Adding to the recombineering toolbox
interrogating internal transcriptional complexity in Caenorhabditis elegans operons via recombineered fosmid-based reporters

Hirani, Nisha

Awarding institution:
King's College London

Title: Adding to the recombineering toolbox
interrogating internal transcriptional complexity in Caenorhabditis elegans operons via recombineered fosmid-based reporters

Author: Nisha Hirani

Adding to the recombineering toolbox: interrogating internal transcriptional complexity in *Caenorhabditis elegans* operons via recombineered fosmid-based reporters

By Nisha Hirani

Thesis submitted for the degree of Doctorate of Philosophy

Kings College London
Institute of Pharmaceutical Science
Franklin-Wilkins Building
150 Stamford Street
London
SE1 9NH
UK

December 2012

Table of Contents

List of Figures

List of Tables

Abstract

Operons are common in prokaryotic genomes but, except for Nematoda, are uncommon in eukaryotes. Nematoda operons comprise gene clusters, containing between two to eight genes, with short, approximately. 100 bp, intergenic regions. Operons are common in *C. elegans* with an estimated 17% of genes clustered in operons. The significance of Nematoda operons is unclear although evidence suggests they may have arisen during genome evolution and compaction – the so-called "easy come, slow go" theory. Conventionally a discrete nascent transcript, transcribed from a major promoter upstream of the first gene in the operon, undergoes *trans*-splicing and 5′ addition of one of two splice-leader sequences followed by *cis*-splicing of the resulting pre-mRNAs to form mature, discrete mRNAs. However evidence suggests operon gene transcription is likely to be more complex. For example, microarray data indicated that correlated operon gene expression is weaker with increasing intergenic distance suggesting expression is influenced by sequence elements located internally within the operon structure. Further evidence, derived from comparison of expression data generated with fluorescent protein (FP) transcriptional reporters driven by either the upstream "operon promoter" alone or by such potential intergenic "internal promoter" sequences, generated different expression patterns in different tissues. Such data indicated the presence of internal regulatory elements within a subset of operons which were subsequently termed "hybrid operons". An alternative method to further dissect the presence of such internal regulatory elements would be to utilize translational style reporters built directly from genomic clones in which an entire operon structure is contained within the insert sequence. Tagging each operon gene with a different FP would permit operon gene expression analysis within the genomic context of that operon. To facilitate construction of fosmid-based translational reporters via seamless counter-selection recombineering two resource sets were first constructed. The first of these contained FP coding sequences synthesized *de novo* and codon-optimized for expression in *C. elegans*. The second set comprised a series of constructs designed to streamline the recombineering method. A subset of these resources was used to insert FP coding sequences seamlessly into the 3' ends of genes within three- (CEOP1312) and two- (CEOP1358) gene-containing operons each of which was located centrally on a different fosmid genomic clone insert. Comparative genomic analyses indicated that operon CEOP1312 was conserved across all available *Caenorhabditis* spp whereas, in contrast, operon CEOP1358 was not. Further investigation of the conservation of CEOP1312 within the *Caenorhabditis* spp lead to

the identification of a GC-rich non-coding sequence conserved in the intergenic region between the second and third genes in all the CEOP1312 orthologous operons. Making use of the available resource sets counter-selection recombineering was utilized to construct a series of fosmid-based translational reporters in which this conserved sequence was precisely, and seamlessly, manipulated. Although the absolute conservation of this non-coding sequence is strongly suggestive of functionality worms transgenic for these different constructs displayed essentially equivalent expression patterns for all CEOP1312 genes indicating that any such function may not, in this case, involve internal transcriptional regulation of one or more of the operon genes.

Acknowledgements

First of all I would like to thank my supervisor Dr Colin T Dolphin for his support, advise, encouragement and guidance throughout this project. I would also like to thank Dr Marcel Westenberg and Dr Minaxi Gami for their encouragement, guidance and assistance on the experiments. Thank you to a fellow colleague and dear friend, Miss Neeta Patel for her continuous support through this journey during the good and difficult times.

I would like to extend my thanks to our collaborator Professor Ian Hope, for his support and advise throughout the project. Many thanks to Dr Jayna Raghwani, Miss Roshni Shah, Miss Reshma Naker and Dr Taheera Ferdous for their friendship, patience, support and encouragement during the PhD. A very special thanks to Mr Mahesh Mistry for his words of encouragement early on during the PhD - they will never be forgotten.

Finally, I would like to express my gratefulness to my parents Prem and Devu Hirani and also my brother Kishen Hirani for their continuous love and support through the PhD.

Abbreviations List

AA = amino acid

Amp = Ampicillin

CDS = coding sequence

CFP = cyan fluorescent protein

Cm = chloramphenicol

EST = expressed sequence tag

FP = fluorescent protein

GFP = green fluorescent protein

GOI = gene of interest

HR = homologous recombination

Mb = megabase

mRNA = messenger RNA

NCS = non-coding sequence

NLS = nuclear localization signal

nt = nucleotide

ODN = oligonucleotide

ORF = open reading frame

Sm = streptomycin

TAP = tandem affinity purification

Tet(A) = Tetracycline

TF = transcription factor

YFP = yellow fluorescent protein

1

Background

1.1 *Caenorhabditis elegans*

The free-living, non parasitic, soil nematode *Caenorhabditis elegans* was identified by Sydney Brenner in 1974 and he proposed its use as a suitable multicellular model animal. In this study he described methods used to isolate and map *C. elegans* mutants and also characterized 300 mutants via morphology and behavior deformities. In 1998, this nematode became the first multicellular organism to have its 100 megabase (Mb) genome fully sequenced (*C. elegans* Sequencing Consortium, 1998) which revealed over 20,000 protein coding genes. From detailed analysis, and alignment of a portion of the protein coding genes, many genes were identified as orthologs to vertebrate and human genes illustrating a high conservation of biological mechanisms across the animal kingdom. This conservation demonstrated a direct link between *C. elegans* and vertebrate systems and a *C. elegans* homologue can usually be located for the vertebrates' gene of interest (Bürglin et al., 1998; Chalfie and Jorgensen, 1998; Kim, 2001).

The majority of the *C. elegans* population are hermaphrodite, producing both oocytes and sperm, and able to reproduce by self-fertilization making the generation of fully homozygotic populations (clones of the parent) possible. In comparison to the hermaphrodite, only 0.1% of the population naturally occurs as males. The hermaphrodites and males can be visually distinguished from their sizes, adult males are slightly smaller and thinner that the adult hermaphrodites, and also from their tail shape. The tail of the hermaphrodite is tapered in contrast to the male tail which has a fan like structure (Figure 1).

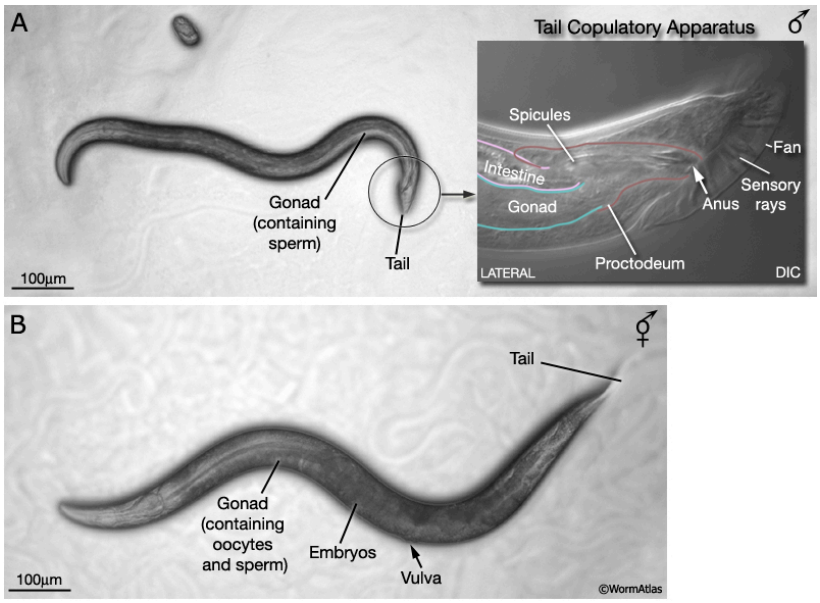

Figure 1. Brightfield images of an adult male (panel A) and an adult hermaphrodite (panel B) *C. elegans* (Lints and Hall, 2005).

As the maintenance requirements of *C. elegans* are a humid environment, ambient temperature and *Escherichia coli* bacteria (OP50) as a food source, the nematode is easy and inexpensive to maintain. Under laboratory conditions, the development of *C. elegans* is rapid. When maintained at 20 °C, development from an egg to a gravid adult, takes approximately. 3.5 days (Figure 2) and each gravid adult can potentially produce more than 300 progeny. The growth and development of *C. elegans* can be sped up/slowed down by adjusting the temperature at which the nematodes are maintained at, for example at 25 °C, they develop from egg to adult in under 3 days.

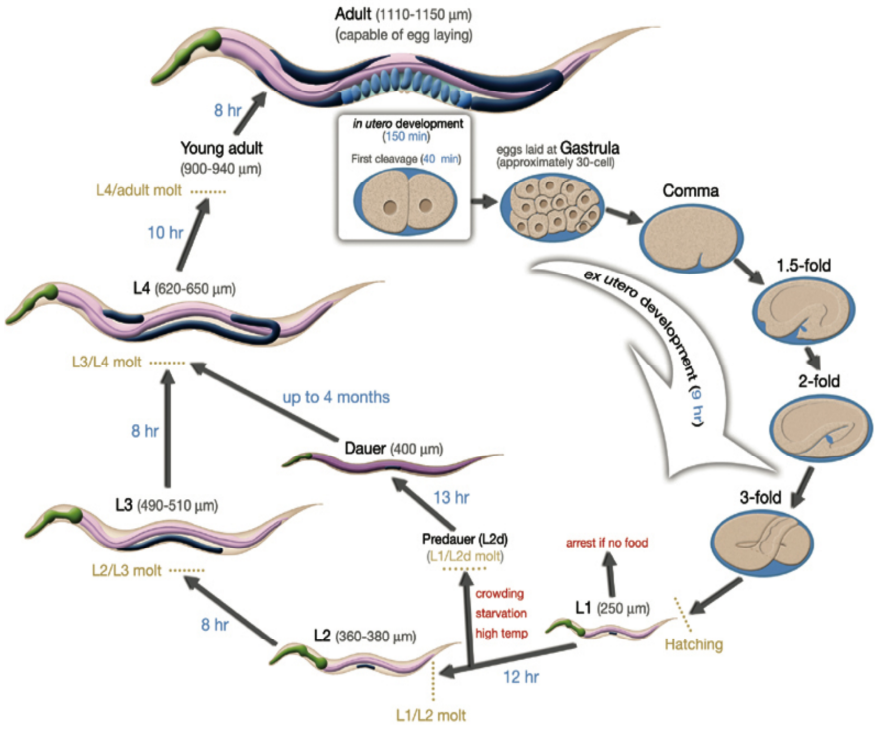

Figure 2. Life cycle of *C. elegans* at 22 °C. A diagram illustrating the life cycle, and time taken to develop to the next stage, of *C. elegans* when maintained at 22 °C (Lints and Hall, 2005).

The entire adult nematode consists of 959 cells (excluding those destined to become either egg or sperm) of which 302 are neurons and 95 make up the body wall muscles. The anatomy of the worm has been extensively studied and has been published in detail in the *C. elegans* atlas, also known as wormatlas (http://www.wormatlas.org/). The nematode is a tractable eukaryotic organism as Sulston et al in 1983 determined its complete cell lineage. Due to its transparent body, specific cell number, small fully sequenced genome and mode of reproduction the generation of strains of *C. elegans* transgenic for either a transcriptional or a translational reporter containing a fluorescent protein (FP) is relatively quick and easy (Antoshechkin and Sternberg, 2007). The ability to perform forward genetics, for example to yield key insights into conserved developmental processes, and reverse genetics, characterize the function of a target gene by inactivating the gene by RNAi (RNA interference), make the nematode an attractive model organism (Boulin and Hobert, 2011).

The transparent body of the nematode simplifies the process of tracking the protein product of a gene of interest (*goi*) via immunohistostaining or utilizing a transcriptional or translational reporter construct. Due to the short life cycle of the worm, transgenic animals can be generated within weeks, where it could take months in other model organisms, and are routinely done so via two well developed and routinely used

methods - microinjection (Mello et al., 1991; Mello and Fire 1995) and ballistic bombardment (Berezikov et al., 2004).

The microinjection technique as described by Mello et al (1999) detailed injecting the DNA into the gonad, more specifically into the core cytoplasm of the ovary, of the hermaphrodite worms as opposed to microinjecting into the nuclei of maturing oocytes (Fire, 1986). The cytoplasm in the gonad contains germ nuclei which have arrested in meiosis I. As the oocytes mature in the proximal region of the ovary, membranes encase individual nuclei along with portions of the cytoplasm, including the injected DNA. Due to the size of the gonad, in comparison to the worm, it is an easy target to microinject and successful injection into the gonad can be visualized by the influx of DNA which distorts the gonad structure. Mello et al (1991) also described the creation of a plasmid, pRF4, containing a mutant collagen gene, *rol-6* (*su*1006), which is used as a microinjection co-marker. Transgenic worms containing the semi-dominant marker, rol-6 (su1006) have a helically twisted cuticle and instead of moving in the usual sinusoidal motion, the transgenic worms roll over and more in circles.

Large numbers of worms can be quickly and easily generated within days making genetic studies less time consuming than other multicellular organisms. Both forward and reverse genetic screens can be performed with *C. elegans*. In forward screens, worms with a mutant phenotype are isolated and the genes responsible for the phenotype are identified. In reverse screens, the gene is first targeted through its sequence and then the mutation(s) that alter or inactivate its function is characterized. Reverse genetic screens can be readily performed utilizing the RNAi library (Fraser et al., 2000; Kamath et al., 2001). The RNAi library contains over 16,000 different clones, which cover 87% of the total *C. elegans* genes. The mechanism of action for RNAi occurs via the introduction of dsRNA (double stranded RNA, introduced by feeding) which results in the specific knock-down/inactivation of an endogenous gene and can be utilized for studying *in vivo* gene function and reverse genetics. Other popular experiments carried out with *C. elegans* include utilizing the worm as a tool for 'whole-animal drug screening'. The worms are exposed to small molecules of interest and their effect on the worm are deduced through a range of phenotypic assays such as life-span [to investigate any affects on the collective life-span of a collection of worm] and brood size [determining if the number of progeny produced by an exposed worm is affected].

As the worm is a popular and widely used model organism, there is also an online genome and biology database available to the worm community known as Wormbase (Stein et al., 2001; Harris et al., 2004; York et al., 2012). This invaluable resource contains, to name a few, the genome annotated with gene models, any homologous genes in *C. briggsae*, the sequence in each RNAi construct and the sequence in each fosmid clone.

1.1.1 Additional *Caenorhabditis spp*

An attractive trait of using *C. elegans* as an animal model is the availability of additional *Caenorhabditis* species with sequenced genomes, for example *Caenorhabditis briggsae*, allowing comparative genomic studies to be performed. *C. briggsae* is a natural companion to *C. elegans* for research due to the similarities between the two. The two species diverged from a common ancestor 80 – 100 million years ago (Stein et al., 2003) and yet are nearly identical in appearance and morphology. The phylogenetic relationships of *Caenorhabditis* nematodes revealed that *C. elegans* and *C. briggsae* are not sibling species and the closest relative to *C. briggsae* identified to date is in fact *Caenorhabditis remanei*.

The sequencing of the *C. briggsae* genome revealed many similarities between *C. briggsae* and *C. elegans* including the sizes of their genomes and number of chromosomes (6 chromosomes). The *C. briggsae* genome is slightly larger than the *C. elegans* genome and comparative genomic analysis revealed this was largely due to repetitive DNA content in *C. briggsae's* genome. Approximately 62% of the protein coding genes in *C. briggsae* have orthologs in *C. elegans* (Stein et al., 2003; Gupta et al., 2007).

1.1.2 Genome organization in *C. elegans*

C. elegans has a compact genome of 100 Mb, condensed within six chromosomes (1 to 5 and X) which encodes for over 20,000 protein-encoding genes. These genes were predicted during the genome sequencing via the GeneFinder sequencing project that predicted coding sequences. Over time, these annotations have been modified as

mRNA sequences, ESTs (expressed sequence tags) and OSTs (ORFeome sequence tags) became available. As the sequencing of the genome was going to take several years to complete, and to get an early insight into gene sequences a large project was undertaken by Yuji Kohara which yielded more than 18,500 ESTs representing more than 3,500 genes (Riddle et al., 1997). The ORFeome cloning project provided confirmation of approximately. 55% of the predicted protein-coding ORFs (open reading frames) and now there are approximately. 12,500 *C. elegans* ORF clones available for further studies (Lamesch et al., 2004).

