doi:10.1016/j.jmb.2004.09.052 J. Mol. Biol. (2004) 344, 409–417

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Expression Analysis of ABC Transporters Reveals Differential Functions of Tandemly Duplicated Genes in *Caenorhabditis elegans*

Zhongying Zhao[1], Jonathan A. Sheps[2]*, Victor Ling[2], Lily L. Fang[1] and David L. Baillie[1]

[1]*Department of Molecular Biology and Biochemistry Simon Fraser University 8888 University Drive, Burnaby BC, Canada V5A 1S6*

[2]*BC Cancer Research Centre BC Cancer Agency, 601 West 10th Avenue, Vancouver, BC Canada V5Z 1L6*

**Corresponding author*

We have previously identified 60 predicted ABC transporter genes in the *Caenorhabditis elegans* genome and classified them into eight groups. As an initial step towards understanding how these putative ABC genes work in worms, we generated promoter–fluorescent protein fusions for the entire family to address when and where these genes are turned on *in vivo*. Both *Aequoria* green fluorescent protein (GFP) and *Discosoma* red fluorescent protein (RFP) were used as reporters in our transgenic assay. Observable expression is more frequently seen from fusions to genes in subfamilies B, C, D and E than those in subfamilies A and G. Sixteen worm ABC genes are found in tandem duplications, forming two four-gene clusters and four two-gene clusters. Fifteen out of the 16 duplicated gene promoters drove different or partially overlapping expression patterns, suggesting active functions for these duplicated genes. Furthermore, our results suggest that an internal promoter can cause differential expression of genes within an operon. Finally, our observations suggest that it is possible for coding sequences to function as a regulatory region for a neighbouring gene.

Keywords: ABC transporter; *Caenorhabditis elegans*; promoter; green fluorescence protein (GFP); expression

Introduction

ATP-binding cassette (ABC) transporters constitute one of the largest protein families in both prokaryotes and eukaryotes. These proteins bind ATP and use its energy to drive the transport of various molecules across the plasma membrane or the intracellular membranes of organelles, such as the endoplasmic reticulum, peroxisomes, and mitochondria.[1,2] Mutation of ABC transporters can result in various diseases in humans or in hypersensitivity to drugs. For example, mutations in the *ABCA1* gene cause very low levels of HDL in plasma and a build-up of cholesterol in macrophages.[3–5] Loss of ABCC7 function is responsible for cystic fibrosis.[6] Some ABC proteins, such as MDR1 (ABCB1), MRP1 (ABCC1) and BCRP (ABCG2), are commonly responsible for the development of drug-resistant tumours,[7] a substantial challenge for chemotherapy.

A functional ABC transporter generally consists of at least one evolutionarily conserved ABC domain, also known as a nucleotide-binding domain (NBD), which are about 200 amino acid residues long, and several α-helical transmembrane domains (TMD). The NBD contains three conserved features: Walker A and B motifs which can be found in many ATP-binding proteins, and a signature (C) motif, located just upstream of the Walker B site.[8] The C motif is diagnostic of ABC transporters and distinguishes them from other Walker ATPases. Many ABC transporters are organized as either full transporters containing two TMD and two NBD or as half transporters containing one of each domain,[8] or occasionally only as solitary NBDs. The half transporters work as either homodimers or heterodimers. Phylogenetic analysis has divided the gene family into eight different subfamilies, named ABCA through H, most of which have well-characterized human homologues.[9] Sixty ABC

Abbreviations used: ABC, ATP-binding cassette; NBD, nucleotide-binding domain; TMD, transmembrane domains; SAGE, serial analysis of gene expression; RNAi, RNA interference; RFP, red fluorescent protein; GFP, green fluorescent protein.

