal, C. (2000) Characterization of RNABP, an RNA binding protein that associates with RNase L. *J. Interferon Cytokine Res*. **20**: 635-644.

Lercher, M. J., Blumenthal, T. and Hurst, L. D. (2003) Coexpression of neighboring genes in *Caenorhabditis elegans* is mostly due to operons and duplicate genes. *Genome Research* **13**: 238-43.

Li, S. et al. (2004) A map of the interactome network of the metazoan *C. elegans*, *Science*. **303**(5657) 540-3.

Lincke, C.R., Broeks, A., The, I., Plasterk, R. H. A. and Borst, P. (1993) The expression of two P-glycoprotein (pgp) genes in transgenic *Caenorhabditis elegans* is confined to intestinal cells. *EMBO J*. **12**: 1615-1620.

Maglich, J. M., Sluder, A., Guan, X., Shi, Y., McKee, D. D., Carrick, K., Kamdar, K., Willson, T. M., Moore, J. T. (2001) Comparison of complete nuclear receptor sets from the human, *C. elegans* and Drosophila genomes. *Genome Biol*. **2**, research0029.1–0029.7

Marton, M.J., Vazquez de Aldana, C.R., Qiu, H., Chakraburtty, K., Hinnebusch, A.G. (1997) Evidence that GCN1 and GCN20, translationalregulators of GCN4, function on elongating ribosomes in activation of eIF2alpha kinase GCN2, *Mol. Cell. Biol*. **17**: 4474 - 4489.

McKay, S. J., Johnsen, R., Khattra, J., Asano, J., Baillie, D. L., Chan, S., Dube, N., Fang, L., Goszczynski, B., Ha, E et al.,. (2003) Gene Expression Profiling of Cells, Tissues and Developmental Stages of the Nematode *C. elegans*. Cold Spring Harbor Symposium LXVIII: The Genome of *Homo Sapiens*, Cold Spring Harbor Press, Cold Spring Harbor, New York.

Miller, D. M. 3rd, Desai, N. S., Hardin, D. C., Piston, D. W., Patterson, G. H., Fleenor, J., Xu, S., Fire, A. (1999) Two-color GFP expression system for *C. elegans*. *Biotechniques* **26**(5): 914-918, 920-921.

Mitsuhashi, N., Miki, T., Senbongi, H., Yokoi, N., Yano, H., Miyazaki, M., Nakajima, N., Iwanaga, T., Yokoyama, Y., Shibata, T. and Seino, S. (2000) MTABC3, a novel mitochondrial ATP-binding cassette protein involved in iron homeostasis. *J. Biol. Chem.* **275**: 17536-17540.

Mounsey, A., Petra Bauer and Hope, I. A. (2002) Evidence suggesting that a fifth of annotated *Caenorhabditis elegans* genes may be pseudogenes. *Genome Research* **12**: 770-775.

Mulder, N. J., et al (2003) The InterPro Database, 2003 brings increased coverage and new features. *Nucl. Acids. Res.* **31**: 315-318.

Muller, C., Leutz, A. (2001) Chromatin remodeling in development and differentiation. *Curr. Opin. Genet. Dev.* **11**: 167–174.

Nelson, F. K., Riddle, D. L. (1984) Functional study of the *C. elegans* secretory-excretory system using laser microsurgery. *J. Exp. Zool.* **231**(1): 45-56.

Nguyen, K. T., Liu, B., Ueda, K., Gottesman, M. M., Pastan, I., Chin, K. V. (1994) Transactivation of the human multidrug resistance (MDR1) gene promoter by p53 mutants. *Oncol Res.* **6**(2):71-7.

Nicholas, K.B., Nicholas H.B. Jr., and Deerfield, D.W. II. (1997) GeneDoc: Analysis and Visualization of Genetic Variation, EMBNEW.NEWS 4:14

Ohno, S. (1999) Gene duplication and the uniqueness of vertebrate genomes circa 1970-1999. *Semin Cell Dev Biol.* **10**(5):517-22.

Orphanides, G., Lagrange, T., Reinberg, D. (1996) The general transcription factors of RNA polymerase II. *Genes Dev.* **10**: 2657–2683.

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

Quandt, K., Frech, K., Karas, H., Wingender, E., Werner, T. (1995) MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res.* **23**: 4878-4884.

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

Riordan, J. R., Deuchars, K., Kartner, N., Alon, N., Trent, J., Ling, V. (1985) Amplification of P-glycoprotein genes in multidrug-resistant mammalian cell lines. *Nature* **316**: 817-819, 1985.

Riordan, J. R., Rommens, J. M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J. L., Drumm, M. L., Iannuzzi, M. C., Collins, F. S., Tsui, L.-C. (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **245**: 1066-1073.