Genes within the *C. elegans* genome are found on both strands, and in instances have been found within introns of other genes (both on the opposite and same strand) and in some rare cases, exons from two distinct genes can overlap. The gene organization in *C. elegans* is both unusual and interesting for a multicellular eukaryotic organism as a portion of the genes has been found in closely spaced clusters and these gene-clusters have been named operons.

1.1.3 Splicing in *C. elegans*

Transcripts within the worm genome, as with other eukaryotic organisms, undergo *cis*-splicing. Most genes, in *C. elegans*, are single-gene transcription units and frequently contain introns which are *cis*-spliced during the formation of the mature transcript. In addition to *cis*-splicing, *C. elegans* genes also undergo trans-splicing.

1.1.3.1 *Trans*-splicing

Trans-splicing is a common occurrence within the *C. elegans* genome and approximately. 70% of genes are *trans*-spliced via SL1 (Spliced leader 1), the major spliced leader. *Trans*-splicing is catalyzed by spliceosomes which include U2, U4, U5 and U6 small nuclear ribonucleoprotein (snRNP) but not U1. The donor in *trans*-splicing is a snRNP and there are two types named SL1 and SL2 (Spliced leader 2), although there are several variants of SL2 which have been termed SL3-SL12 (MacMorris et al., 2007). *Trans*-splicing of operon genes occur via two different snRNPs, SL1 (GGTTTAATTACCCAAGTTTGAG) and SL2

(GGTTTTAACCCAGTTACTCAAG). Both SL1 and SL2 are 22 nt sequences, which are donated by the snRNP to the 5′ mRNA molecule adding a TMG (trimethylguanosine) cap to protect the mRNA from exonucleolytic degradation, and the mRNA is also 3' polyadenylated and subsequently *cis*-spliced (introns removed) to form a mature mRNA unit which is then translated. The mechanism of *trans*-splicing is very similar to that of *cis*-splicing (Figure 3). The *trans*-splicing initiating sequence on mRNA is the same as the intron 3' splice site, also known as the outron, when present at the 5′ of mRNA. The initiation of the process of *trans*-splicing and *cis*-splicing are differentiated by the absence or presence of the upstream 5′ splice site respectively (Conrad et al., 1991; Conrad et al., 1993; Conrad et al., 1995).

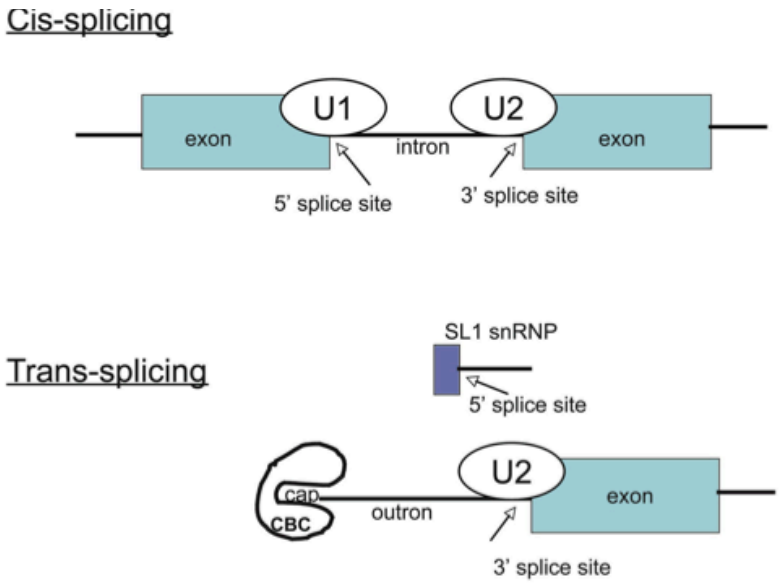

Figure 3. Comparison of *cis*- and *trans*- splicing. In *cis*-splicing, the U1 snRNP pairs with the 5′ splice site and the U2 snRNP pairs with the 3' splice site leading to the intron being excised and the two exons are bought together. In *trans*-splicing there is no 5′ splice site and instead the donor SL snRNP provides the site. The SL snRNP interacts with the U2 snRNP, located at the 3' splice site, resulting in the formation of an exon on the pre-mRNA. The region between the 5′ cap and *trans*-splice site is called the outron and the CBC illustrates the nuclear cap binding complex (Blumenthal, 2005).

In *C. elegans* the *trans*-splicing occurs extremely close to the initiating methionine of the protein. The TMG cap present on the SL RNA becomes the TMG cap which protects the 5′ end of the mRNA, after *trans*-splicing, and the presence of this TMG cap activates translation (Krause and Hirsh, 1987). The *C. elegans* genome contains 110 SL1 RNA genes where as only 10 SL2 genes have been identified (Stein et al., 2003). *Trans*-splicing occurs in many species in the nematode phylum and conservation of the

SL sequences has been observed (Blumenthal and Steward, 1997). With over 17% of the total genes in the *C. elegans* genome contained within operons, these structures are a common occurrence. Two types of operons have been identified in the *C. elegans* genome, SL1-type operon and SL2-type operons.

1.1.4 Operons

Traditionally, operon gene organization was exclusive to prokaryotic organisms where multiple genes were under the control of a single major upstream promoter and the clustered genes functionally related i.e. all the gene protein products were required in the same metabolic pathway or at the same time in the same organelle. A well-defined example of this is the *lac* operon in bacteria that contains genes responsible for the transportation and metabolism of lactose (Beckwith, 1967).

Operons in eukaryotes were first described in Trypanosomes in 1987 (Imboden et al., 1987) and also primarily described in animals, namely the nematode *C. elegans*, in 1993 (Spieth et al., 1993). Additional eukaryotic organisms have been found to contain two gene-operons (Drosophila (Andrews et al., 1996; Liu et al., 2000), vertebrates (Sloan et al., 1999; Corcoran et al., 2004) and plants (Garcia-Rios et al., 1997; Thimmapuram et al., 2005)). In the instances of eukaryotic organisms (Drosophila, vertebrates and plants) containing only two-gene operons, the dicistronic mRNA is transcribed from the operon gene sequence and is transported to the cytoplasm where it undergoes translation. There is no *trans*-splicing of the dicistronic mRNA to form monocistronic mRNAs before translation.

Polycistronic gene clusters, or operons, were discovered in *C. elegans* in 1993 (Spieth et al., 1993) and more than 17% of genes within the genome are found in such arrangements. Operons in *C. elegans* were discovered due to the tightly clustered genes, in the same orientation, with the downstream genes in these clusters being *trans*-spliced with the rare SL2 *trans*-splice leader. These polycistronic gene clusters contain between two to eight genes, on average an operon contains 2.6 genes, with short intergenic regions of approximately. 100 – 400 bp and are found on all chromosomes. The genes within the clusters are transcribed as a single nascent mRNA transcript, which immediately undergoes *trans*-splicing to form monocistronic mRNAs.

The formation of operons is essentially a one-way process as once they have formed it would be very difficult to break the structure as the breakage would leave the downstream operon genes without a promoter and hence would not be expressed – the so-called "easy come, slow go" theory (Qian and Zhang, 2008). Operons in the *Caenorhabditis* genus are highly conserved, especially between *C. elegans* and *C. briggsae* species as 90% of the *C. elegans* operons are conserved in *C. briggsae*, and these two species are thought to have diverged more than 100 million years ago (Stein et al., 2003).

There are two main theories to explain the phenomena of operons in the *Caenorhabditis* genus: grouping of functional genes and simple genome compaction. The first theory is operon structures have grouped genes that are from the same gene class e.g. mitochondrial proteins (Blumenthal and Gleason, 2003) or are functionally related genes i.e. those whose protein products are required for transcription, splicing and translation. Classes of genes that tend to be found in operons are those encoding proteins which are involved in RNA decay, mitochondrial functioning and DNA replication (Blumenthal and Gleason, 2003). Some operons contain functionally co-regulated genes for example two subunits of the acetylcholine receptors, genes *deg-3* and *des-2*, are located in operon CEOP5284 (Treinin et al., 1998). Others operons have been identified as containing related genes for example genes encoding RNA polymerase I and a regulator of ribosome synthesis have been found within a single operon. However, in many instances operons contain genes with no discernable relationship between them. More recently genes that are expressed in the oogenic germline, not genes expressed in spermatogenesis, have been found in operons suggesting genes which are highly expressed can be grouped within operons (Reinke and Cutter, 2009).

The second theory is the operons have simply formed due to compaction of the genome (Qian and Zhang, 2008). The genes which have been clustered have no need to be regulated as their expression is always required, or is regulated at the mRNA or protein level, and thus removing the large promoter regions of genes which are closely spaced together would reduce the DNA between the genes and slightly compact the genome.

Recent evidence has reexamined operons resulting in a degree of heterogeneity. This has led to a classification of the *C. elegans* operons into two types, SL1-type operons and SL2-type operons.

1.1.4.1 SL1-type operons

In this type of operon, the downstream genes are *trans*-spliced via SL1 alone. The poly-adenylation site and the *trans*-splice site are adjacent and therefore there is no intergenic sequence. To date very few operons of this type have been identified (approximately. 20) and the pre-mRNA from these either gives rise to mRNA of the upstream or downstream gene in the operon but not to both (Williams et al., 1999; Blumenthal, 2005).

1.1.4.2 SL2-type operons

The majority of operons were identified via the rare SL2 *trans*-splicing sequence through a genome-spanning microarray (Blumenthal et al., 2002). This microarray, performed on over 94% of confirmed and predicted genes, identified over 1000 genes (approximately. 17% of the genome) that are found within operon clusters with each cluster containing between two to eight genes (Blumenthal et al., 2002). *C. elegans* operons produce a polycistronic mRNA molecule that immediately undergoes *trans*-splicing via SL1 and/or SL2 to produce discrete monocistronic pre-mRNAs which are then *cis*-spliced to create mature mRNAs. In the majority of operons the first gene is *trans*-spliced to the SL1 spliced leader and all subsequent genes in the operon are *trans*-spliced by SL2 (Figure 4).

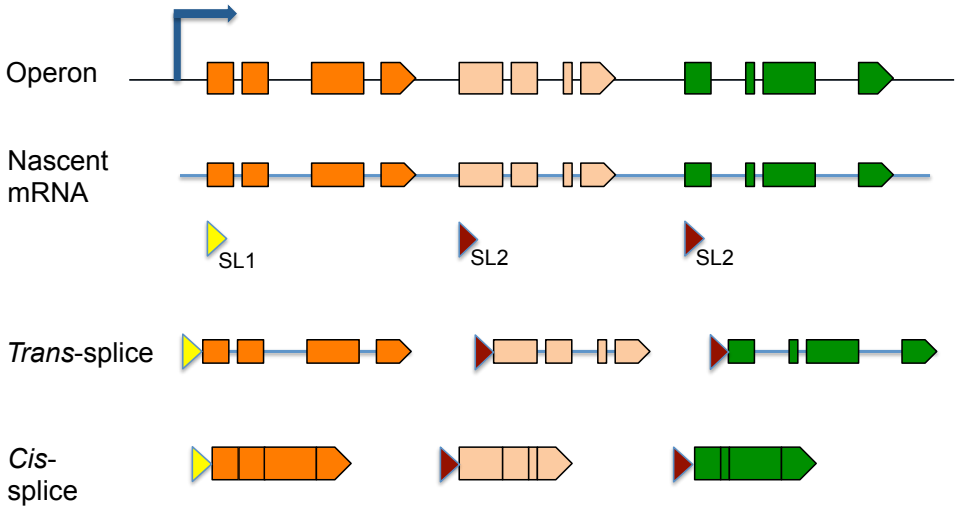

Figure 4. Summary of the *trans*- and *cis*- splicing of operon genes. The above summarizes the procedure operon genes go through to generate protein. The genes within the operon are transcribed as nascent polycistronic pre-mRNA which undergoes *trans*-splicing, via either SL1

or SL2, to generate monocistronic mRNAs and is finally *cis*-spliced (to remove introns) creating a mature mRNA that can be translated.

The pre-mRNA molecules, are rarely detected, via RT-PCR or protein western blot, suggesting that either the *trans*-splicing occurs immediately or occurs concurrently with the initial transcription.

In addition to the splice site there are two sequences within the pre-mRNA that are important in SL2 *trans*-splicing and these are the two signals for the 3' end formation. A sequence 5′ of the cleavage site, AAUAAA, is required for 3' end formations and binds CPSF (cleavage and polyadenylation specificity factor) and a U-rich sequence 3' of the cleavage site binds CstF (cleavage site stimulatory factor). When these two factors bind, they cooperate and bring the two sites together to define the cleavage site. The AAUAAA site is not required when SL2 *trans*-splicing occurs. The protein which binds to the U-rich region, CstF, performs two functions - it protects the downstream RNA from exonucleolytic attack and degradation and also recruits the SL2 snRNP (Liu et al., 2003; Blumenthal, 2005).

Many examples have been identified where the genes within a single operon in *C. elegans* are co-expressed, in both time and space, however there are also numerous examples where operon-clustered genes exhibit different and distinct patterns of spatio-temporal expression and many theories have arisen to explain the discrepancies. One theory suggests it is because the mRNAs have different stabilities, leading to different half-lives. These operon genes are regulated post-transcriptionally leading to differences in individual mRNA stabilities. It is also possible that the translation of the individual mRNAs could be controlled, for example micro(mi)RNA mediated inhibition (Gu and Kay, 2010). However there are also clear examples where the genes within the operon are of related function and co-expressed by the single promoter. For example operon CEOP5284 contains *des-2* and *deg-3* which encode subunits of the acetylcholine receptor (Treinin et al., 1998). Another observation made regarding genes within operons is that some classes of genes are clearly found in operons in comparison to others which are rarely located in operons or are completely missing (Blumenthal and Gleason, 2003). In 2009 work was published which suggested that the driving force for including genes within operons is if it is necessary for the gene to be highly expressed in the female germ line (Reinke and Cutter, 2009).

Conventionally a discrete nascent transcript, transcribed from a major operon upstream promoter, undergoes *trans*-splicing followed by *cis*-splicing to form mature mRNA. However several different sources suggest that operon gene transcription is more complex. Microarray data revealed a correlation between gene expression level and distance from the operon promoter. For example, Reinke and Cutter (2009) examined microarray expression data and revealed that the further away the gene is from the promoter, the weaker the expression suggesting expression of the distant genes in the operon is influenced by intergenic sequence-located elements (Reinke and Cutter, 2009). However, downstream genes in some operons have been found to use SL1 alongside SL2. Evidence, in the form of transcriptional GFP (green fluorescent protein) reporters of the major upstream promoter and intergenic regions, revealed different GFP expression patterns with some being expressed in different tissues (Vázquez-Manrique et al., 2007; Huang et al., 2007). Huang et al (2007) coined the term 'hybrid operon' to describe operons with such internal promoter elements. This adds an additional layer of complexity to operon gene structures in *C. elegans* as it demonstrates that within the operon structure there are regulatory elements in the intergenic sequences that influence, and may in some cases drive, the expression of some operon genes (Huang et al., 2007). Additional evidence in the form of HTZ (histone variant H2A.Z which is a functionally specialized region of chromatin) ChIP-chip analysis revealed 37% of operons contained internal peaks suggesting as many as 37% of operons contain internal promoters (Whittle et al., 2008). Allen et al (2011) demonstrated via analysis of RNA deep-sequencing data that intercistronic distances of genes with internal regulatory elements in hybrid operons are on average longer than the intercistronic distances in operons. It was also hypothesized that genes with internal promoter elements within hybrid operons when transcribed from the operon promoter receives SL2 and when transcribed from the proximal promoter receives SL1 (Allen et al., 2011).

1.2 Gene expression analysis methods in *Caenorhabditis elegans*

Analysis of gene expression in *C. elegans* can be performed via many different methods. These methods can range from detecting changes in gene expression for a few selected

genes to a more high-throughput methods allowing detection of expression levels of nearly all genes within the *C. elegans* genome.