E-mail address of the corresponding author: jsheps@bccancer.bc.ca

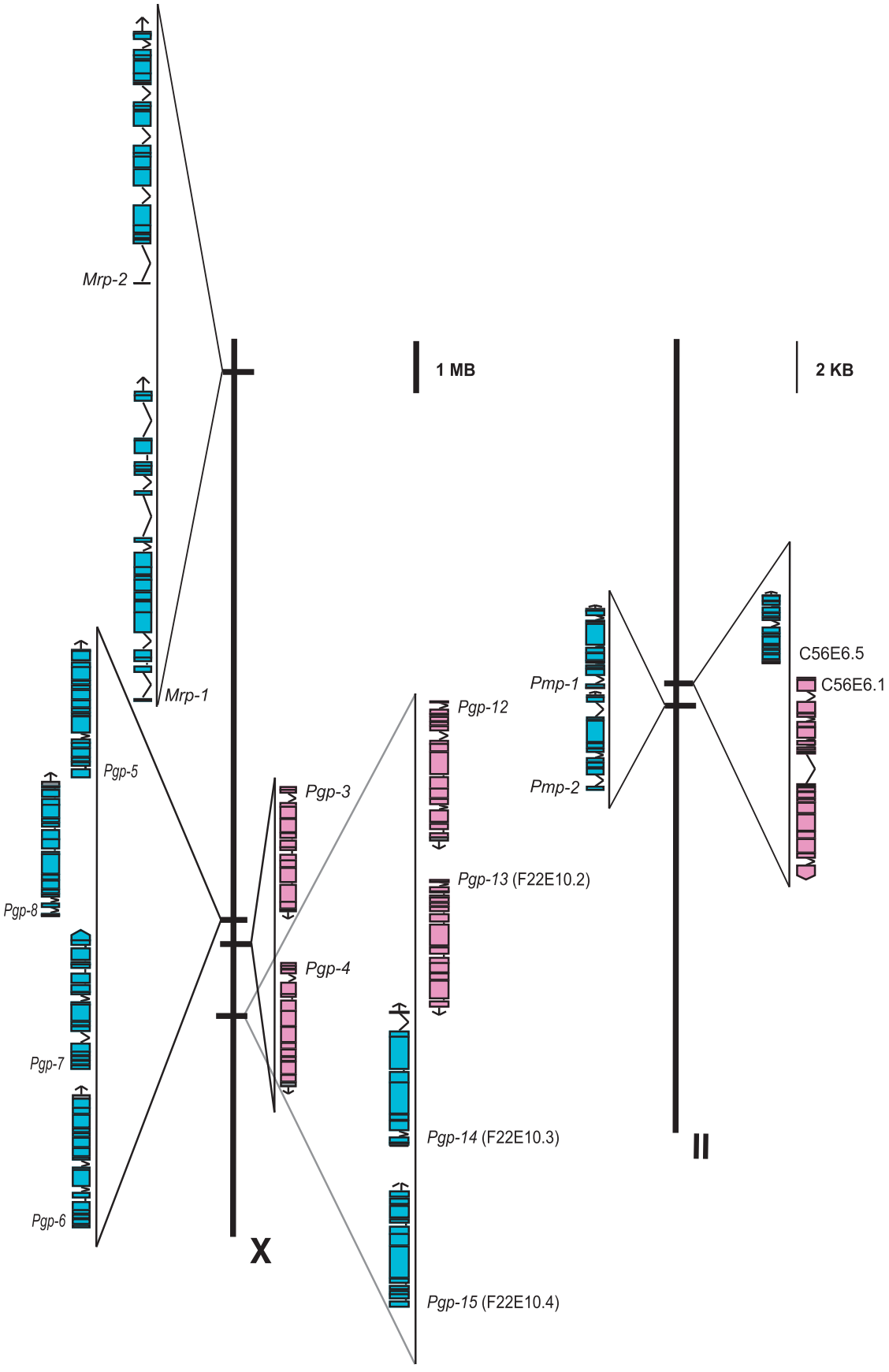

Figure 1. Schematic representation of 16 tandemly duplicated ABC genes on chromosomes II and X of the *C. elegans* genome. These form six clusters, which are illustrated using images drawn from Wormbase release WS130 (http://wormbase.org/). The two ABCH subfamily genes identified with cosmid clone C56E6 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 ABCD subfamily.

transporters have been identified in *Caenorhabditis elegans* with members grouped into each of the eight established subfamilies. Several ABC transporters have been functionally characterized in *C. elegans*. The expression of two P-glycoprotein (*Pgp*) genes (subfamily ABCB) in transgenic *C. elegans* is confined to intestinal cells.[10] *Pgp-1* (K08E7.9) and *Pgp-3* (ZK455.7) are responsible for resistance to chloroquine and colchicines.[11] ABCC subfamily member *Mrp-1* (F57C12.5) and *Pgp-1*are both involved in heavy metal resistance.[12] The ABCA subfamily gene *Ced-7* (C48B4.4) encodes a protein involved in cell corpse engulfment.[13]

Much of the genomes of flies and worms consist of duplicated genes.[14] It is becoming common to find computational investigations into the evolution of genes or gene families with the increasing number of genome sequences available.[15–17] 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. 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. Our 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 in which to address these questions because there are 16 tandemly positioned ABC genes in