Robertson, H. M. (2000) The large srh family of chemoreceptor genes in *Caenorhabditis nematodes* reveals processes of genome evolution involving large duplications and deletions and intron gains and losses. *Genome Research* **10**: 192-203.

Rozen S and Skaletsky H. (2000) Primer3 on the WWW for general users and for biologist programmers.*Methods Mol Biol.* **132**:365-86.

Rubin, G. M., et al. (2000) Comparative genomics of the eukaryotes. *Science* **287**: 2204-2215.

Runswick, M.J., Walker, J.E.(1983) The amino acid sequence of the beta-subunit of ATP synthase from bovine heart mitochondria. *J Biol Chem.* **10**;258(5):3081-9.

Saitou, N. andNei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406-425.

Sanchez-Fernandez, R., Davies, T. G., Coleman, J. O., Rea, P. A. (2001) The Arabidopsis thaliana ABC Protein Superfamily, a Complete Inventory. *J. Bio. Chem.* **276**: 30231-44.

Saurin, W., Hofnung, M., Dassa, E. (1999) Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. *J. Mol. Evol.*, **48**: 22-41.

Savary, S., Allikmets, R., Denizot, F., Luciani, M.-F., Mattei, M.-G., Dean, M., Chimini, G. (1997) Isolation and chromosomal mapping of a novel ATP-binding cassette transporter conserved in mouse and human. *Genomics.* **41**: 275-278.

Schwartz, S., Zhang, Z., Frazer, K. A., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R. and Miller, W. (2000) PipMaker-A Web Server for Aligning Two Genomic DNA Sequences. *Genome Research* **10**: 577-586.

Scotto, K. W. (2003) Transcriptional regulation of ABC drug transporters. *Oncogene* **20**; 22(47): 7496-511.

Shen, D. W., Fojo, A., Chin, J. E., Roninson, I. B., Richert, N., Pastan, I. and Gottesman, M. M. (1986) Human multidrug-resistant cell lines: increased mdr1 expression can precede gene amplification. *Science* **232**: 643–645.

Sheps, J. A., Ralph, S., Zhao, Z., Baillie, D. L., Ling, V. (2004) The ABC transporter gene family of *C. elegans* has implications for the evolutionary dynamics of multidrug resistance in eukaryotes. *Genome Biol.* **5** R15.

Simmer, F., Moorman, C., Van Der Linden, A. M., Kuijk, E., Van Den Berghe, P. V., Kamath, R., Fraser, A. G., Ahringer, J., Plasterk, R. H. (2003) Genome-Wide RNAi of *C. elegans* Using the Hypersensitive rrf-3 Strain Reveals Novel Gene Functions. *PLoS Biol.* **1**(1) E12. Epub.

Sluder, A., Mathews, S. W., Hough, D., Yin, V. P., Maina, C. V. (1999) The nuclear receptor superfamily has undergone extensive proliferation and diversification in nematodes. *Genome Res.* **9**: 103– 120.

Sonnhammer, E. L., Durbin, R. (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **167**(1-2): GC1-10.

Spieth, J., MacMorris, M., Broverman, S., Greenspoon, S., Blumenthal, T. (1988) Regulated expression of a vitellogenin fusion gene in transgenic nematodes. *Dev Biol.* **130**(1): 285-293.

Stein, L. D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M. R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., et al., . (2003) The Genome Sequence of *Caenorhabditis briggsae*: A Platform for Comparative Genomics. *PLoS Bi*ol. **1**(2):E45.

Stringham, E., Pujol, N., Vandekerckhove, J., Bogaert, T. (2002) unc-53 controls longitudinal migration in *C. elegans*. *Development* **129**(14): 3367-3379.

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

Sundseth, R., MacDonald, G., Ting, J., King, A. C. (1997) DNA elements recognizing NF-Y and Sp1 regulate the human multidrug-resistance gene promoter. *Mol Pharmacol.* **51**(6): 963-71.

Tachikui, H., Saitou, N., Nakajima, T., Hayasaka, I., Ishida, T., Inoue, I. (2003) Lineage-specific homogenization of the polyubiquitin gene among human and great apes. *J Mol Evol.* **57**(6): 737-44.

Taglicht, D and Michaelis,S. (1998) *Saccharomyces cerevisiae* ABC proteins and their relevance to human health and disease. *Methods Enzymol.***292**:130-62.

Tamames, J., G. Casari, C. Ouzonis, and A. Valencia. 1997. Conserved clusters of functionally related genes in two bacterial genomes. *Journal of Molecular Evolution* **44**: 66-73.