1.2.1 Microarray

Microarray immobilizes miniature DNA probes (short specific sequences) that span the genome to assess the gene expression level of nearly every gene in the *C. elegans* genome. Microarray was first introduced into the *C. elegans* community by Kim et al (2001) at Stanford as they developed the first *C. elegans* whole-genome arrays (Kim et al., 2001). They also provided a free resource to the community of hybridization and scanning services. However this resource is no longer available free to the community and subsequently, as one of the pioneers of commercial microarrays, the Affymetrix *C. elegans* gene chip is a popular choice in the community. Due to the sensitivity of this method, subtle changes within the genome when exposed to difference environmental factors, including particular growth conditions, or exposure to compounds of interest can be detected and the genes affected identified for further investigation. A genome-wide microarray allows rapid assessment of >20,000 *C. elegans* genes in an extremely short amount of time. However, this method is limited to only assess the genes which are represented, via the DNA probes, on the chip (whether the chip has been commercially purchased or has been specially generated) thus excluding unannotated genes and alternative transcripts from analysis (Portman, 2006).

1.2.2 SAGE

SAGE (serial analysis of gene expression) is another method that allows large-scale profiling of transcript levels for thousands of genes across the genome. SAGE works by ligating 3' ends of complementary DNA (cDNA) fragments together to form concatamers, which are subsequently sequenced and the sequences reads obtained are mapped to the mRNA template of the cDNA. Advantages of this technique include no requirement of specialist equipment, apart from a high-throughput DNA sequencer, and its ability to detect unannotated genes and also allowing direct quantitation of relative expression levels of additional genes. However, its biggest drawback is that it can be difficult to control for biological reproducibility (Velculescu et al., 1995).

1.2.3 Immunofluorescence staining

Immunofluorescence staining illustrates a direct link between the distribution, cellular and subcellular, of the protein of interest. Antibodies, conjugated to a visual tag, are used to illustrate the distribution of the genes' protein product. Limitations in this method include the requirement for specific antibodies and also the need to fix animals due to the impermeable cuticle of the worm (Duerr, 2006).

1.2.4 Reporter gene fusions

1.2.4.1 LacZ

LacZ, one of the three structural genes of the *E. coli lac* operon, encodes the enzyme β-galactosidase (β-gal) that cleaves the disaccharide lactose into glucose and galactose. In tandem with the chromogenic substrate 5-bromo-4-chloro-indolyl-β-D-galactopyranoside (X-gal), β-gal has been widely used as a reporter protein in a number of *in vitro* and *in vivo* systems. β-gal-catalyzed X-gal hydrolysis generates an insoluble, intense blue product thereby localizing the reporter protein. Fire et al. (1990) generated a comprehensive plasmid-based resource for gene expression analysis in *C. elegans* that utilized β-gal as the reporter. Although the catalytic nature of the β-gal reporter ensures sensitivity its use requires that worms are permeabilized and fixed and therefore the approach cannot be used in live specimens (Boulin et al., 2006).

1.2.4.2 Fluorescent proteins

In 1994 Martin Chalfie (Chalfie et al., 1994) reported on their development of green fluorescent protein (GFP), isolated from the jellyfish *Aequorea victoria*, as a gene expression marker demonstrating its utility in *C. elegans* by expressing GFP in the sensory neurons. In the same year the ability to use GFP as a gene expression reporter was also demonstrated in the prokaryotic cells *E. coli* (Inouye and Tsuji 1994).

GFP is a 27 kDa β-barrel protein made up of 11 β-sheets which surround a central α helix and the chromophore responsible for the fluorescence is comprised of only three amino acids – Serine 65, Tyrosine 66 and Glycine 67 (Figure 5). In its natural state, GFP can form dimers but mutating amino acid at position 206 from Alanine (A) to Lysine (K) (Zacharias et al., 2002) solved this and prevents dimerization from occurring.

The biophysical characteristics of GFP make it a valuable reporter which can be utilized in many animal models. The small globular structure of GFP minimizes interactions with any potential protein partners. GFP also does not require any co-factors for functioning *in vivo* and therefore would not deprive the host cell of any co-factors. It is also non-toxic as host cells can tolerate high concentration of GFP without any detrimental affect. The high brightness and photostability of GFP also allows imaging to be performed over time and proteins expressed at low levels to be detected.

The inherent "brightness" of a fluorescent protein can be defined as the product of the molar extinction coefficient and the quantum yield $(mM \times cm)^{-1}$ and is usually compared to the well-characterized and well-used standard enhanced green fluorescent protein (eGFP). The molar extinction coefficient, also known as the molar absorption coefficient or molar absorptivity, is a measurement of how strongly a chemical species absorbs light at a given wavelength and this value can be predicted from the amino acid sequence. The quantum yield is the efficiency of the fluorescence process which is defined as the ratio of the number of photons emitted to the number of photons absorbed.

Photostability is the stability of the protein, and the fluorescence expressed, during imaging. The stability is influenced by the photobleaching as it is defined as the time required to bleach from an emission rate of 1000 photons per second (t ½) in a widefield fluorescence microscope.

Following on from the work that demonstrated the usefulness of GFP (Tsien, 1998), the original GFP was subsequently engineered to produce additional *Aequorea*-based fluorescent proteins (FPs), which emit in the blue (BFP – blue fluorescent protein) (Heim et al. 1994, Heim et al., 1996), cyan (CFP – cyan fluorescent protein) (Heim et al. 1996; Tsien, 1998) and yellow (YFP – yellow fluorescent protein) (Ormo et al. 1996; Wachter et al. 1998) wavelength regions. The blue and cyan variants were

created by modification of the chromophore structure alone in comparison to the yellow variants which retain the GFP chromophore but are a result of a single amino acid substitution of T203Y which produced a more red-shifted FP (Wachter et al., 1998). However, no *Aequorea*-based fluorescent proteins could be engineered with orange or red wavelength emission peaks.

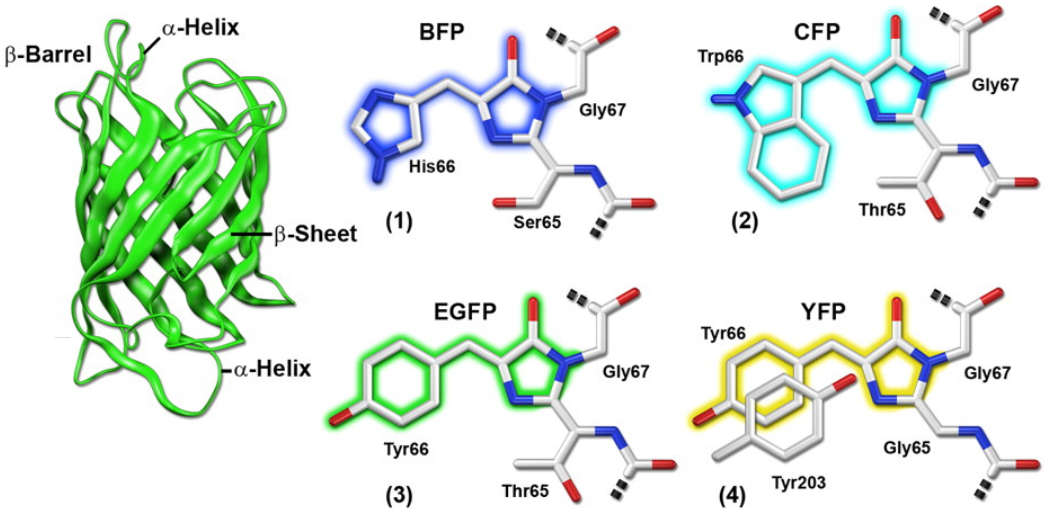

Figure 5. Tertiary and chromophore structure of *Aequorea*-based FPs. This illustration depicts the tertiary structure of GFP and all variants derived from GFP. The variations in the three amino acid chromophore are shown for BFP (1), CFP (2), EGFP (3) and YFP (4). Image modified from (Shaner et al., 2007).

Subsequent refinements of CFP and YFP yielded structures that had increased photostability and brightness, CFP variants included enhanced CFP (eCFP, Cubitt et al., 1995) and mCerulean (Rizzo et al., 2004) and YFP variants included mVenus with reported faster maturation (Nagai et al., 2002), and mCitrine with greater photostability and brightness (Griesbeck et al., 2001).

FPs in the orange and red regions remained elusive until the discovery of bioluminescent reef corals (Matz et al., 1999). The majority of FPs sourced from reef coral naturally exists in dimeric or tetrameric conformations and as these conformations are large in size and tend to aggregate they can potentially interfere with the localization of the tagged protein. Extensive research was carried out and the first red FP, monomeric red fluorescent protein (mRFP), a monomer resulting from 33 mutations made to the β-barrel structure, was derived from *Discosoma striata* (Campbell et al., 2002). However, photostability was still an issue and efforts were continued to produce a FP in the red region with properties that were comparable to eGFP. One line of experimental effort focused on the addition of amino acids from eGFP to both

the N- and C- terminal of the protein alongside directed mutations of residues near/surrounding the chromophore. This effort resulted in the FP, mCherry, that was red-shifted in the emission wavelength and possessed better photostability (Shaner et al., 2004).

In 2004 Shaner et al., utilizing site-directed mutagenesis of mRFP1, a development of mRFP (Campbell et al., 2002), and screening via fluorescence-activated cell sorting, developed a range of eight new FPs, six of which were monomers, that spanned the yellow, orange and red spectra and were termed the fruit FPs (as many of these FPs were named after fruits i.e. mCherry) (Shaner et al., 2004). These FPs exhibited several improved properties including improved tolerance to N- and C-terminal fusions, increased extinction co-efficient and also increased photostability. However, none of the new FPs was optimal for all these properties. A dimer fruit FP, tdTomato, was also created possessing the highest photostability and brightness out of all the fruit FPs. As tdTomato is a dimer the size and molecular weight of the final fusion protein could interfere with its localization and limit its use (Shaner et al., 2004).

Because they would posses less phototoxicity and would also allow deeper imaging into biological tissues, for example whole-animal imaging, there was also a need for FPs in the far-red region of the spectrum. Expansion of the Fruit FPs yielded the first potentially useful far-red FP, mPlum (Wang et al., 2004). Limited in its brightness but with high photostability, mPlum could still be utilized in multi-label and FRET experiments.

In 2007, a combination of random and site-specific mutagenesis was applied to create a number of TurboRFP variants that contained mutations at specific positions around the chromophore. The screening of the TurboRFP variants identified Katushka the brightest far-red FP described to date (Shcherbo et al., 2007). However, as Katushka is a dimer protein the mutations surrounding the chromophore were engineered into TagRFP, a monomeric red FP (RFP) generated from a wild-type RFP sourced from a coral (*Entacmaea quadricolor*), (Merzlyak et al., 2007) generating a far-red FP with reported high brightness and exceptional photostability called mKate (Shcherbo et al., 2007).

FPs in the yellow, orange and red regions are not the only ones to be derived from a coral source and FPs with emission spectras in the cyan and green regions have also been isolated. For example, monomeric teal fluorescent protein (mTFP1, Ai et al.,

2006), a cyan-emitting FP, was generated from a naturally forming tetrameric CFP from *Clavularia* coral. With its high photostability, brightness, resistance to pH changes and high quantum yield, mTFP1 is an attractive candidate for use as a fluorescent reporter.

1.2.5 Tandem affinity purification (TAP) tags

A research team at EMBL invented the concept of tandem affinity purification (TAP) tags in the late 1990s (Riguat et al., 1999; Puig et al., 2001). TAP is a technique used to study protein-protein interactions via a fusion protein with a TAP tag on either the N- or C-terminus of the protein. TAP tags are composed of two different tags separated by a protease cleavage site. The external, primary tag exhibits strong binding affinity to its binding partner via an irreversible interaction. The protein of interest fused to a TAP tag is passed through a column filled with agarose beads coated in the binding partner of the first external tag capturing the protein. A protease is then passed through the column, cleaving at the protease site, releasing the captured protein from the column. The elute is passed through a second column filled with agarose beads coated with the binding partner of the second tag allowing an intact protein complex to be captured. The interaction between the second tag and its binding partner is reversible and in the presence of the correct ligand allows elution of the captured complex off the beads under gentle conditions. A large variety of primary and secondary tags have been developed with their function in TAP tags assessed. The actual tags selected for use in the TAP tags are dependent upon the system that they will be used in (reviewed by Li, 2010; Xu et al., 2010; Li, 2011).

1.2.6 Fluorescent proteins: use as *C. elegans* gene expression reporters

In part because of the requirement that samples need to be fixed and permeabilizied when *LacZ* is used as a reporter, FPs increased in popularity as the fluorescent signal could be visualized and followed in live samples. Additionally the small size of the nematode (length, width and thickness) and transparency of the body, throughout all stages of development, are ideal conditions for using fluorescent proteins.

Soon after the original report of Chalfie et al. (1994), GFP and variants were incorporated into the previously generated *C. elegans* expression plasmid set (Fire et al., 1990), resulting in a range of plasmids containing CFP, GFP, YFP and β-gal reporters, and from here on in when referring to these CDSs, they will be prefixed with F-. Generated as different sets of constructs either lacking a promoter sequence, thus enabling the insertion of a DNA sequence derived from a gene of interest, or built with a variety of useful promoter sequences, such as those driving the endogenous *hsp-x* (heat shock protein) genes, the complete resource has since proved to be of immense utility to worm researchers the world over. Subsequent iterations of this resource set included the insertion into the reporter coding sequence of differing numbers of small, artificial introns as the presence of such sequences, thus generating a form of mini-gene, are known to increase overall protein expression levels (Buchman and Berg, 1998). Inclusion of such introns proved to indeed increase both FP and β-gal expression from all of the so-called Fire vectors.

In addition to the 'Fire vectors' readily used by the worm community, a sequence encoding mCherry (Shaner et al., 2004) was codon-optimized for expression in *C. elegans* and, in addition, contained three equally-spaced artificial introns (from here on in will be referred to as Mc-mCherry, McNally et al., 2006). The codon-optimization of the sequence required several rounds of manual intervention for optimal expression in *C. elegans* (J. Audhya, pers. comm.).

1.2.7 Gene expression reporters

FPs can be used as reporters to demonstrate gene expression and are primarily utilized in two types of reporter constructs, "transcriptional" and "translational". In the former the promoter region is typically defined as the 5′ intergenic region of the gene of interest (*goi*). This region is used as promoter fragments a few kb immediately upstream from the start codon contains a significant amount of *cis*-regulatory information to drive expression of the reporter and provide a tentative expression pattern for the *goi*. In a translational reporter the reporter gene is fused, in-frame, to the *goi* coding sequence and the upstream promoter region and translation of the entire construct will result in the GOI protein fused to the FP which in most cases is GFP. These constructs are usually larger and more difficult to generate using traditional cloning techniques in comparison to transcriptional reporters.

1.2.7.1 Transcriptional reporters

Transcriptional reporters comprise the intergenic sequence upstream of the start codon of the *goi* placed in front of gfp CDS to drive its expression. These constructs can be made via the traditional cloning methods using restriction enzymes and ligation. Further efficient methods, PCR (polymerase chain reaction) stitching and Gateway-based cloning, have been developed to streamline the process of generating transcriptional reporters. PCR stitching is a simple method whereby the promoter region of the *goi*, and the *gfp* CDS and the generically used, but not in all cases, 3' untranslated region (3' UTR) of *unc-54* (both the CDS and 3'UTR are present in the standard fire vector pPD95.75) are PCR-amplified separately. The two products are then joined together by a second PCR step, resulting in a complete fusion gene, which can be used directly for microinjection (Hobert, 2002). The Gateway method utilizes two specific *attB* sequences that flank the promoter region of interest; these *attB* sequences are introduced when PCR-amplifying the promoter region. These *attB* sequences are substrates for the recombination that results in the promoter region being introduced into a vector plasmid, containing only gfp CDS, to generate a promoter construct. The recombination reaction results in an additional 21 bp sequence; however there is nothing to suggest that it has an effect on the expression patterns obtained (Bamps and Hope, 2008).

1.2.7.2 Translational reporters

Translational reporters are in-frame gene fusions between *gfp* and the *goi*, ideally including its 5′ upstream region, exons, introns and also its 3' UTR. As previous work has shown, although the genome of the worm is compact not all regulatory elements associated with a specific gene are necessarily located simply 5′ of the start codon. Evidence suggests that potential transcription regulatory elements can be located further than the 5′ 'promoter' sequence, for example in introns or in 3' UTRs. In particular, the 3' UTR may be of significance due to potential post-transcriptional regulation by microRNAs (miRNAs) that function to repress target mRNAs by binding specifically in this region. Hence, translational reporters are more likely to recapitulate an expression pattern close to that of the endogenous gene, because of the inclusion in the resulting construct of many, if not all, of the associated sequence elements that collectively regulate expression of the *goi*.