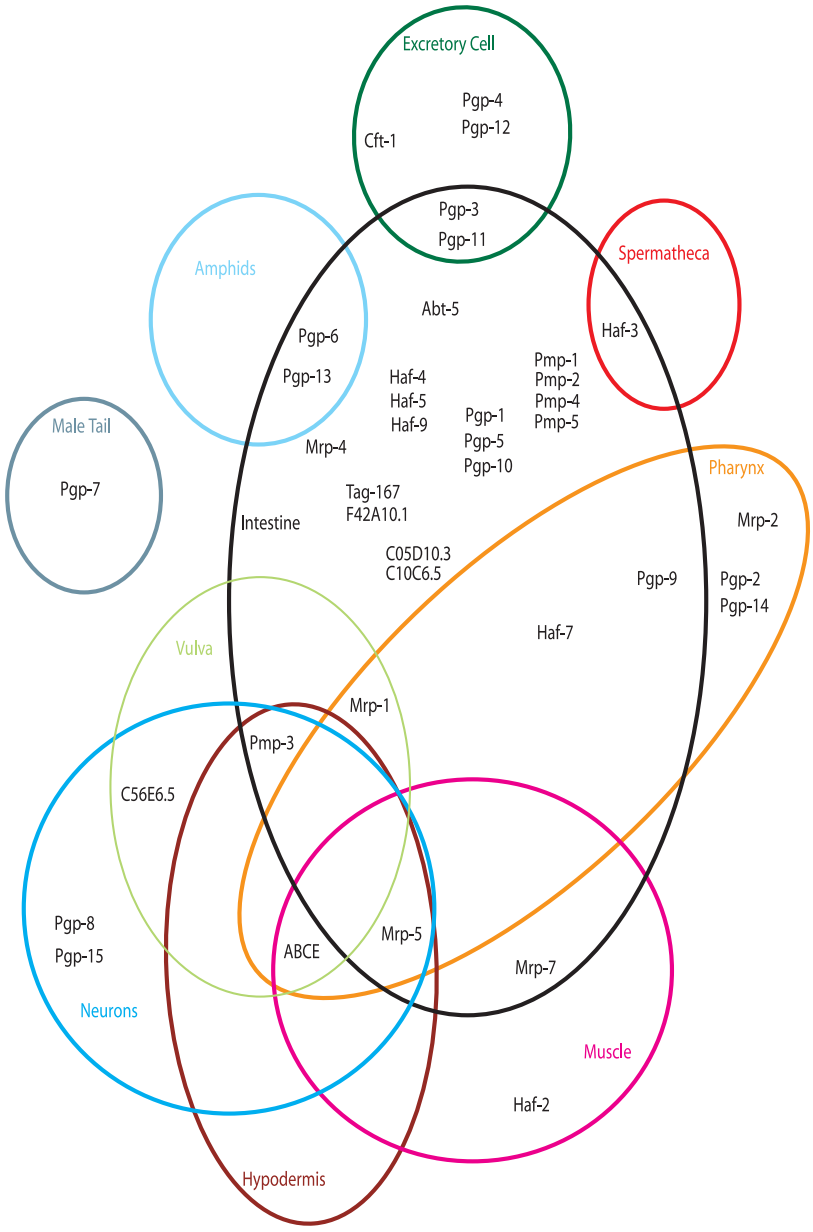

Figure 2. Venn diagram showing the tissue distribution of expression of 39 ABC transporters in *C. elegans* mostly deduced from the expression of promoter-GFP (or RFP) fusion proteins. Not all genes shown are expressed in all, or the same, developmental stages. See online Supplementary data for full details on the timing and relative strengths of expression observed.

two-gene or four-gene clusters on different chromosomes (Figure 1). The presence of tandem duplications is not unique to the worm ABC genes. For instance, there are five-gene ABCA clusters in both human and mouse.[18] As a preliminary investigation into the roles of ABC transporter genes in *C. elegans* we have performed transgenic expression analysis of all ABC transporters in *C. elegans*, using promoter-driven GFP or red fluorescent protein (RFP) reporters, except for *Pgp-1*, *Mrp-5* (F14F4.3) and *Mrp-1*, for which tissue expression patterns have already been determined.[11,19] Our 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.