Teeter, L. D., Petersen, D. D., Nebert, D. W., Kuo, M. T. (1991) Murine mdr-1, mdr-2, and mdr-3 gene expression: no coinduction with the Cyp1a-1 and Nmo-1 genes in liver by 2,3,7,8-tetrachlorodibenzo-p-dioxin. *DNA Cell Biol.* **10**(6): 433-41.

The C. elegans Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science.* **282**(5396): 2012-8.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**: 4876-4882.

Thottassery, J. V., Zambetti, G. P., Arimori, K., Schuetz, E. G., Schuetz, J. D. (1997) p53-dependent regulation of MDR1 gene expression causes selective resistance to chemotherapeutic agents. *Proc Natl Acad Sci* **94**(20): 11037-42.

Tyzack, J. K., Wang, X., Belsham, G. J., Proud, C. G. (2000) ABC50 interacts with eukaryotic initiation factor 2 and associates with the ribosome in an ATP-dependent manner. *J. Biol. Chem.* **275**(44) 34131-9.

Ueda, K.; Pastan, I.; Gottesman, M. M. (1986) Isolation and sequence of the promoter region of the human multidrug-resistance (P-glycoprotein) gene. *J. Biol. Chem.* **262**: 17432-17436.

Van Helvoort, A., Smith, A. J., Sprong, H., Fritzsche, I., Schinkel, A. H., Borst, P. and Van Meer, G. (1996) MDR1 P-glycoprotein is a lipid translocase of broad specificity, while MDR3 P-glycoprotein specifically translocates phosphatidylcholine. *Cell* **87**: 507–517.

van Noort, V., Snel, B., Huynen, M. A. (2003) Predicting gene function by conserved co-expression. *Trends Genet.* **19**: 238–242.

Wade, P. A., Gegonne, A., Jones, P. L., Ballestar, E., Aubry, F., Wolffe, A. P. (1999) Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. *Nat. Genet.* **23**: 62-66.

Wang X and Chamberlin H.M. (2002) Multiple regulatory changes contribute to the evolution of the *Caenorhabditis elegans* lin-48 ovo gene. *Genes & Dev.***16**(18):2345-9.

Wang, X., Chamberlin, H. M. (2004) Evolutionary innovation of the excretory system in *Caenorhabditis elegans*. *Nat Genet.* **36**(3): 231-232.

Watterson, G. A. (1983) On the time for gene silencing at duplicate loci. *Genetics* **105**: 745–766.

Webb, C. T., Shabalina, S. A., Ogurtsov, A. Y. and Kondrashov, A. S. (2002) Analysis of similarity within 142 pairs of orthologous intergenic regions of *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *Nucleic Acids Research* **130**: 1233-1239.

Wijnholds, J., Mol, C. A., van Deemter, L., de Haas, M., Scheffer, G. L., Baas, F., Beijnen, J. H., Scheper, R. J., Hatse, S., De Clercq, E., Balzarini, J. and Borst, P. (2000) Multidrug-resistance protein 5 is a multispecific organic anion transporter able to transport nucleotide analogs. *Proc. Natl. Acad. Sci.* USA **97**: 7476-7481.

Wingender, E., Chen, X., Fricke, E., Geffers, R., Hehl, R., Liebich, I., Krull, M., Matys, V., Michael, H., Ohnhäuser, R., Prüß, M., Schacherer, F., Thiele, S., Urbach, S. (2001) The TRANSFAC system on gene expression regulation. *Nucleic Acids Res.* **29**: 281-283.

Wolf, E., Kim, P. S. and Berger, B. (1997) Multicoil: A program for predicting two-three-stranded coils. *Protein Sci.* **6**:1179-1189.

Wu, Y. and Horvitz, H. R. (1998) The *C. elegans* cell corpse engulfment gene ced-7 encodes a protein similar to ABC transporters. *Cell* **93**: 951-960.

Yang, J. M., Vassil, A. D., Hait, W. N. (2001) Activation of phospholipase C induces the expression of the multidrug resistance (MDR1) gene through the Raf-MAPK pathway. *Mol Pharmacol.* **60**(4): 674-80.

Yeh, L. C., Lee, J. C. (1998) Yeast ribosomal proteins L4, L17, L20, and L25 exhibit different binding characteristics for the yeast 35S precursor rRNA. *Biochem Biophys Acta.* **1443**(1-2):139-48.

Zhao Z, Fang LL, Johnsen R, Baillie DL. (2004a) ATP-binding cassette protein E is involved in gene transcription and translation in *Caenorhabditis elegans*. *Biochem Biophys Res Commun.* **323**(1):104-11.

Zhao, Z, Sheps, J.A, Ling, V, Fang, L.L, Baillie, D.L. (2004b) Expression analysis of ABC transporters reveals differential functions of tandemly duplicated genes in *Caenorhabditis elegans*. *J Mol Biol.***344**(2):409-17.