1.2.7.3 Genomic Clones

Genomic clones for *C. elegans* are available in the form of YACs (yeast artificial chromosomes), cosmids and fosmids. YACs are vectors which can be used to clone between 100 kb to 3000 kb of DNA sequence and can only be propagated in yeast. However there are stability issues with them as there is a 50% chance of chimerism, artifacts where the sequence of the cloned DNA doesn't correspond to a single genomic location but to multiple locations. Other problems can include deletions and rearrangements of the genomic sequence as, due to the size, the clone may have undergone intermolecular recombination. Cosmids are plasmids that contain the lambda phage *cos* sequence and are able to contain inserts of 35 kb to 50 kb. Cosmids replicate as plasmids, depending upon the origin of replication, and multiple copies would be present in a host cell. Fosmids are similar to cosmids but are based on bacterial F-plasmid and can contain DNA inserts of approximately. 40 kb. Unlike cosmids, fosmids are usually maintained at single or low copy numbers in the host cell and hence the ability to perform standard cloning can be limited. However the low copy number of fosmid clones offers a higher stability, in comparison to cosmids, and will preserve the sequence of the DNA insert.

A fosmid library was created (D. Moerman, pers. comm.) and covers approximately. 80% of the genome and provides at least one clone for approximately. 90 % of protein-coding genes. The fosmid library comprises genomic clones (for wild type N2 DNA) with 35 to 40 kb inserts in the inducible backbone pCC1FOS (Epicentre). Each fosmid clone, on average contains 8 complete genes, and therefore is more likely to contain all relevant regulatory elements and as the library covers the genome approximately. 5 times, the chance of identifying a fosmid clone with the centrally located gene of interest is relatively high. This valuable resource is available to the worm community and comprises fosmid clones that can be manipulated for generating translational reporters to investigate gene expression. Due to the large size of each fosmid clone, it is impractical to use traditional genetic engineering to manipulate the clones and to insert a reporter sequence into a chosen site as the possibility of locating a restriction enzyme site, unique within the whole clone, at the target location is extremely low.

An additional technique has been described which would allow one to modify large genomic clones i.e. fosmids, at a specific target site with ease and this technique is known as recombineering (Copeland et al., 2001). This method has been employed by several groups (Dolphin and Hope, 2006; Sarov et al., 2006; Zhang et al., 2008;

Tursun et al., 2009) to modify large *C. elegans* genomic clones, for example by introducing a fluorescent reporter, and using the final construct as a translational reporter. The use of these fosmid-based translational reporters reduces the possibility that any of the regulatory elements for the GOI have been excluded in the final reporter construct. The different methods that have been described utilize different selection cassettes.

1.3 Genetic Engineering: Recombineering

Also known as genetic modification, genetic engineering is the alteration of genetic material to create recombinant DNA. Traditional genetic engineering involves the manipulation of genetic code to create new recombinant DNA that can be used to experimentally investigate novel genes and sequences. Traditional genetic engineering uses restriction enzymes and ligase, more commonly referred to as cloning, and relies on the presence of unique restriction sites to introduce DNA sequences. However this technique becomes limited as the size of the vector increases as the probability of locating an appropriate restriction site at the required location decreases.

In the late 1990s, techniques based on homologous recombination (HR) were developed to simplify and speed up genetic engineering in *E. coli* (Murphy, 1998; Zhang et al., 1998; Muyrers et al., 1999). Recombination-mediated genetic engineering, more commonly known as recombineering is the result of the techniques developed.

The inherent capacity for certain species to undergo HR has been harnessed to develop non-traditional approaches to mediate genetic modification of target sequences. Such a method was first demonstrated in *Saccharomyces cerevisiae* (*S. cerevisiae*) where it was shown that linear plasmid DNA could be used to either target genes on the genome or be used to clone genes directly from the genome (Orr-Weaver et al., 1983). HR with single stranded oligos (ssDNA) was also first shown in *S. cerevisiae* and was accomplished using synthetic oligos as short as 20 bp but the highest efficiency was seen with oligos 40 – 50 bp (Moerschell et al., 1988).

Recombineering in *E. coli* is based on HR, which is controlled by bacteriophage proteins, employing both *RecE* and *RecT* from Rac prophage or Red *exo*, *beta* and *gam*

from bacteriophage lambda (λ). *RecE*, from the Rac prophage, and Red *exo*, from the bacteriophage lambda, both encode a 5′ -> 3’ exonuclease that degrades linear dsDNA in a 5′ to 3’ direction to create 3’ overhangs. *RecT*, from the Rac prophage, and Red *beta*, from λ prophage, code for ssDNA annealing protein which bind to the 3’ overhangs (created by the 5′-3’ exonuclease) and mediates its pairing with its complementary target sequence. The defective λ bacteriophage also contains *gam*, which inhibits the host nucleases, RedBCD and SbcCD, protecting dsDNA substrates for recombineering in the *E. coli* cell. The promoters of both prophages are temperature sensitive and shifting the temperature of the culture induces expression of the genes required for HR.

The λ *gam* protects the linear DNA introduced into the cells (PCR-amplified selection cassette) from the RecBCD nuclease, which is responsible for degrading linear DNA. The λ Exo then binds to the double stranded DNA (dsDNA) and degrades one strand in a 5′-3’ direction whilst remaining bound to the other strand and preserving it creating 3’ overhangs (Reviewed by Sawitzke et al., 2007). The BETA protein binds to the 3’ overhangs, single stranded DNA (ssDNA), protecting the ssDNA, and the 3’ overhang ends, from nuclease attack. λ beta binding to the ssDNA improves the stability and directs the annealing of the complementary DNA strands (the ssDNA to the complementary sequence in the target DNA). The beta protein anneals the ssDNA to the complementary strand near the DNA replication fork to incorporate the “new” sequence into the target DNA and generate the recombinant (Ellis, et al., 2001).

1.3.1 Recombineering: basic principles

Recombineering is a genetic and molecular technique based on homologous recombination techniques to combine DNA sequences. This technique allows the insertion of a DNA sequence at a precise location within a target DNA construct. This can be performed in two ways, by positive selection or by counter-selection (Figure 6). If performed via positive selection, the DNA introduced would include an additional marker allowing the new recombinants to be selected. If recombineering is performed via counter-selection the recombinants are created in two steps. A recombineering cassette containing an antibiotic resistance gene is first introduced at the target site and in the second step is subsequently replaced with the required DNA sequence.

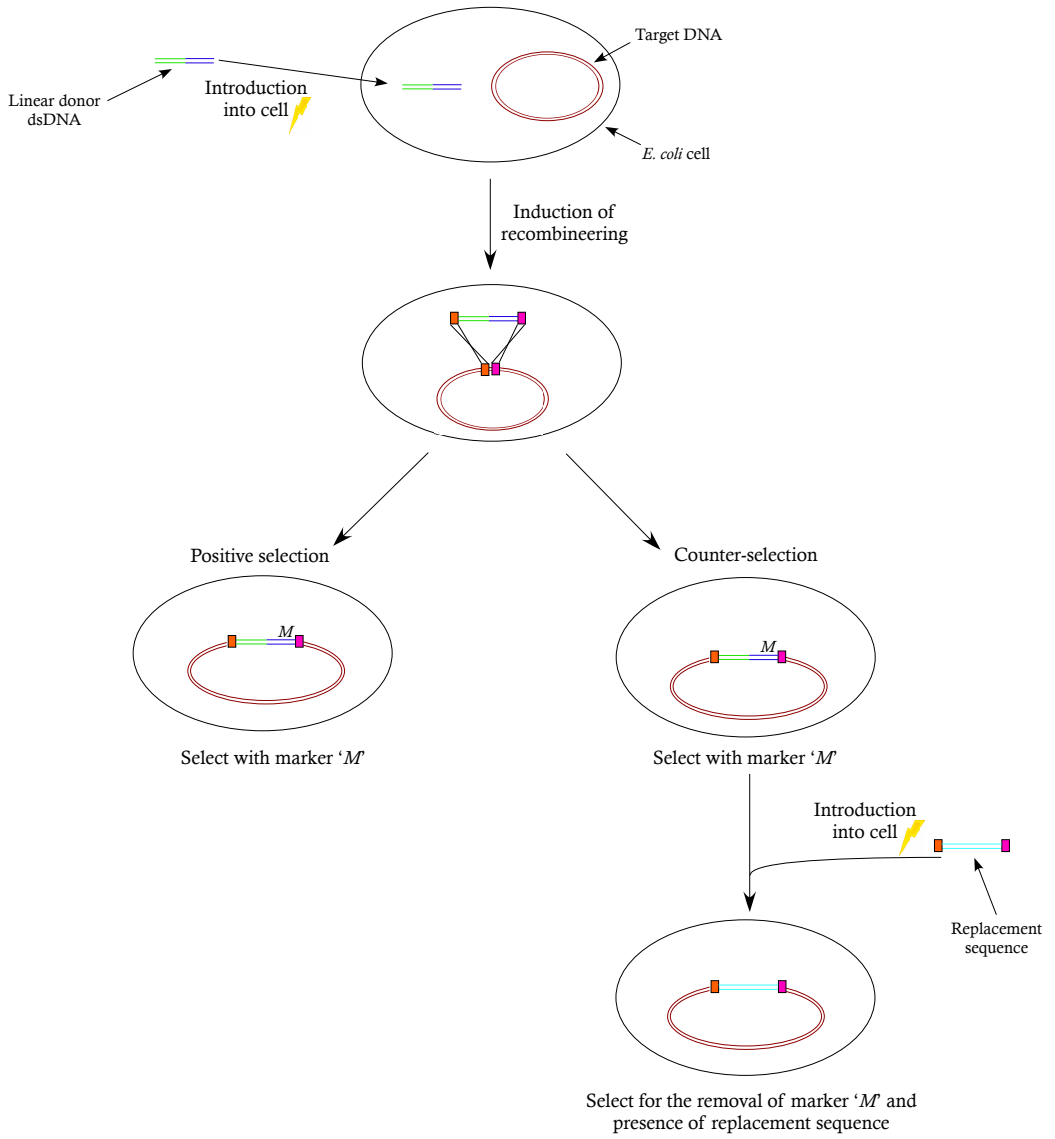

Figure 6. Recombineering: a schematic. Linear donor DNA, containing the coding sequence for a marker, is firstly introduced into *E. coli* cells containing the target DNA, via electroporation and cells are then induced for expression of EXO, BETA and GAM to allow recombineering to be performed. If performing recombineering only via positive selection, successful recombinants will be selected by the presence of the newly introduced marker. For counter-selection, the replacement sequence is introduced via electroporation into *E. coli* cells containing the recombinant and cells are induced for recombineering. Successful recombinants are identified by the removal of the marker.

1.3.2 Defective λ red prophage

The transcription of genes *exo*, *beta* and *gam*, on the defective λ red prophage, are controlled by a temperature sensitive repressor cI857. The repressor only dissociates at

42 °C to allow the transcription of *exo*, *beta* and *gam* and therefore all preparation work and incubations of *E. coli* cells when performing recombineering are carried out at 32° C (Figure 7). When expression of *exo*, *beta* and *gam* are required the *E. coli* cells are briefly heat-shocked to 42 °C before being rapidly cooled to halt transcription via binding of the repressor.

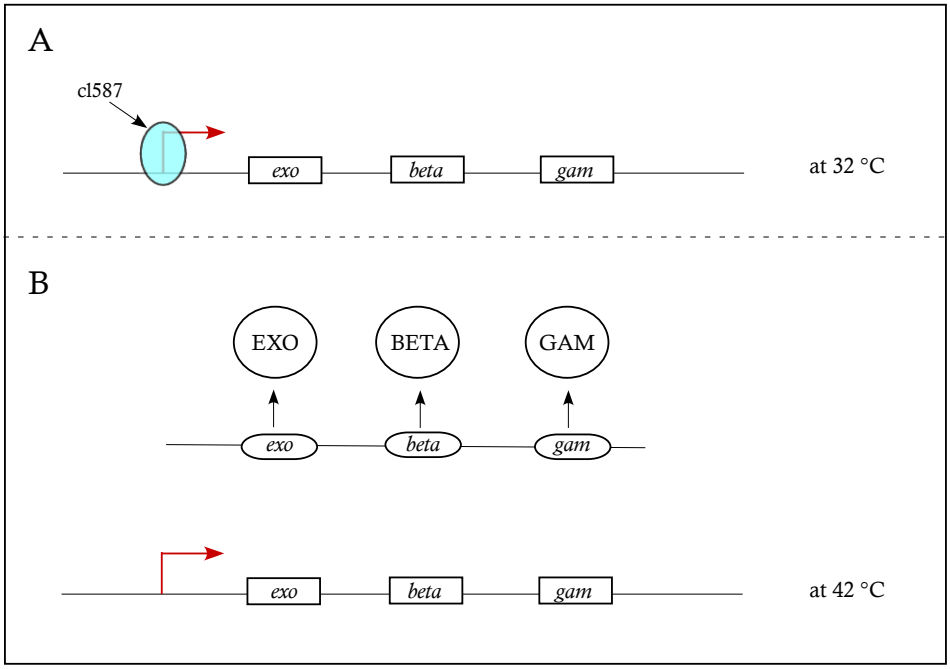

Figure 7. The defective λ red prophage. The above diagram shows the λ red prophage required for recombineering. When the *E. coli* cells are maintained at 32 °C, the temperature sensitive repressor (cl587) binds to the promoter region preventing transcription and translation of *exo*, *beta* and *gam* (Panel A). When the cells are briefly incubated at 42 °C, the temperature sensitive repressor (cl857) dissociates from the promoter allowing transient transcription and translation of *exo*, *beta* and *gam*.

1.3.3 Counter-selection recombineering

Counter-selection recombineering is a two step positive-negative selection approach and can be carried out using a antibiotic marker which can be used in both the positive and negative steps i.e. galK or can be carried out using a selection cassette that comprises two different markers, one for each selection step i.e the RT-Cassette (*rpsL – Tet(A)* counter-selection cassette). Counter-selection recombineering involves the PCR-amplification of a selection marker or cassette with oligonucleotides (ODNs) containing sequence homologous to the insertion site. These same homologous sequences are also included on the ODNs that will be used to PCR-amplify the

replacement sequence to ensure that the replacement sequence takes the exact place of the counter-selection cassette. As this technique relies on the homology sequences in the ODNs to target the first selection cassette to the insertion site, this method can be applied to manipulate a fosmid clone at any specific target site for example insert, delete or replace sequences.

Recombineering can be performed in any *E. coli* strain as long as the phage recombinases are available to be transiently expressed i.e. integrated into the *E. coli* genome or on a plasmid. The phage recombinases are contained within the defective λ red prophage as this methodology requires *exo*, *beta* and *gam* genes under the control of λ PL promoter (Figure 7). The plasmid prophage delivery system is more flexible as it can be transferred between strains, however the tighter controlled and coordinated expression of integrated prophage system was adopted as it is considered more controllable and efficient (Dolphin and Hope, 2006).

Recombineering can be performed via either a simple positive selection marker or as, adopted here, utilizing a single cassette for counter-selection. A positive selection marker would include an antibiotic resistance gene allowing recombinants to be selected via the presence of the new marker. There are many negative selection markers available, however only two have been reported in counter-selection recombineering. These are *rpsL*, in the RT-Cassette (Dolphin and Hope, 2006), and *galK* (galactokinase gene, Zhang et al., 2008).

The galK cassette is a selection cassette comprised one gene that can be used for both positive and negative selection steps. For the positive selection step, cells are plated onto agar plates containing galactose, which acts as the sole source of carbon, therefore only the *E. coli* cells which have successfully incorporated the selection marker will be able to utilize galactose and grow. In the negative selection step bacteria are spread onto, special minimal media, plates containing deoxy-galactose (DOG) which is metabolized by galK generating a toxic intermediate inhibiting *E. coli* growth. To simplify and streamline the recombineering protocol, several groups have generated 'recombineering toolkits' for the community, which both utilize the galK (Zhang et al., 2008; Tursun et al., 2009) selection marker.

The first set of galK resources to be published (Zhang et al., 2008) described the creation of constructs where the galK cassette is flanked by 50 bp of either gfp and/or

N-/C- terminal affinity tag sequence. This would reduce the number of primers required for PCR- amplification of both the selection cassette and the replacement sequence to one pair. The second toolkit (Tursun et al., 2009) generated a larger number of constructs and the galK cassette was introduced into the second intron of CFP, GFP, Venus and Mc-mCherry. As the second recombineering step relies on flp recombinase to recognize the FRT (Flp recognition target) sites (resulting in recombination between the FRT sites and removing the sequence between them), which flank the galK cassette and excise it out, leaving behind a 34 bp scar sequence in the second intron, to create what has been termed a virtually seamless insertion. Additional constructs were generated to allow dual labeling in one fosmid clone. A modified FRT sequence was used to flank the galK cassette to minimize any chances of any intramolecular recombination occurring between the scar sequences. Additional resources including affinity tags were also included in this toolkit.