Results

Many of the genes in *C. elegans* are found in pairs of adjacent sister genes. In order to have a general idea of how these, tandem duplications of homologous genes, function *in vivo* we have tried to address when and where the worm ABC transporter genes are turned on *in vivo.* To this end we have used GFP or RFP markers driven by gene-specific promoters. All the promoters were derived from genomic DNA. 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). We have generated GFP transgenic strains for 57 out of the 60 ABC genes in *C. elegans*. GFP constructs successfully produced visible expression patterns in 36 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 faithfully reproduce the expression patterns seen by GFP assay, indicating that the choice of reporter itself does not make a difference in promoter-driven expression pattern. For the majority of genes their expression patterns are first reported here. The average intergenic space upstream of *C. elegans* ABC transporter genes is 3118 bp and the average size of our putative promoter construct was 1575 bp. The largest promoter sequence used was 3422 bp, for *Pgp-12* (F22E10.1); the smallest promoter construct was only 250 bp, for Y49E10.9 (full details available as an online Supplementary data Table). Up to five attempts to obtain successful expression were made for each ABC transporter gene. Transgene expression was most frequently seen in the intestine, pharynx and excretory cell (Figure 2, full details in the online Supplementary data Table).

Tandemly duplicated ABC genes tend to be differentially expressed

Sixteen worm ABC genes are tandemly arranged, forming two four-gene clusters and four two-gene clusters (Figure 1). Twelve of 16 ABCs on the X chromosome are present in tandem. Fifteen out of these 16 ABC genes gave observable expression in our promoter driven GFP/RFP assay (Figure 2 and online Supplementary data Table). 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 (Table 1), except for C56E6.1, which is arranged in a head-to-head orientation with C56E6.5, and yielded no observable expression. Among the four ABC genes clustered on cosmid clone F22E10 (having 65–75% amino acid sequence identity) all are expressed in different tissues (Figure 3; Table 1). *Pgp-12* 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. A similar situation prevails in the four-gene cluster,

Table 1. Expression patterns and similarity between tandem duplicate ABC genes

Cosmid ID	Gene name	Expression pattern	% Similarity
C56E6.1	NA	No observable expression	19
C56E6.5	NA	Larval and adult head and tail neurons, vulva	
C44B7.8	*Pmp-1*	Strong larval and weak adult intestine	70
C44B7.9	*Pmp-2*	Strong larval and adult intestine	
F22E10.1	*Pgp-12*	Larval and adult excretory cell, embryo	68
F22E10.2	*Pgp-13*	Adult posterior intestine, amphids	
F22E10.3	*Pgp-14*	Larval and adult pharynx	
F22E10.4	*Pgp-15*	Adult head and tail neurons, embryo	
T21E8.1	*Pgp-6*	Larval and adult intestine, adult amphids	63
T21E8.2	*Pgp-7*	Adult male tail rays	
T21E8.3	*Pgp-8*	Adult head neurons	
C05A9.1	*Pgp-5*	Larval and adult anterior intestine, embryo	
F42E11.1	*Pgp-4*	Larval excretory cell	78
ZK455.7	*Pgp-3*	Larval and adult excretory cell, intestine	
F57C12.4	*Mrp-2*	Adult pharynx	74
F57C12.5	*Mrp-1*	Adult pharynx, intestine, vulva[19]	