1.3.3.1 RT-Cassette

The RT-Cassette is a dual selection cassette from pBAC-RT which was originally designed for use in BACs (single copy constucts) (Stavropoulos and Strathdee, 2001). The *tetA* and *rspL* genes are under the control of a ompF promoter (Figure 8). The *tetA* gene, isolated from TN10, confers tetracyline resistance. The protein product of tetA is an efflux pump protein, localized to the inner membrane, which transports magnesium-tetracycline complexes out of the cell and in exchange internalize extracellular protons. The *rpsL* gene confers streptomycin sensitivity and when within an *E. coli* strain which otherwise is streptomycin resistant, due to a mutated *rpsL* allele engineered into the bacterial genome, produces a dominant phenotype resulting in the *E. coli* strain conferring sensitivity to streptomycin (Stavropoulos and Strathdee, 2001). The endogenous ompF promoter directs expression of an outer membrane porin protein and is one of the most abundant proteins in the bacterial cell illustrating that ompF promoter is a strong promoter and is also osmotically regulated. The ompF has a high level of activity under low osmolarity conditions and thus the selection during the negative step is performed on no salt LB (NSLB) agar plates (Stavropoulos and Strathdee, 2001)

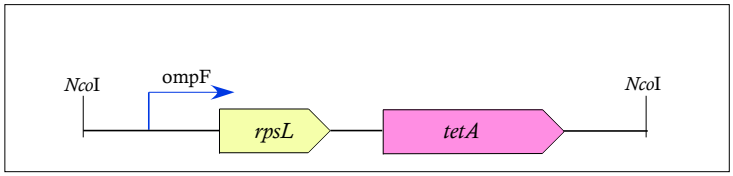

Figure 8. RT-Cassette: a schematic. The counter-selection cassette, the RT-Cassette, contains the *rpsL* and *tetA* genes, hence the name RT, with their expression controlled by the ompF promoter. The RT-Cassette can be isolated as a single *Nco*I fragment and used as a PCR template.

This cassette can be used for dual selection in a two-step recombineering protocol as previously described (Dolphin and Hope, 2006). The first step is positive selection as the presence of the *tetA* gene is selected for and the subsequent negative selection step is screening for the removal of the RT-Cassette and restoration of the streptomycin resistance due to the removal of the wildtype *rpsL* allele which exerts the dominant phenotype of streptomycin sensitivity over a mutant *rpsL* allele in the genome.

Aims and Objectives

Operons were first identified in *C. elegans* in 1993 (Spieth et al., 1993). Increasing experimental data (Huang et al., 2007; Allen et al., 2011) demonstrated operons within the *C. elegans* genome are not similar to the simple structures defined as bacterial operons. Approximately. 25% of the total operons in *C. elegans* have been classified as "hybrid-operons" as internal regulatory elements (also known as internal promoters), located in the upstream intergenic region for the gene within the operon, have been identified. In many instances these internal promoters were demonstrated to possess the capability of driving *gfp* expression in transcriptional reporters. Comparison of the GFP expression patterns from the "major upstream operon promoter" and "internal promoters" from the same operon revealed different expression patterns and in some cases the expression was observed in different tissues. Here, we intend to investigate the expression patterns of genes contained within operons in the context of the operon structure, by generating fosmid-based translational reporters, to determine if the genes within the selected operons have any relation in their expression i.e. expressed in the same tissue or at the same time and to further investigate if there are any discrepancies in the expression patterns observed. FP CDSs (excluding the stop codon of the FP) will be seamlessly recombineered 6 codons upstream from the stop codon of the target gene and expression pattern data for each gene within the operon investigated. The operons will then undergo detailed analysis to identify any potential promoter elements and their function dissected out via fosmid-based reporters. In order to facilitate this task, two resource sets will be created including some FPs novel to *C. elegans* research (mCerulean, mTFP1 and mCitrine) to make the generation of multiple fosmid-based translational reporters in parallel easier and quicker.

This project is split into two parts. In the first half, four FPs will be selected for multiple tagging, of up to four fluorescent reporters, in a single fosmid-based construct. The FP CDSs will be designed, with conveniently placed restriction enzyme sites to facilitate subcloning steps which will result in multiple variants for each FP chosen for multiple tagging, and CDSs codon-optimized for expression in *C. elegans* and the resulting sequences commercially synthesized. A subset of variants, generated via the subcloning steps, will be validated for future use as a reporter with and without TAP tags and nuclear localization signals to expand the utility of the resource set. An additional resource set, utilizing the codon-optimized synthesized FPs variants and the readily used by the worm community F-CFP, F-YFP and Mc-mCherry, will also be

generated to streamline the recombineering process when employing the RT-cassette when performing counter-selection recombineering and these two resource collections will be made available to the worm community, via addgene, as part of the funding requisite.

For the second part of the project, a selection of the recombineering resources will subsequently be utilized to investigate expression patterns of genes within two carefully selected operons via fosmid-based translational reporters. The genes within the two operons will be iteratively tagged with a FP reporter which will be seamlessly recombineered 6 codons before the genes stop codon within the fosmid clones. These two operons will also be interrogated via bioinformatics analysis to determine if a) they are conserved in the *Caenorhabditis* genus and b) to identify any conserved non-coding sequence(s) which could potentially indicate an internal promoter/regulatory element. If any conserved non-coding sequences are identified, their role will be dissected with fosmid-based translational reporters.

2

Materials and Methods

All materials and methods have been collated into a single chapter (Chapter 2) to minimize repetition. However relevant sections, relating to individual chapters, are indicated at the start of that chapter.

2.1 General molecular biology techniques

2.1.1 Plasmid and fosmid DNA isolation

Plasmid and Fosmid DNAs were isolated from cultures using a commercial isolation system (Sigma GenElute) following the manufacturer's protocol. Plasmid DNAs were obtained from 5 ml LB medium cultures supplemented with the appropriate antibiotics (appendix 2), grown at 37 °C with shaking at 220 rpm for 17 hours. Fosmid DNAs were obtained from a 10 ml LB medium culture supplemented with the appropriate antibiotic and Copy control fosmid auto-induction solution (Epicentre), grown at 32 °C with shaking at 150 rpm for 19 hours. All isolated DNA was eluted in 50 μl elution buffer. DNA concentrations were determined by gel electrophoresis, by visual comparison to a standard DNA ladder, and spectrophotometrically (NanoDrop 1000).

2.1.2 Agarose gel electrophoresis

DNA samples were separated by electrophoresis through agarose gels (0.8% to 2.0% w/v) with 1 x TAE running buffer (appendix 2) and visualized by ethidium bromide staining. Where separated DNAs were to be excised for subsequent purification, DNA was stained with SYBR Green and visualized with visible blue light (Dark Reader transilluminator, Clare Chemical) to eliminate the potential of UV-mediated DNA damage.

2.1.3 Restriction enzyme incubation

Plasmid and fosmid DNAs were digested with restriction enzymes from NEB (UK) according to the manufacturer's recommendations (unless otherwise stated).

2.1.4 DNA purification: gel slices and PCR products

PCR products and DNA restriction fragments, excised in an agarose gel slice, were purified with a commercial kit (Wizard SV gel and PCR clean-up system, Promega), performed according to the manufacturer's recommendations, and resulting DNA resuspended in nuclease-free water (50 μl).

2.1.5 Ligations

Ligations were set-up as follows: 50-100 ng of vector backbone was ligated with a 3 to 5 fold molar excess of insert fragment. The amount of insert DNA required was dependent upon the length of the fragment and was calculated using the following equation:

$$\frac{\text{Vector backbone (ng) x size of insert (kb)}}{\text{Size of vector backbone (kb)}} \quad \text{x} \quad \text{ratio of } \frac{\text{insert}}{\text{vector}} \quad = \text{ng of insert}$$

Ligation reaction mixes (10 μl) were catalyzed by incubation (15 °C, 16 hours) with T4 DNA ligase (3 units, Promega, UK).

2.1.6 Production of electrocompetent *E. coli* host cells

For each strain of *E. coli* a liquid culture (100 ml, SOB [- Mg^{2+}] (appendix 2) plus appropriate antibiotic(s)), inoculated with an aliquot (1 ml) from a small starter culture (5 ml, LB medium, 16 h at 220 r.p.m., 32 °C (MW005) or 37 °C (EPI300, DH5α)) was incubated, using the starter culture conditions, to an OD_{600} of 0.7. The culture was divided equally between two pre-cooled Falcon tubes, incubated on ice (20 min), pelleted by centrifugation (5,000 rpm, 4 °C), resuspended gently in ice-cold glycerol (50 ml, 10% v/v), re-pelleted, washed once more and the final pellet resuspended in ice-cold glycerol (500 μl) before being stored (-80 °C) in aliquots (100 μl).

2.1.7 Transformations

Aliquots (2 μl) from each ligation reaction were electroporated into electrocompetent DH5α cells (100 μl) using electroporation cuvettes (2mm gap, 2.3kv, Eppendorf Electroporator 2510). Cells were recovered in 1 ml SOC (-Mg^{2+}) medium (appendix 2) for 1 hour at either 32 °C (MW005) or 37 °C (DH5α or EPI300) with shaking (220 rpm). Finally aliquots (50 μl) of recovered cells were spread on LB agar plates, with appropriate antibiotics and incubated at either 32 °C (MW005) or 37 °C (DH5α or EPI300) for 18 hours.

2.1.8 DNA amplification: PCR

2.1.8.1 Colony screening PCR

Discreet colony cells were diluted into 100 μl ddH_2O and 1 μl used as template. The initial denaturation was at 94.0 °C for 3 mins to disrupt the cellular structures and expose the DNA template, followed by 30 cycles of 30 sec at 94.0 °C denaturation, 30 sec at 58.0 °C annealing followed by extension at 72.0 °C with time determined by product size (1 kb/minute). The PCR reaction (50 μl) consisted of 3 mM $MgCl_2$, 1x PCR buffer (Promega), 200 μM of each dNTP, 1 unit of Go-Taq polymerase (Promega) and 15 pmol of each forward and reverse primer.

2.1.8.2 PCR amplification of RT-Cassette/FP CDSs

PCR template (1 ng) was amplified using 1 unit of Phusion Polymerase (NEB), HF Phusion reaction buffer with 15 pmol of each forward and reverse ODN (ultramer grade, IDT) and 200 μM of each dNTP (NEB). The intial denaturation was at 98.0 °C for 30 seconds, followed by 30 cycles of 30 sec at 98.0 °C denaturation, 10 sec at 58.0 °C annealing followed by an extension time of 2 minutes at 72.0 °C (extension rate of 1 kb/30 secs). The concentrations of the PCR products were determined by gel electrophoresis, by visual comparison to a standard DNA ladder, and spectrophotometrically (NanoDrop 1000).

2.2 *Caenorhabditis elegans*

2.2.1 Maintenance

Wild-type (N2) *Caenorhabditis elegans* and transgenic strains were maintained as previously described by Brenner in 1974. Briefly nematodes were maintained on NGM (Nematode growth media) agar plates (appendix 2) seeded with OP50 (uracil deficient *Escherichia coli* strain) and maintained at 15 – 20 °C. OP50 was streaked onto LB plate, incubated (16 h, 37 °C) and a single discrete colony used to inoculate 50 ml LB medium (16 h, 37 °C). The miniculture of OP50 was used to seed 35 mm (100 μl) and 55 mm (200 μl) NGM plates. The seeded plates were incubated at 20 °C for 24 hours prior to use or storage at 4 °C.

2.2.2 Microinjection

Microinjections were carried out essentially as previously described by Mello *et al.* in 1991. Briefly, young gravid hermaphrodite N2 worms, adhered to agarose (2 % w/v) pads under halocarbon oil 700 (20 μl, Sigma), were injected, in a single gonadal arm, with DNA mixtures (plasmid: (50 – 100 ng/μl) or fosmid (5 – 10 ng/μl) mixed with co-marker pRF4 (*rol-6* marker, 100 ng/μl)) via glass injection needles (borosilicate glass capillaries, pulled using a Sutter puller, model P-97). The microinjection system comprised an inverted microscope (Leitz fluovert FU) plus microinjector (Eppendorf Femtojet, model 5247 v2, injection pressure (P*i*) 800 - 2000 hPa and compensation pressure (P*c*) of 175 hPa). Microinjected worms were recovered (20 μl M9 buffer, appendix 2), transferred onto freshly seeded NGM plates, incubated (16 h, 15 °C) and subsequently maintained at 20 °C.

2.3 Imaging

2.3.1 Microscope

As the fluorescent reporters will ultimately be made available to the worm community, and as, in addition, we had limited access to confocal microscopes, transgenic worms were visualized on an upright motorized epifluorescent microscope (Olympus BX-61), equipped with DIC optics, and images captured with a monochrome camera (Olympus F-View II). Resulting images were analyzed with proprietary software (Olympus CellSens Dimensions software v1.6).

2.3.2 Filter sets

Images of FP immunocaptured on agarose beads and worms transgenic for the different reporters utilized were captured with filter sets (Table 1).

Table 1. Filter sets utilized for capturing fluorescent signals from the different reporters used.

FP	Filter set	Excitation filter	Dichroic	Emission filter
F-CFP	CFP-2432A (Semrock)	438/24-25	458	483/32-25
mCerulean				
mTFP1				
GFP	N-41020 (Chroma)	480/20-20	495	510/20-20
F-YFP	YFP-2427A (Semrock)	500/24-25	520	542/27-25
mCitrine				
mCherry	mCherry-A (Semrock)	562/40-25	593	593/LP-25

2.3.3 Spectral separation

Filter sets chosen to image fluorescent signal for the reporters utilized were selected to capture emission over a wide range of wavelength to ensure that maximum signal from each FP is detected. Due to the significant emission overlap between mTFP1 and mCitrine (Figure 30), a computational approach was utilized to separate the individual channels using software developed on algorithms described by Zimmermman et al. (2005) and available in the spectral unmixing module present in the image analysis

package. Individual fluorophores imaged in worms transgenic for multiple reporter constructs, and that resulted in possible overlapping emission spectra, were spectrally separated following acquisition of respective reference images.

2.4 Derivation of FP coding sequence-containing constructs

2.4.1 Design considerations

The FP genes were designed using the software package Gene Designer (DNA 2.0, Villalobos et al., 2006). Gene Designer enables the user to construct an amino acid sequence and then generate the corresponding encoding DNA sequence by reverse translation. By specifying sequence requirements and/or restrictions the encoding sequence can be forced to include, or exclude, specific features such as, for example, specific restriction enzyme recognition sites. In addition, codon usage can be optimized for a particular species – in this case *C. elegans*. In this manner, each of the four FP amino acid sequences were compiled and coding sequences (CDSs) generated. All such CDSs were designed to be flanked by, and contain, unique restriction sites resulting in a modular assembly designed to facilitate all of the envisaged downstream sequence manipulations. Flanking *Not*I sites were included to allow facile excision of the complete coding sequence. A number of blunt-cutting restriction enzyme sites were included in the coding sequences to allow future insertion of artificial introns. All coding sequences, including those regions encoding the FPs and TAP-tag sections, were optimized for codon usage in *C. elegans* using a default threshold codon frequency of 12%.

The primary sequence of each FP was as published apart from the following differences. In both mCerulean and mCitrine, the A206K mutation, demonstrated to promote monomerization (Zacharias et al., 2002), was included. As regions of these proteins shared identical, or highly similar, amino acid sequences the reverse translation stage included a requirement that the four corresponding CDSs should, within the constraints of the other sequence specifications, also be different from each other. This design step was included to minimize any potential *in vivo* dimerization

and intra-molecular recombination when two or more sequences would be present in the same final construct. In addition, an N-TAP tag and C-TAP tag were also designed for either the N- or C- termini of each FP. Both versions of the TAP tags were designed to fuse to the FP via a flexible linker [$(G_4S)_3$] followed by the internal tag Strep Tag II [WSHPQFEK], a short linker [G_4S], a PreScission protease cleavage site [LEVLFQGP], a second short linker [G_4S] and the external S-Tag [KETAAAKFERQHMDS]. Furthermore, two sequences of SV40 nuclear localization signal (2x NLS) were also designed for the C-TAP tag variant and all coding sequences contained two stop codons (TAA-TAG).

2.4.2 Commercial synthesis

Following their design the CDSs were synthesized commercially and provided in proprietary plasmid vectors. Sequences encoding C-TAP tag::2x NLS and N-TAP tag::mTFP1 (Figure 9, Panels A and B) were synthesized (Gene Oracle) and supplied as the respective plasmids pNH001 and pNH002 based in the common restriction enzyme site-depleted vector pGOv5. mCitrine, mCherry and mCerulean coding sequences (Figure 9, Panels C, D and E) were synthesized (Mr Gene) and supplied as the respective plasmids pNH003, pNH004 and pNH005 based in vector pMK.

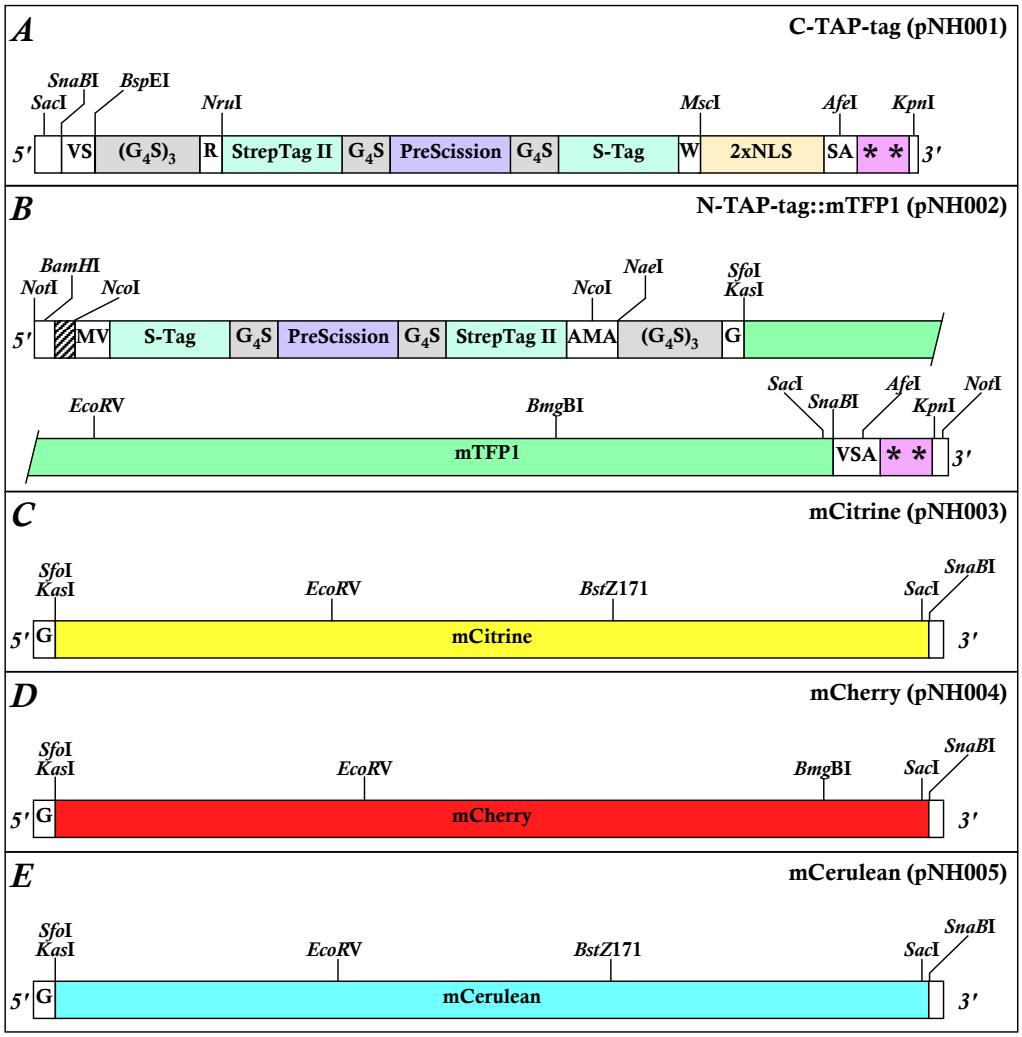

Figure 9. Design and synthesis of FP and TAP-tag CDSs. Sequences encoding a C-TAP tag (panel A), an N-TAP tag fused via a large linker $(G_4S)_3$ to mTFP1 (panel B), mCitrine (panel C), mCherry (panel D), and mCerulean (panel E) were designed with the synthetic gene design tool GeneDesigner (DNA 2.0) and commercially synthesized. Both C- and N- terminally located TAP-tags comprised S-Tag (external) and Strep Tag II (internal) affinity tags separated by a PreScission protease site. A large linker $(G_4S)_3$ was included to separate the TAP-tag from the FP and two small linkers (G_4S) to separate the individual components in the TAP-tag. The C-TAP tag included two stop codons (TAA-TAG, two asterisks) and a pair of nuclear localization signal sequences (2x NLS). The N-TAP-tag included the translation initiation signal aaaaccATG (hatched box in Panel B). Synthesized C-terminal TAP-tag and N-terminal TAP-tag::mTFP1 modules resulted in, respectively, plasmids pNH001 (Panel A) and pNH002 (Panel B). mCitrine, mCherry and mCerulean CDS units, each containing two unique blunt-cutting restriction enzyme sites to facilitate subsequent artificial intron insertion, were synthesized generating, respectively, plasmids pNH003 (Panel C), pNH004 (Panel D) and pNH005 (Panel E). In all sequences the inclusion of unique restriction enzyme sites, introduced to facilitate down- stream sub-cloning and construct manipulation, occasionally required the introduction of one or more additional amino- acid residues absent from the native sequences (open boxes with single letter residues).

2.4.3 Derivation of pGOv5-based FP-encoding sub-clones.

Briefly the sequences encoding each FP contained within pNH003, pNH004 and pNH005 were excised using *Kas*I and *Sac*I and ligated into the *Kas*I-*Sac*I purified backbone of pNH002 to generate pNH006, pNH007 and pNH008 (Figure 10, Panel A). The 48 bp large linker [$(G_4S)_3$] was released from pNH002 and pNH006 - pNH008 upon incubation with Turbo *Nae*I (Promega) and *Sfo*I and the resulting backbone ligated to create plasmids pNH009 – pNH012 (Figure 10, Panel B). Constructs pNH009 – pNH012 were incubated with *Nco*I (NEB) to excise the in-frame N-TAP tag. The backbone was gel purified and ligated to make constructs pNH013 – pNH016 (Figure 10, Panel C). The C-TAP tag and 2xNLS was released from pNH001 upon incubation with *Sac*I and *Kpn*I and the 250 bp fragment was gel purified. The purified fragment was subsequently ligated into the *Sac*I – *Kpn*I backbone of constructs pNH013 – pNH016 to generate constructs pNH017 – pNH020 (Figure 10, Panel D). The 2x NLS was excised from constructs pNH017 - pNH020 by incubating with *Msc*I and *Afe*I and ligation of the backbone resulted in constructs pNH026 - pNH029 respectively. The C-TAP tag was removed from constructs pNH017 - pNH020 upon excision with *Nru*I and *Msc*I and the resulting backbone was rejoined to create plasmids pNH030 – pNH033 (Figure 10, Panel E).

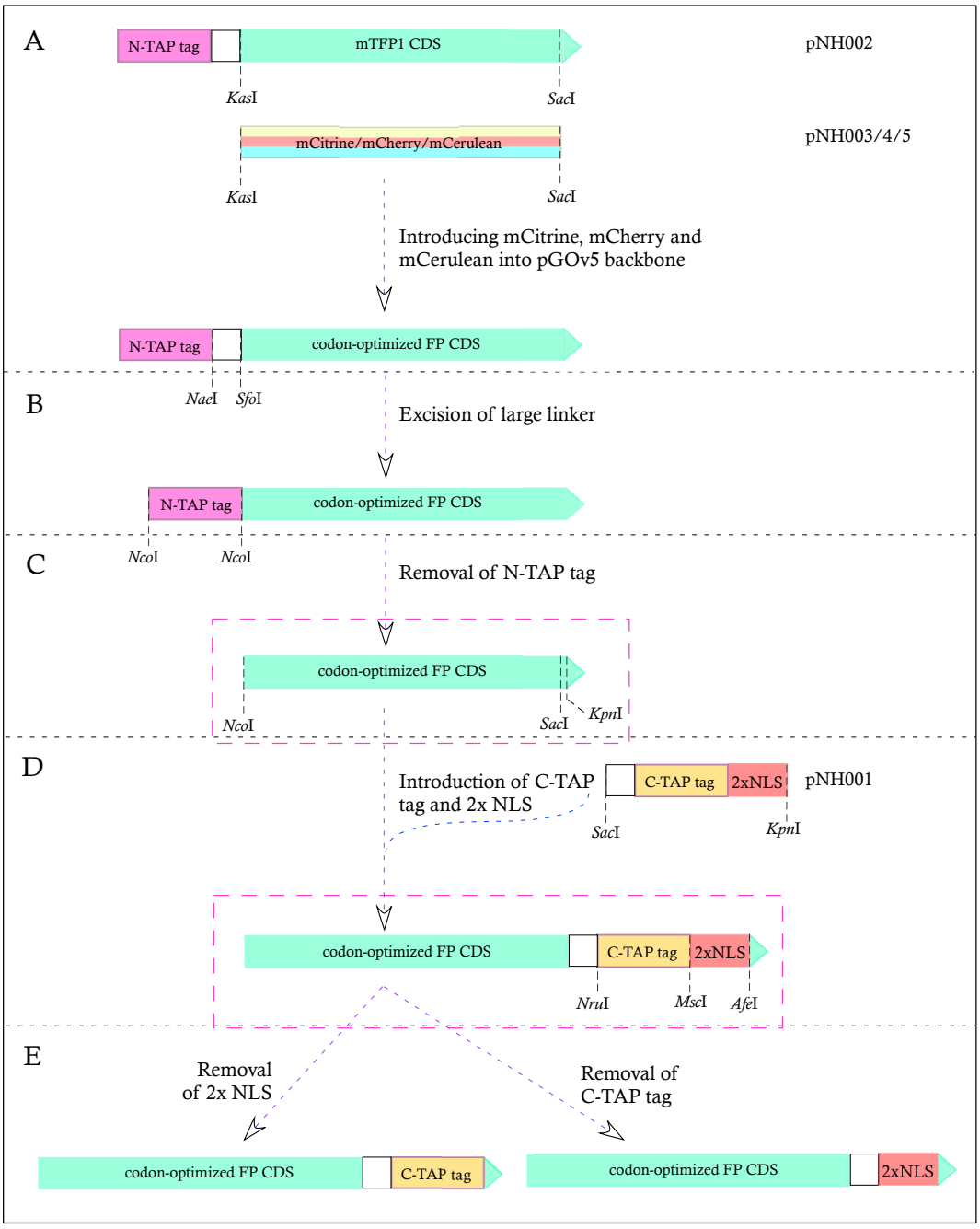

Figure 10. Schematic representation of the subcloning workflow. mCitrine, mCherry and mCerulean, from pNH003 – pNH005, were excised with *Kas*I-*Sac*I and ligated into the *Kas*I-*Sac*I backbone of pNH002 (Panel A). The large linker was removed upon incubation with *Nae*I and *Sfo*I with the resulting backbone ligated (Panel B). The N-TAP tag was excised with *Nco*I and the backbone closed (Panel C). The *C-TAP tag::2x NLS* was introduced into the FP CDSs by digesting pNH013 - pNH016 and pNH001 with *Sac*I-*Kpn*I and ligating the resulting fragment from pNH001 into the FP CDSs (Panel D). From these constructs, either the *2x NLS* or the *C-TAP tag* were removed upon incubation with *Nru*I-*Msc*I and *Msc*I-*Afe*I respectively (Panel E). The constructs enclosed within the pink dash-line box were used for *in vivo* validation.

2.4.4 Introduction of artificial introns

Artificial introns were introduced via iterative rounds of cloning, with the first round to introduce the first artificial introns and the next round to introduce the second set of introns (Figure 11). Short, blunt-ended artificial intron sequences were generated by annealing (95 °C for 15 min, cooling to room temp. in 1x T4 DNA ligase buffer (Promega)) pairs of 5′ phosphorylated ODNs (Table 1), via overlapping 3' ends, and filling-in (1U T4 DNA polymerase (NEB) in the presence of dNTPs (2 mM)). Introns B, α, β and γ (Table 2) were introduced into linearized (*Eco*RV) and dephosphorylated (1U CIAP (Promega), 37 °C, 1h) pNH013 - pNH016 generating the single intron-containing derivatives pNH078 – pNH081, respectively (Figure 11, Panel A). Introns C and E were introduced into *Bmg*BI linearized pNH078 and pNH080 to create, respectively, pNH082 and pNH084 and introns D and F were introduced into *Bst*Z171 linearized pNH079 and pNH081 to generate, respectively, pNH083 and pNH085 (Figure 11, Panel B).

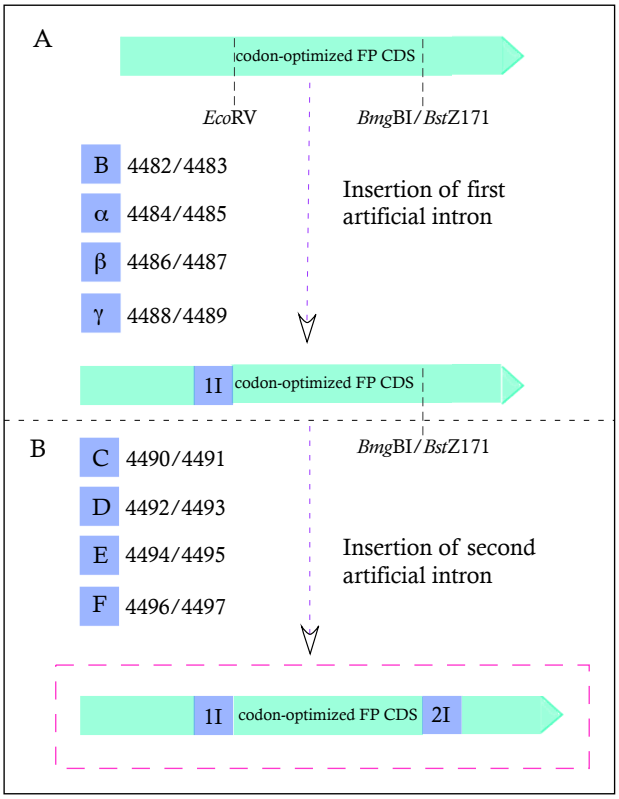

Figure 11. Intron insertion: schematic. Each codon-optimized FP CDS containing construct was linearized by digestion with *Eco*RV and subsequently dephosphorylated. Artificial intron sequences B, α, β and γ were inserted into mTFP1, mCitrine, mCherry and mCerulean CDSs, respectively to generate pNH078 – pNH081 (Panel A). These constructs were subsequently linearized with either *Bmg*BI or *Bst*Z171 and artificial intron sequences C, D, E and F were inserted into mTFP1, mCitrine, mCherry and mCerulean CDSs, respectively (Panel B). The

resulting construct containing two artificial introns, enclosed by the pink box, was utilized for *in vivo* validation.