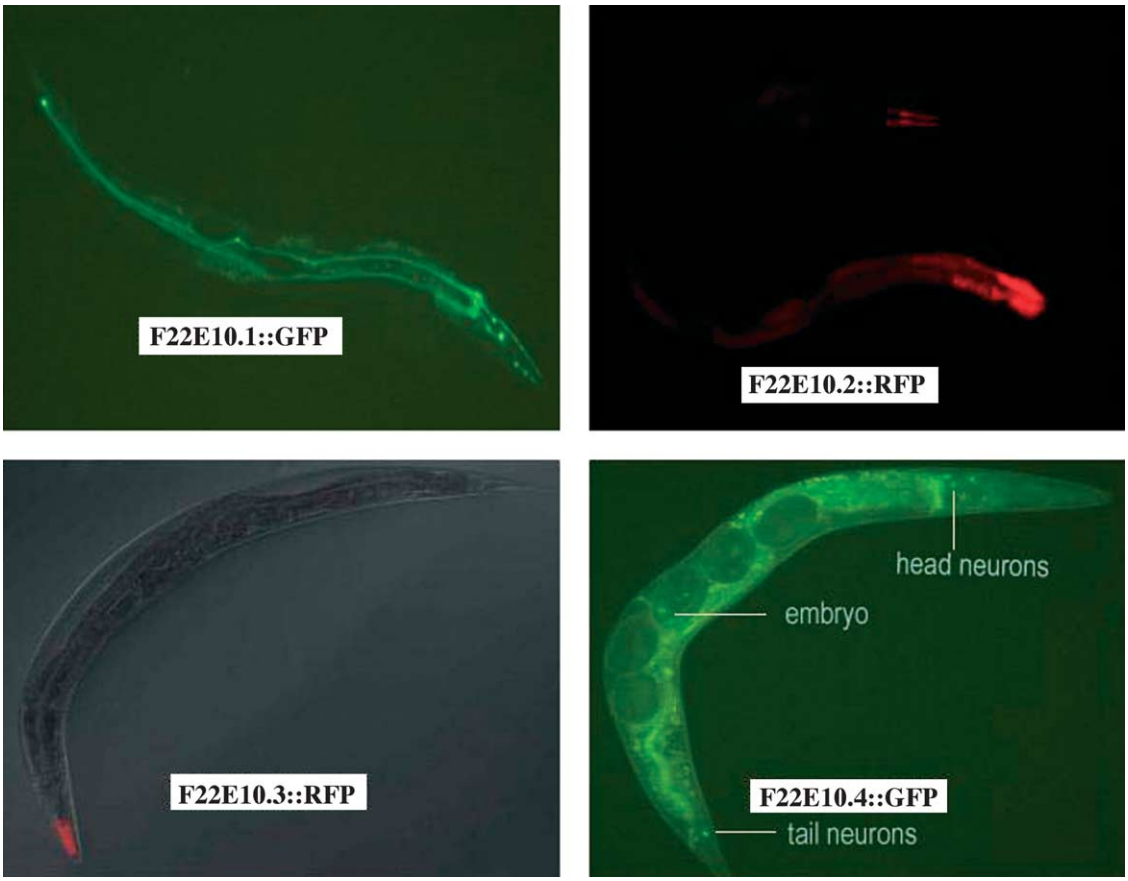

Figure 3. Photomicrographs of GFP or RFP expression driven by promoters derived from the four ABC genes in cluster 6. Promoter–GFP fusion constructs were built as described.[18] *Pgp-12* (F22E10.1) showed strong expression in excretory cell in all stages (only adult shown); *Pgp-13* (F22E10.2) yielded 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.

which includes C05A9.1, T21E8.1, T21E8.2 and T21E8.3 (*Pgp-5*, *-6*, *-7* and *-8*) which all show expression in different tissues (Figure 2). The expression patterns of *Mrp-2* (F57C12.4) and *Mrp-1* overlap, while stage specific expression is observed for *Pmp-1* (C44B7.8) and *Pmp-2* (C44B7.9) though in the same tissues (Figure 2 and online Supplementary Table). Most of the duplicated ABCs are members of the ABCB subfamily (10 out of 16), including all ABCs within four-gene clusters (*Pgp-12* to *-15*, and *Pgp-5* to *-8*) and two of the duplicated ABCs transcribed in the same orientation, *Pgp-3* (ZK455.7) and *Pgp-4* (F42E11.1). Our observed expression patterns, in combination with EST data in GenBank†, suggest that none of these tandemly duplicated ABC genes are pseudo-genes. In other words, they are expected to be functional *in vivo*. ABCB genes found in tandem duplications are most often expressed in the worm's pharynx, gut or excretory cell. Some members of the ABC subfamily are known to be involved in drug resistance.[11,12] These duplicated genes might provide the worm fuller protection against xenobiotics, as is the case with their homologues in other organisms. Few of the duplicated worm ABC genes have a visible RNAi phenotype,[20] 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 function has yet to be determined.

Coding sequences of neighbouring genes contain regulatory elements

It is common in *C. elegans* for intergenic spaces to be 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 (D. Moerman, personal communication) and have been amplified by the ORFeome project,[21] but failed to give observable GFP expression when intergenic sequences alone were used as putative promoters.

† http://www.ncbi.nlm.nih.gov/

We 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 we used for GFP fusions were initially 858 bp and 1519 bp for F42A10.1 and *Pmp-3*, respectively. We then included 2928 bp and 2914 bp, respectively, of intergenic sequences, including partial (opposite strand) coding sequences from the neighbouring genes, as our putative promoters for F42A10.1 and *Pmp-3* (online Supplementary Figure 1). Both extended promoters were found to effectively drive strong RFP expression in either the larval or adult stages. This suggests that transcriptional elements in *C. elegans* can be found buried within coding sequences.

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

We have generated transgenic strains for 57 worm ABCs. A similar expression assay for the other three *C. elegans* ABC transporters *Pgp-1*, *Mrp-5* and *Mrp-1* has been done.[11,19] 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 eight 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 tandem duplications within the ABCA subfamily whereas ten 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.[19] 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 of nine ABCG genes can be described as recently duplicated (data not shown). This observation reinforces the impression, already explored in our previous paper,[9] that the evolutionary dynamics and functional inter-relationships of the ABC transporter family are rather atypical of multi-gene families in general.