Table 2. Artificial Intron Sequences

ODNs	Sequence (5′-3′)	Intron	Sequence (5′-3′)
4482	GTAAGTTTAAACAGATCCATACTAACTAACTTGTTC		
4483	CTGAAAATTATGTCAGAACAAGTTAGTTAGTATGGA	B	GTAAGTTTAAACAGATCCATACTAACTAACTTGTTCTGACATAATTTTCAG
4484	GTAAGTTTAAACATATATATACTAACTAACCCTGAT		
4485	CTGAAAATTTAAATAATCAGGGTTAGTTAGTATATA	α	GTAAGTTTAAACATATATATACTAACTAACCCTGATTATTTAAATTTTCAG
4486	GTAAGTTTAAACAGTTCGGTACTAACTAACCATACA		
4487	CTGAAAATTTAAATATGTATGGTTAGTTAGTACCGA	β	GTAAGTTTAAACAGTTCGGTACTAACTAACCATACATATTTAAATTTTCAG
4488	GTAAGTTTAAACATGATTTTACTAACTAACTAATCT		
4489	CTGAAAATTTAAATCAGATTAGTTAGTTAGTAAAAT	γ	GTAAGTTTAAACATGATTTTACTAACTAACTAATCTGATTTAAATTTTCAG
4490	GTAAGTTTAATTAAGTTGATACTAACTAACAAAGAT		
4491	CTGAAAATTAATCAGATCTTTGTTAGTTAGTATCAA	C	GTAAGTTTAATTAAGTTGATACTAACTAACAAAGATCTGATTAATTTTCAG
4492	GTAAGTTTAAACAGGATCTTACTAACTAACATGCTA		
4493	CTGAAAATTCAGTGTTAGCATGTTAGTTAGTAAGAT	D	GTAAGTTTAAACAGGATCTTACTAACTAACATGCTAACACTGAATTTTCAG
4494	GTAAGTTTAAACTATTCGTTACTAACTAACTTTAAA		
4495	CTGAAAATTTAAATGTTTAAAGTTAGTTAGTAACGA	E	GTAAGTTTAAACTATTCGTTACTAACTAACTTTAAACATTTAAATTTTCAG
4496	GTAAGTTTAAACAGTTGAATACTAACTAACGGAGAT		
4497	CTGAAAATTTCAAAGATCTCCGTTAGTTAGTATTCA	F	GTAAGTTTAAACAGTTGAATACTAACTAACGGAGATCTTTGAAATTTTCAG

2.4.5 Building *myo-3*^{PROM}-driven transcriptional reporters.

The *FP::C-TAP-tag::2xNLS* CDSs from pNH017 (mTFP1), pNH018 (mCitrine), pNH019 (mCherry) and pNH020 (mCerulean) were excised with *Bam*HI and *Kpn*I and ligated into the *Bam*HI-*Kpn*I backbone of pPD95.86 (Fire et al, 1990) to generate, respectively, pNH022 - pNH025. The FP CDSs from pNH013 - pNH016 were released as *Bam*HI-*Kpn*I fragments and inserted into the gel-purified *Bam*HI-*Kpn*I backbone of pPD95.86 to generate constructs pNH047, pNH048, pNH035 and pNH049. The dual-intron containing FP CDSs were excised, with *Bam*HI and *Kpn*I and introduced into the backbone of pPD95.86 to generate, respectively, pNH086, pNH087, pNH088 and pNH089 (Figure 12). To serve as controls, the F-CFP, F-YFP and Mc-mCherry CDSs were PCR-amplified from pPD136.61, pPD136.64 (Fire et al., 1990) and pAA64 (McNally et al., 2006) with *Bam*HI-containing forward and *Kpn*I-containing reverse ODNs and cloned into their respective RE sites in pPD95.86 generating pNH074, pNH076 and pNH077 (Figure 12, page 68 and Table 4).

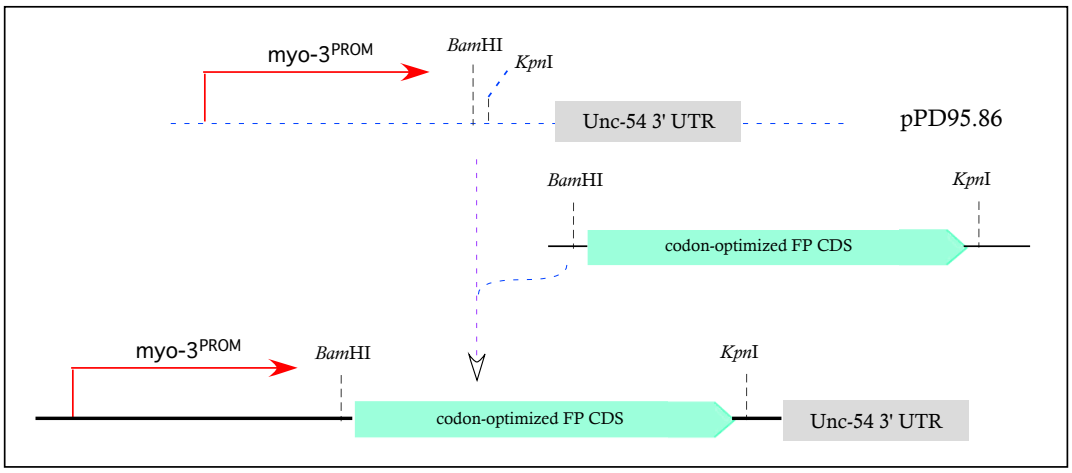

Figure 12. Building *myo-3*^{PROM} constructs: schematic. The FP CDSs, *FP*, *FP containing two artificial introns* and *FP::C-TAP-tag::2xNLS*, for mCerulean, mTFP1, mCitrine and mCherry were excised with *Bam*HI and *Kpn*I from the respective constructs and ligated into the *Bam*HI-*Kpn*I backbone of pPD95.86.

2.4.6 Immunocapture of *in vitro* transcription-translation products

Templates for *in vitro* transcription-translation, with a T7 promoter, were generated by PCR-amplifying codon-optimized FP CDSs (using constructs pNH013 – pNH016 as templates, 1 ng) with ODNs 12036/12037 (Table 4). As positive controls, equivalent templates were generated of eGFP (Clontech, ODNs 12038/12039 (Table 4)), eGFP CDS codon-optimized for expression in *C. elegans* (pMG001 (unpublished), ODNs 12040/12041 (Table 4)) and TagRFP-T (Shaner et al., 2008) CDS codon-optimized for expression in Sf9 cells (pMW039, ODNs 12042/12043 (Table 4)). An aliquot from each PCR reaction (5 µl) was then transcribed and translated (TNT T7 Quick PCR kit, Promega) following the manufacturer's protocol.

Aliquots (5 µl) from each completed *in vitro* transcription-translation reaction were incubated (16 hr, 4 °C on a rotary mixer) with Protein-G agarose beads (1 µl, Millipore) in buffer (0.2% Triton-X 100 in PBS) and an aliquot (0.25 µl) of the corresponding polyclonal antibody (Living colours full-length polyclonal antibody (Clontech) for mCerulean, mTFP1, eGFP and mCitrine, RFP/DsRed polyclonal antibody (MBL) for mCherry and polyclonal anti-tRFP antibody (Evrogen) for TagRFP-T). Beads were gently pelleted, washed three times with buffer, resuspended in 100 µl and either imaged or stored at 4 °C.

2.4.7 Affinity-capture: S-Tag

Aliquots (5 μl) from each completed *in vitro* transcription-translation reaction incubated (16 hr, 4 °C on a rotary mixer) with S-protein coated agarose beads (1 μl, Novagen) in buffer (0.5 ml Triton-X 100 in PBS (0.2 % v/v)). Beads were pelleted gently, washed three times with buffer, re-suspended (100 μl) and either imaged immediately or stored at 4 °C.

2.4.8 Affinity-capture: PreScission and Strep-TagII

Aliquots (5 μl) of mTFP1 from each completed *in vitro* transcription-translation reaction were incubated with protease (Prescission (GE Healthcare) or HRV 3C (Merck), 10 U each, 4 °C, 20 h). Protein samples were mixed with Laemmli loading dye (Bio-Rad) in a 1:1 ratio and separated via acrylamide (12 % w/v) gel electrophoresis under denaturing conditions (mini Protean system, Bio-Rad) before being transferred into a 0.2 μm PVDF membrane (Bio-Rad Trans-Blot turbo transfer system, Bio-Rad) The membrane was washed twice in 15 ml PBST (0.2 % Tween 20 in PBS, 10 min) prior to blocking in 5 % skimmed milk (in PBST, 2 hr, room temp with gentle shaking). Two subsequent 15 min wash steps were performed before incubating the membrane with primary antibody (1:1000 antibody in 5 % BSA (Sigma) in PBST, 16 hr, 4 °C with shaking). The membrane was washed (3x, PBST, 5 mins) before incubation with the secondary antibody, polyclonal goat anti-mouse HRP (Dako) (PBST, 2 hr, room temp with shaking). Three final 5 min wash steps of PBST were performed before employing ECL (enhanced chemiluminescence, Thermo Scientific) detection methods as per manufacturers protocol. The intensity of the HRP reaction was visualized on Amersham Hyperfilm ECL (GE Healthcare).

2.5 Recombineering

2.5.1 Two-step counter-selection recombineering

All recombineering experiments during this project was performed in *E. coli* strain MW005 (genotype; F- *mcrA* Δ(*mrr- hsd*RMS-*mcr*BC) Φ80d*lac*Z M15 Δ*lac*X74 *deo*R *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*) 7649 *gal*U *gal*K *rsp*L *nup*G [λ*cI*857 (*cro-bio*A) < >*ara*C-PBADt*rfA*]; Westenberg et al., 2010), which permits both λ Red-mediated recombineering and copy-number induction of pCC1FOS-based fosmid clones. Positive selection of successful recombinants was via the RT-Cassette tetracycline resistance (Tc, 5 μg/ml) and chloramphenicol resistance encoded by the pCC1FOS backbone of the fosmid clones (Cm, 10 μg/ml) and on occasion (as indicated in the methods) Ampicillin (Amp 50 μg/ml). Negative selection, when selecting for the replacement of the RT-Cassette with the desired sequence, was achieved with Streptomycin (Sm, 500 μg/ml) and Cm (10 μg/ml) and at the start of the project was essentially performed as described (Dolphin and Hope, 2006).

Access to the resources created subsequently lead to the development of a more streamlined recombineering protocol (Figure 13) requiring only a single PCR to generate the initial RT-Cassette, smaller culture volumes and shorter incubation times and no requirement to sequence final recombinants. A liquid culture, 5 ml SOB (-Mg^{2+}), was inoculated with an aliquot (1 in 10 v/v) of an overnight mini-culture, containing the target fosmid clone, and incubated on (2 hr, 32 °C). The required number of aliquots (1 ml) were transferred into new micro-tubes, the cells were pelleted, resuspended (1 ml pre-warmed (45 °C) SOB (-Mg^{2+})) and incubated in a shaking incubator (45 °C, 5 mins at 100 r.p.m, Labnet VorTemp). Each tube was chilled on ice (10 mins), centrifugated (5,000 r.p.m, 5 mins at 4 °C), and washed twice (1 ml ice-cold ddH_2O) and finally resuspended in 50 μl ice-cold ddH_2O. The cells were electroporated with 500 ng RT-Cassette insert (for positive selection), recovered (3 hr, 32 °C with shaking) in SOC (-Mg^{2+}) medium, and colonies positive for the insertion of the RT-Cassette were selected for on LB agar (Cm and Tc, after incubation for 36 – 48 hr at 32 °C) and identified upon colony PCR screening and subsequent restriction digest analysis. The RT-Cassette was subsequently replaced, via negative-selection recombineering in MW005 cells, as prepared above, with the appropriate gel-purified *Not*I-*Not*I fragment (50 ng) excised from the respective replacement construct (Table

7). Desired recombinant-containing colonies were selected on No Salt (NS)-LB agar (appendix 2) (Sm, Cm, 32 C, 36-48 hrs). Propagation of final recombinants in liquid culture, induction of copy-control number ('CopyControl', Epicentre), and fosmid DNA isolation were performed (Dolphin and Hope, 2006) and the fidelity of final constructs was determined by RE analysis.

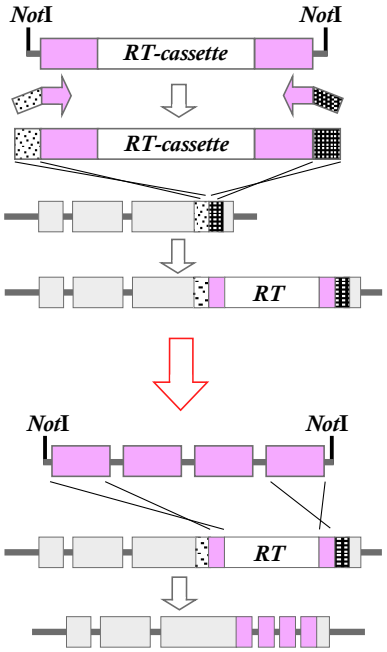

Figure 13. Streamlined counter-selection recombineering protocol: schematic. A *rps*L-*tet*A (RT) counter-selection cassette, PCR-amplified from a *Not*I restriction fragment excised from the appropriate pCC1FOS-based construct, flanked by approximately. 200 bp from the 5′ and 3' ends of the fluorescent protein reporter to be inserted (purple boxes) and terminal 50 bp homology arms (stippled boxes), is recombineered into the insertion site within the target gene *via* tetracycline selection. In the replacement step a *Not*I fragment, containing the fluorescent protein coding sequence, replaces the inserted RT-Cassette, *via* streptomycin selection creating an in-frame *goi::fp* fusion.

2.5.2 Generation of pNH034: a pCC1FOS-based RT-Cassette-containing cassette

To introduce the RT-Cassette into pCC1FOS, two discrete regions of the pGOv5 backbone were PCR-amplified from pNH002 (1 ng). One region was generated using ODNs 12015/12016 (Table 4) (600 bp product) and comprised the 3' end of the *β-lactamase* gene flanked by a 50 bp sequence identical to a region in pCC1FOS and another 50 bp equivalent to the 5′ region of the RT-Cassette sequence. The second

region, generated using ODNs 12017/12018 (Table 4), was a 200 bp region also found in pCC1FOS and a 50 bp sequence identical to the 3' end of the RT-Cassette. These two PCR products were fused with the RT-Cassette via a 3-way overlap PCR (Horton et al., 1989). The resulting purified product (500 ng) was introduced into pCC1FOS, replacing the *lox*P and *Lac*Z regions to generate pNH034 (Figure 14).

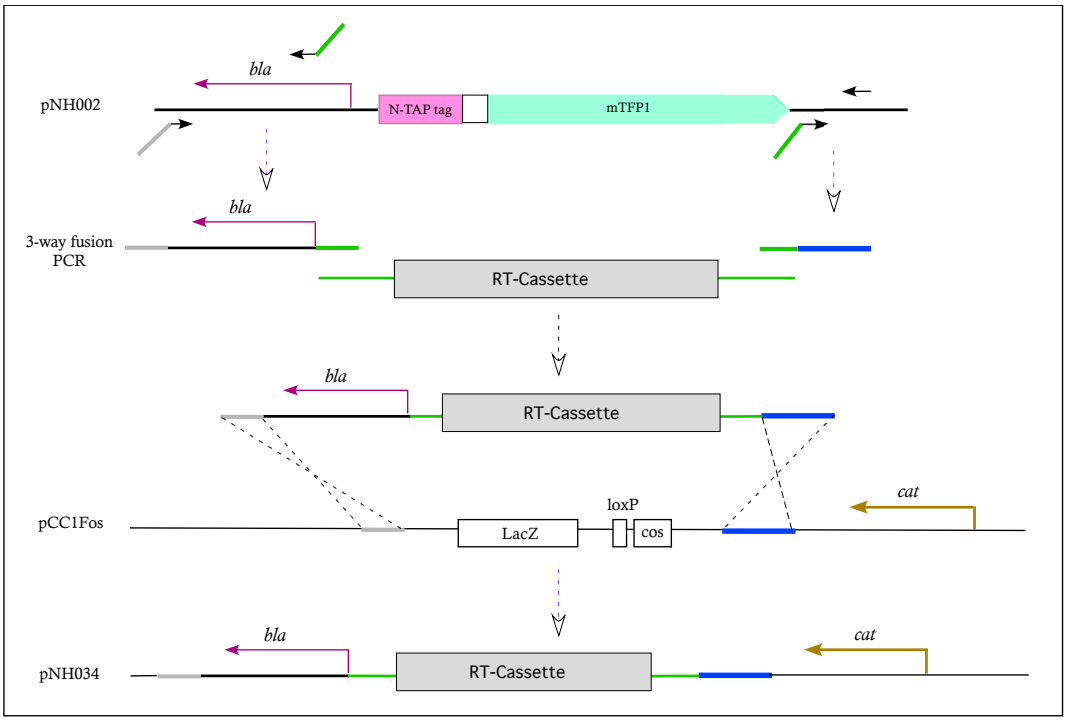

Figure 14. pNH034 generation: schematic. Two regions in pGOv5 were PCR amplified and the resulting amplicons, with sequence overlapping the RT-Cassette-containing sequence, were subsequently utilized in a 3-way fusion PCR, with the RT-Cassette, to create a PCR product where the RT-Cassette is flanked by sequence homologous to two discrete regions in the pCC1FOS backbone. The resulting product, of the RT-Cassette and the 3' end of the *bla* gene, was introduced, via positive selection recombineering, into pCC1FOS replacing the *Lac*Z, *lox*-P and the cos sites to generate pNH034.

2.5.3 Generation of pNH034-based sub-clones.

The RT-Cassette in construct pNH034 was subsequently replaced, via negative selection recombineering, with FP cassettes from the pGOv5-based constructs, excised with *Fsp*I and *Lgu*I, to generate a collection of constructs in which the RT-Cassette had been replaced by a different *Not*I-flanked FP CDS (Figure 15). As the full-length *bla*

gene was reconstructed during RT-Cassette replacement, the successfully created recombinants were selected with Ampicillin. The *Fsp*I-*Lgu*I fragments from constructs pNH002, pNH006 – pNH008, pNH013 - pNH020, and pNH026 - pNH029 were recombineered into pNH034 to generate pNH058 - pNH061, pNH054 - pNH057, pNH062 - pNH065, and pNH090 - pNH093 respectively.