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 we refer to as a head-to-head orientation (Table 2). For these 30 genes, the average intergenic size is 4063 bp; and the average segment used as a promoter in our expression constructs was 1579 bp. 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, and their average intergenic size is 2173 bp, and the average promoter construct derived from them 1572 bp, and only five of these genes failed to yield observable expression. It is not clear why the layout of the source genes in the genome have such an influence on the efficacy of the isolated promoters, but we feel that these results indicate the importance of *cis* element position in the control of expression.

Discussion

Much of the genomes of both flies and worms consist of duplicated genes. Tandem or locally duplicated genes are more often seen in the worm than in the fly genome.[14] It is generally thought that duplicated paralogues 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 16 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. We found that 15 out of 16 tandem ABC genes could effectively drive GFP or RFP expression *in vivo*, which strongly argues against the above hypothesis.

Duplicated ABCs are most often seen in subfamily B and 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.[11] It is tempting to speculate that many other B subfamily members are involved in similar functions. The tissue-specific expression patterns of these duplicated ABC transporters suggest that the individual

Table 2. 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.

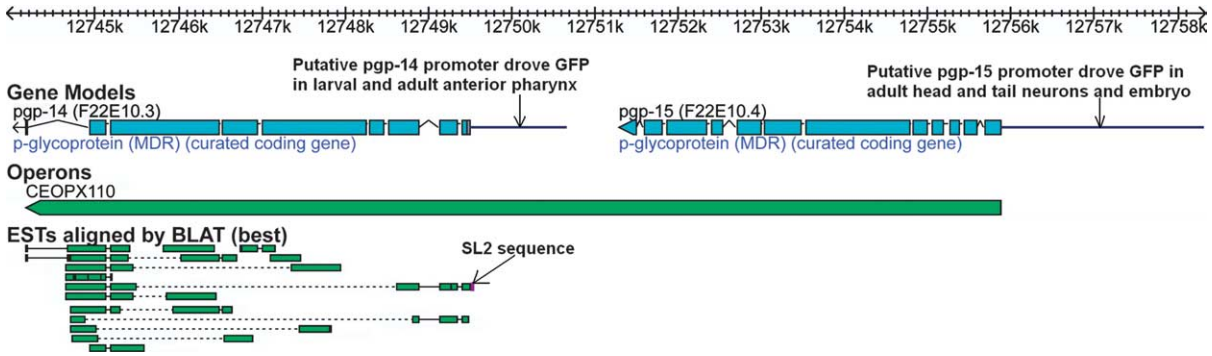

Figure 4. Presence of an SL2 sequence in a cDNA does not necessarily make the gene part of an operon. The picture was generated using the Genome Browse tool on the Wormbase web site (release WS130, available at http://wormbase.org/). The top line is a sequence scale in kilobase pairs. Blue bars represent exons and black lines introns. Purple bars 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 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* is not an operonic gene or else can be transcribed both as an operonic and a non-operonic gene.

genes have evolved to perform different physiological functions after the expansion of the cluster. The presence of multiple, similar, paralogues of drug resistance genes may provide more effective protection against xenobiotics by spatially or temporally differential expression. This is the case for eight members of the two four-gene clusters. Differential expression of five duplicated ABC genes was also observed in the mouse.[18] 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 a GFP expression signal and showed few matching ESTs in GenBank is C56E6.1, a member of subfamily H. This gene, like many poorly expressed in our 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 pseudo-gene because it gave an observable RNAi phenotype[20] and its cDNA has been successfully amplified.[21] Thus, it is possible that this gene is expressed at a low level, or else in the germline, in which tissue transgenes are likely to be silenced.[22] Most members of subfamilies A and G failed to give observable expression. Consistent with this observation, most of these genes had no, or few, ESTs in GenBank and few, if any, SAGE tags[23] are known from them. 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 found (D. Moerman, personal communication). In general, it appears that genes in head-to-tail orientation tend to more easily be expressed in our assay compared to those in a head-to-head orientation. It has been observed that co-expression of neighbouring genes in *C. elegans* is mostly due to the genes forming an operon or being recently duplicated genes.[24] Genes arranged in the same orientation may be more easily co-expressed or be under common regulatory control than those in the reverse arrangement.

We propose that not only can intron 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 neighbouring gene. For example, the 1.5 kb *Pmp-3* promoter cannot drive observable GFP expression whereas a 2.9 kb promoter, extending into the upstream coding region *Rfc-1* (C54G10.2) can effectively drive strong expression in intestine, vulva, hypoderm and neurons. Both *Pmp-3* and *Rfc-1* are well established based on abundant 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. 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[21] (Figure 4). Our transgenic data shows that a 1 kb or 1.7 kb *Pgp-14* promoter

† http://wormbase.org/

can drive strong GFP or RFP expression, respectively, 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 3). The presence of the operon is confirmed both by microarray data[25] and the presence of an SL2 sequence in the 5′ EST clone yk1279h09 (GenBank accession no. BJ121724).

We 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 1 kb, and can effectively drive strong reporter expression in pharynx. The lack of ESTs and RT-PCR products for *Pgp-15* indicates that our transgenic assay is more sensitive than the strictly molecular methods. It will be very interesting to examine whether other genes contained in operons with large intergenic spaces can practice dual transcription.

Materials and Methods

Promoter–GFP fusion construction

All the promoters are derived from wide-type N2 genomic DNA by PCR. Promoter–GFP fusion constructs were built as described.[26] GFP coding sequence was derived from A. Fire's vector, pPD95-67. The primer sequences used for GFP amplification were exactly the same as those used by O. Hobert. Similar C, D and D* primers were used to amplify RFP (Clontech) with sequences: 5′-CGCTCATCAAGAGAAAAATGG-3′, 5′-AAACGCGCGAGACGAAAG-3′ and 5′-GGAAACA GTTATGTTTGGTATATTGGG-3′, respectively.

In order to obtain putative promoters as accurately as possible for each gene, we set the following criteria for picking up promoter primers in terms of gene structure and organization in the genome. The 3′ primer (B) was placed as close as possible to, or spanning, the initiation codon (ATG) of the target gene. If a primer covered the ATG, it was mutated to ATC in order to avoid unnecessary translation initiation, which might cause a frame shift. We picked an ~3 kb region upstream from the ATG as our promoter if the intergenic region was equal to or more than 3 kb in size, based on the assumption that an average promoter size is ~1 kb in *C. elegans*. The full intergenic region was used as a promoter if this region ranged from 500 bp to 3000 bp; the promoters were arbitrarily extended into UTR or coding region of adjacent genes if the intergenic size was less than 500 bp, resulting in a 1–3 kb promoter. If no proper primer site could be found around the ATG, the B primer was shifted upstream, but we tried to include SL1 (usually TTTCAG) sites within the PCR product when present. The 5′ overhang (~20 bp) 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†.

Microinjection

We use the *dpy-5* gene (a gift from Dr Ann Rose) as a marker to screen for transgenic worms. The *dpy-5* rescuing plasmid, pCes-361 is co-injected with the promoter–reporter fusion construct into *dpy-5* mutant worms at concentrations of 100 ng/μl and 10 ng/μl. Wild-type F1 worms are picked after a four-day incubation at 20 °C and checked for wild-type F2 progeny after another three-day incubation. We attempted microinjections at most five times, when they failed to give F2 rescued progeny, with decreasing concentrations of PCR product.

Microscopy

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

† http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi

Acknowledgements

We thank Drs Ann Rose and Colin Thacker for the *dpy-5* rescue plasmid and mutant worms, Dr Andy Fire for the GFP vector, and Dr Don Moerman for unpublished SAGE data. We are grateful to the Sanger Centre and Washington University of St Louis for genome sequencing, and Drs John Tyson and Robert Johnsen for microscopy and helpful comments. This work depended crucially upon free access to the *C. elegans* sequence data in GenBank and WormBase, including EST sequences made available by Yuji Kohara. Funding provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) as well as Genome BC and Genome Canada.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2004.09.052

References

1. Childs, S. & Ling, V. (1994). The MDR superfamily of genes and its biological implications. In *Important Advances in Oncology* (DeVita, V. T., Hellman, S. & Rosenberg, S. A., eds), pp. 21–36, J. B. Lippincott Co., Philadelphia.
2. Dean, M. & Allikmets, R. (1995). Evolution of ATP-binding cassette transporter genes. *Curr. Opin. Genet. Dev.* 1995;, 5.
3. Brooks-Wilson, A., Marcil, M., Clee, S. M., Zhang, L. H., Roomp, K., van Dam, M. *et al*. (1999). Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. *Nature Genet.* **22**, 336–345.
4. Orso, E., Broccardo, C., Kaminski, W. E., Bottcher, A., Liebisch, G., Drobnik, W. *et al*. (2000). Transport of lipids from golgi to plasma membrane is defective in Tangier disease patients and Abc1-deficient mice. *Nature Genet.* **24**, 192–196.

5. Remaley, A. T., Rust, S., Rosier, M., Knapper, C., Naudin, L., Broccardo, C. *et al*. (1999). Human ATP-binding cassette transporter 1 (ABC1): genomic organization and identification of the genetic defect in the original Tangier disease kindred. *Proc. Natl Acad. Sci. USA*, **96**, 12685–12690.
6. Riordan, J. R., Rommens, J. M., Kerem, B., Alon, N. & Rozmahel, R. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*, **245**, 1066–1073.
7. Kuwano, M., Toh, S., Uchiumi, T., Takano, H., Kohno, K. & Wada, M. (1999). Multidrug resistance-associated protein subfamily transporters and drug resistance. *Anticancer Drug. Des.* **14**, 123–131.
8. Hyde, S. C., Emsley, P., Hartshorn, M. J., Mimmack, M. M., Gileadi, U., Pearce, S. R. *et al*. (1990). Structural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport. *Nature*, **346**, 362–365.
9. Sheps, J. A., Ralph, S., Zhao, Z.-Y., Baillie, D. L. & Ling, V. (2004). The ABC transporter gene family of *Caenorhabditis elegans* has implications for the evolutionary dynamics of multidrug resistance in eukaryotes. *Genome Biol.* **5**, R15.
10. Lincke, C. R., Broeks, A., The, I., Plasterk, R. H. A. & 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.
11. Broeks, A., Janssen, H. W. R. M., Calafat, J. & Plasterk, R. H. A. (1995). A P-glycoprotein protects *Caenorhabditis elegans* against natural toxins. *EMBO J.* **14**, 1858–1866.
12. Broeks, A., Gerrard, B., Allikmets, R., Dean, M. & 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.
13. Wu, Y. & Horvitz, H. R. (1998). The *C. elegans* cell corpse engulfment gene ced-7 encodes a protein similar to ABC transporters. *Cell*, **93**, 951–960.
14. Rubin, G. M., Yandell, M. D., Wortman, J. R., Gabor Miklos, G. L., Nelson, C. R., Hariharan, I. K. *et al*. (2000). Comparative genomics of the eukaryotes. *Science*, **287**, 2204–2215.
15. Copley, R. R., Letunic, I. & Bork, P. (2002). Genome and protein evolution in eukaryotes. *Curr. Opin. Chem. Biol.* **6**, 39–45.
16. Harrison, P. M., Echols, N. & Gerstein, M. B. (2001). Digging for dead genes: an analysis of the characteristics of the pseudogene population in the *Caenorhabditis elegans* genome. *Nucl. Acids Res.* **29**, 818–830.
17. 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 Res.* **10**, 192–203.
18. Annilo, T., Chen, Z. Q., Shulenin, S. & Dean, M. (2003). Evolutionary analysis of a cluster of ATP-binding cassette (ABC) genes. *Mamm. Genome*, **14**, 7–20.
19. Mounsey, A., Bauer, P. & Hope, I. A. (2002). Evidence suggesting that a fifth of annotated *Caenorhabditis elegans* genes may be pseudogenes. *Genome Res.* **12**, 770–775.
20. Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M. *et al*. (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature*, **421**, 231–237.
21. Reboul, J., Vaglio, P., Tzellas, N., Thierry-Mieg, N., Moore, T., Jackson, C. *et al*. (2001). Open-reading-frame sequence tags (OSTs) support the existence of at least 17,300 genes in *C. elegans*. *Nature Genet.* **27**, 332–336.
22. Kelly, W. G., Xu, S. Q., Montgomery, M. K. & Fire, A. (1997). Distinct requirements for somatic and germline expression of a generally expressed *Caernorhabditis elegans* gene. *Genetics*, **146**, 227–238.
23. McKay, S. J., Johnsen, R., Khattra, J., Asano, J., Baillie, D. L., Chan, S. *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, NY.
24. Lercher, M. J., Blumenthal, T. & Hurst, L. D. (2003). Coexpression of neighboring genes in *Caenorhabditis elegans* is mostly due to operons and duplicate genes. *Genome Res.* **13**, 238–243.
25. Blumenthal, T., Evans, D., Link, C. D., Guffanti, A., Lawson, D., Thierry-MIieg, J. *et al*. (2002). A global analysis of *Caenorhabditis elegans* operons. *Nature*, **417**, 851–854.
26. Hobert, O. (2002). PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans*. *BioTechniques*, **32**, 728–730.

Edited by J. Karn

(Received 20 August 2004; received in revised form 21 September 2004; accepted 21 September 2004)