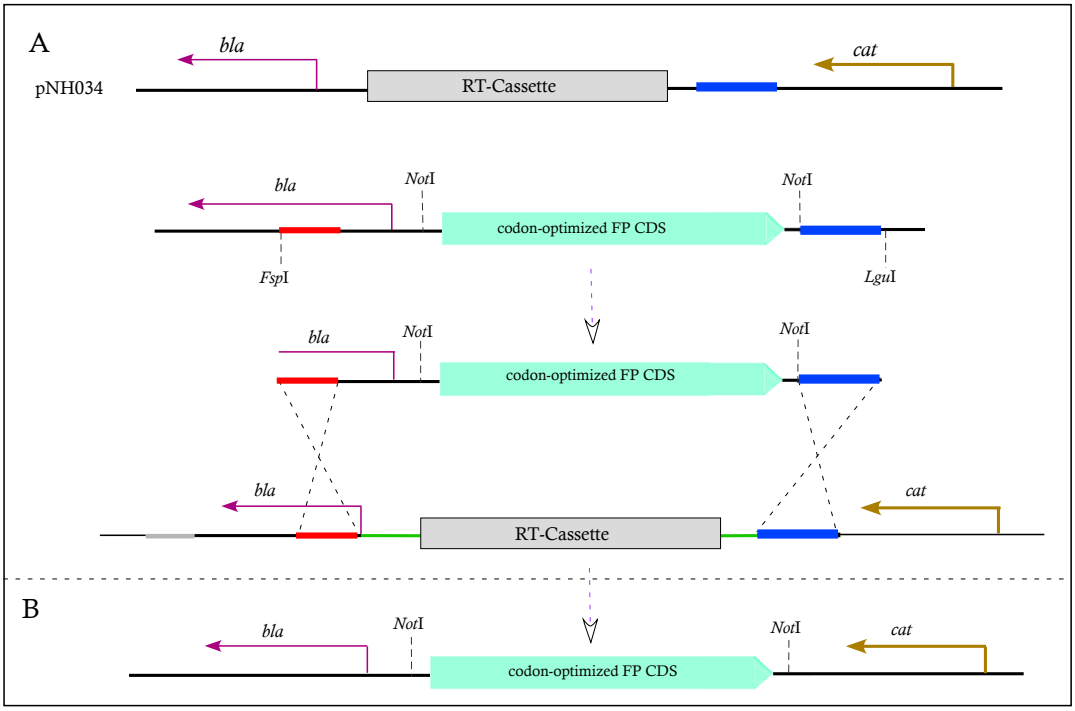

Figure 15. Generation of pNH034-based sub-clones. The pGOv5 based-constructs were digested with *Fsp*I and *Lgu*I with the resulting 2 kb fragment gel purified and via positive selection recombineering introduced into pNH034 (Panel A). The resulting constructs had a functional *bla* with the FP CDS flanked by *Not*I sites (Panel B).

The RT-Cassette was PCR-amplified (ODNs 12024/12025, Table 4) and the column purified amplicons were centrally introduced by recombineering into the FP gene sequence, leaving approximately. 200 bp of FP gene sequence flanking the RT-Cassette, to generate the corresponding RT-Cassette containing constructs (Figure 17). To differentiate between the RT-Cassette and the RT-Cassette flanked by FP CDS, the latter will be referred to as the FP-RT-FP Cassette (Figure 16).

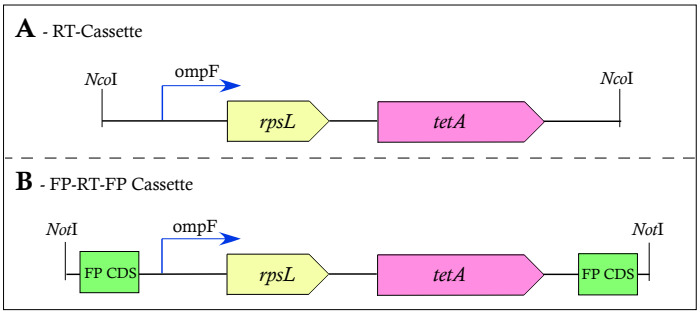

Figure 16. RT-Cassette and FP-RT-FP Cassette: schematic. A schematic to illustrate the difference between the RT-Cassette (Panel A) and FP-RT-FP Cassette (Panel B).

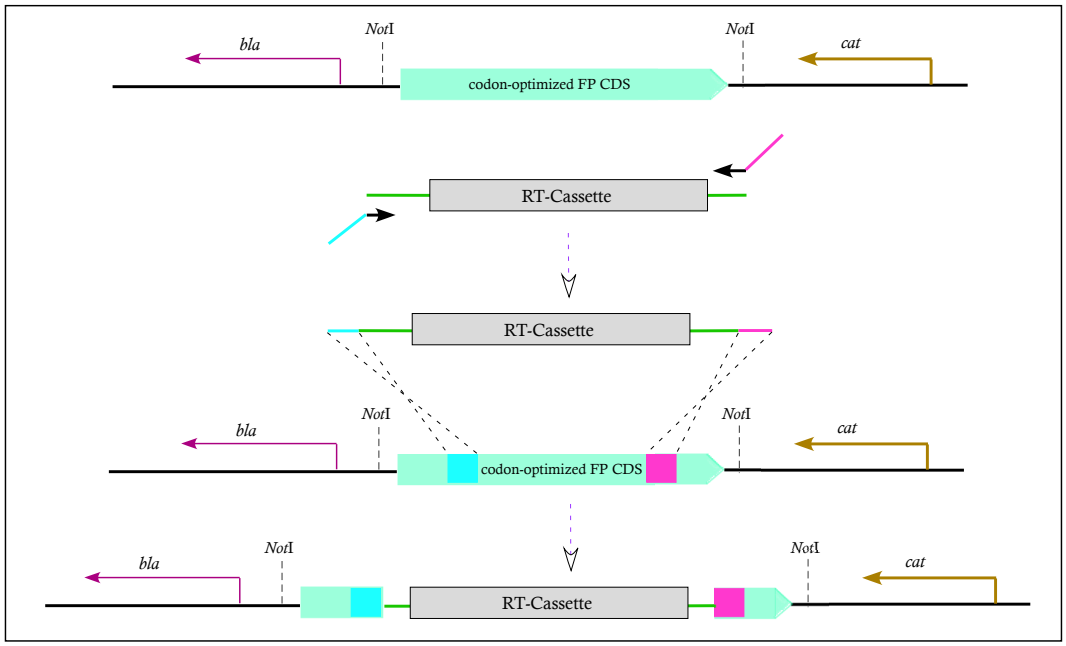

Figure 17. Schematic representation of introducing the RT-Cassette centrally into the FP CDS. The RT-Cassette was PCR amplified with ODNs containing homology arms to introduce the RT-Cassette into the middle of the FP CDS, with approximately. 200 bp of FP CDS flanking the RT-Cassette. The RT-Cassette was introduced, via positive selection, into the corresponding constructs to generate the final set of constructs.

2.5.4 Introducing F-CFP, F-GFP, F-YFP and Mc-mCherry CDSs into pNH034

The RT-Cassette was PCR-amplified (ODNs 12024/12025, Table 4) and the purified product introduced into pNH043 *via* positive selection recombineering to create the intermediate pNH038. Next F-CFP, F-GFP, F-YFP CDSs, from constructs

pPD136.61, pPD95.77 and pPD136.64 using ODNs 12028/12029 (Table 4), and Mc-mCherry CDSs, from pAA64 using ODNs 12026/12027 (Table 4), were PCR amplified. The purified PCR products subsequently replaced the RT-Cassette in pNH038 (figure 18, Panel A, via negative selection recombineering, to generate pNH039 – pNH042 (Figure 18, Panel B). The RT-Cassette was amplified with ODNs (12030/12031 for F-FPs and 12032/12033 for Mc-mCherry, Table 4) that were designed to seamlessly insert the RT-Cassette centrally in the FP CDS, with approximately. 200 bp of FP gene sequence flanking the RT-Cassette. The purified RT-Cassettes were recombineered into constructs pNH039 - pNH042 (via positive selection recombineering) to generate constructs pNH050 – pNH053 (Figure 18, Panel C).

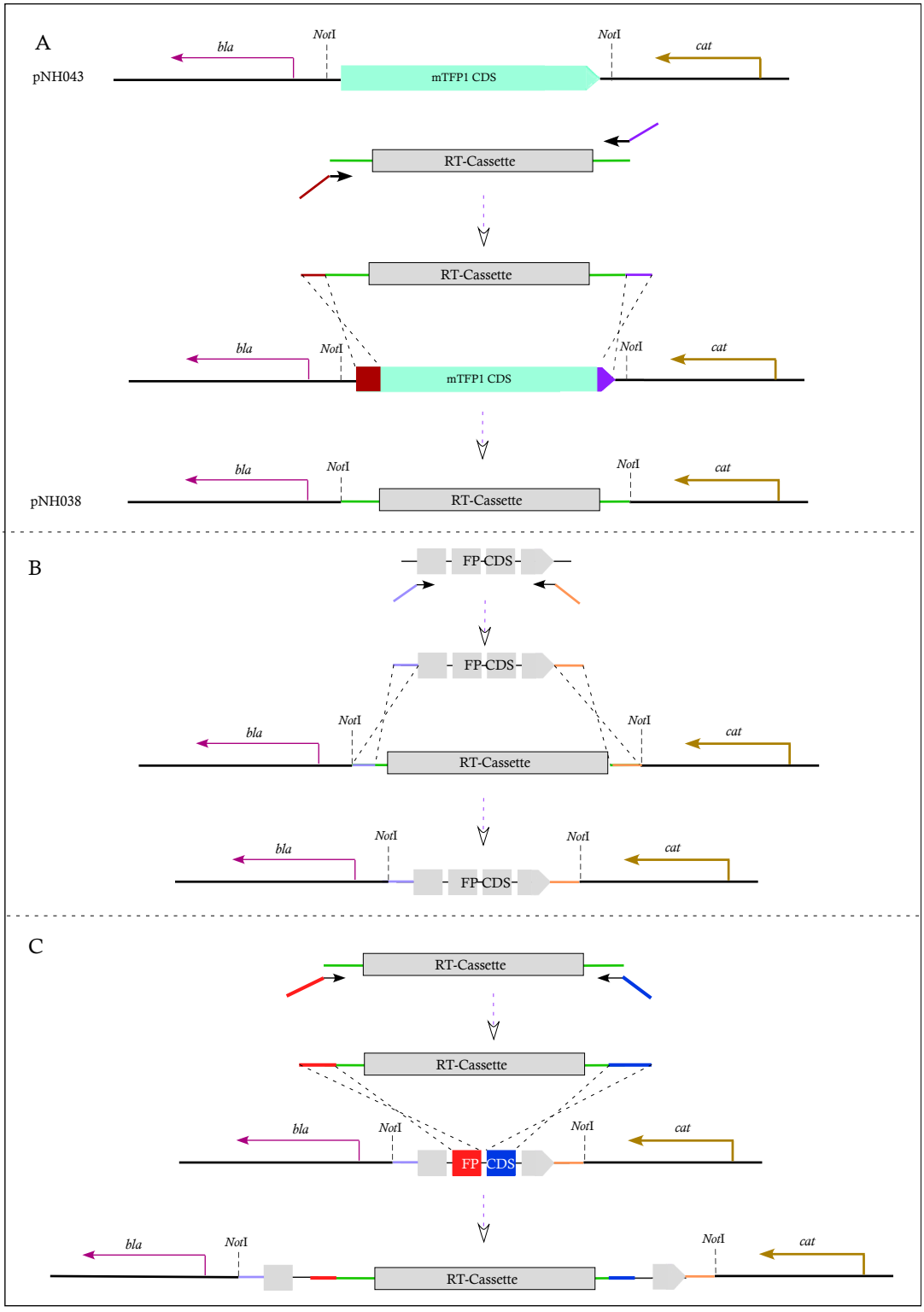

Figure 18. Generation of F-CFP, F-GFP, F-YFP and Mc-mCherry CDS-containing pCC1FOS-based constructs. PCR-amplified RT-Cassette replaced mTFP1 in pNH043 via positive selection to create pNH038 (Panel A). F-CFP, F-GFP, F-YFP and Mc-mCherry CDSs were PCR amplified and introduced into pNH038 via negative selection recombination to generate pNH039 – pNH042 (Panel B). RT-Cassette, with homology arms which would centrally introduce the RT and be flanked by 200 bp of the FP gene, was introduced via positive selection recombineering creating pNH050 – pNH053 (Panel C).

2.5.5 Identification of fosmid target

To validate the recombineering resource set described and generated, a fosmid target needed to be identified. Data from Wormbase (WS226, Yook et al., 2012) was examined to locate candidate pCC1FOS-based genomic fosmids that contained a cluster of genes with published expression pattern data. Filters were applied to the data to remove any data that did not have an associated reference to the expression pattern data. The extrapolated data was compiled into a list of genes by the level of connectivity in the ACeDB (Eeckman and Durbin, 1995) schema developed by the wormbase project. The gene co-ordinate data were used to align with the fosmid clone genomic positions by performing a GFF (General Feature Format) union to generate Gene::Fosmid connections. These data were then clustered and sorted on fosmid ID to produce a candidate list of over 4000 clones that contain three or more genes each of which has a published expression pattern. This led to the identification of fosmid clone WRM069dD11 which contains three centrally-located genes (*F09E5.3, F09E5.15* and *EEED8.6*, respectively) with previously published expression patterns.

2.5.6 Construction of single and multiple-tagged fosmid-based reporters via iterative counter-selection recombineering

Upon generation of the recombineering resources, a subset of the constructs were utilized in iterative rounds of counter-selection recombineering to tag *F09E5.3*, *F09E5.15* (*prdx-2*) and *EEED8.6* (*ccpp-6*), centrally located on fosmid clone WRM069dD11, each with a different codon-optimized FP CDS, both contiguous and dual-intron containing CDSs, to ultimately create a triple-tagged fosmid-based translational reporters. Fosmid-based constructs were generated to also incorporate the F-CFP, F-YFP and Mc-mCherry FP CDSs (Figures 19 and 20).

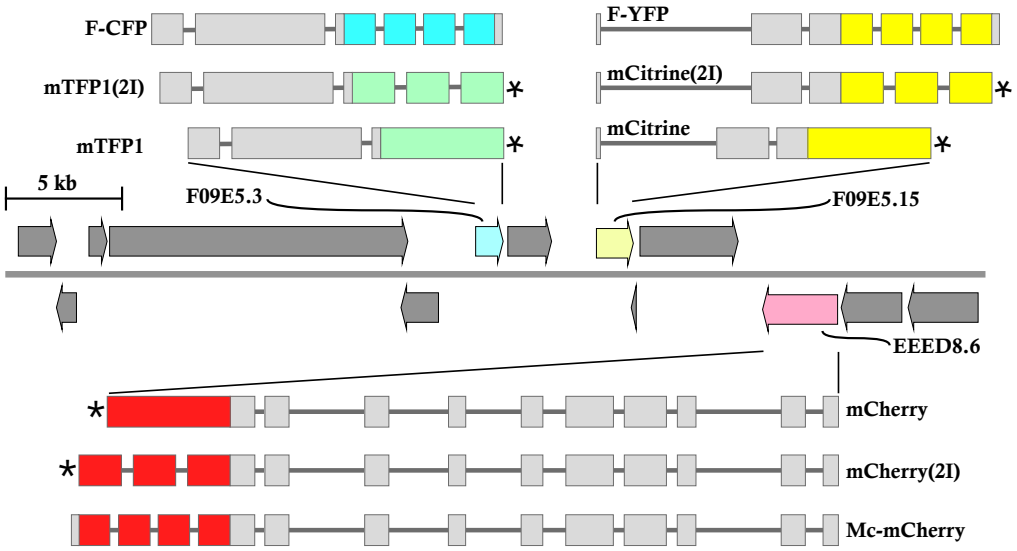

Figure 19. Counter-selection recombineering-mediated generation of fosmid-based FP-fusion reporter constructs. Schematic representation of sizes and orientations of the genes located on the insert of the fosmid clone WRM069dD11. Genes *F09E5.3*, *F09E5.15* and *EEED8.6* (coloured arrows) were tagged, respectively, with CDSs of either F-CFP, mTFP1(2I) or mTFP1, F-YFP, mCitrine(2I) or mCitrine, or Mc-mCherry, mCherry(2I) or mCherry. Genes fused with reporters marked with an asterisk lack the 5 terminal codons of the target gene.

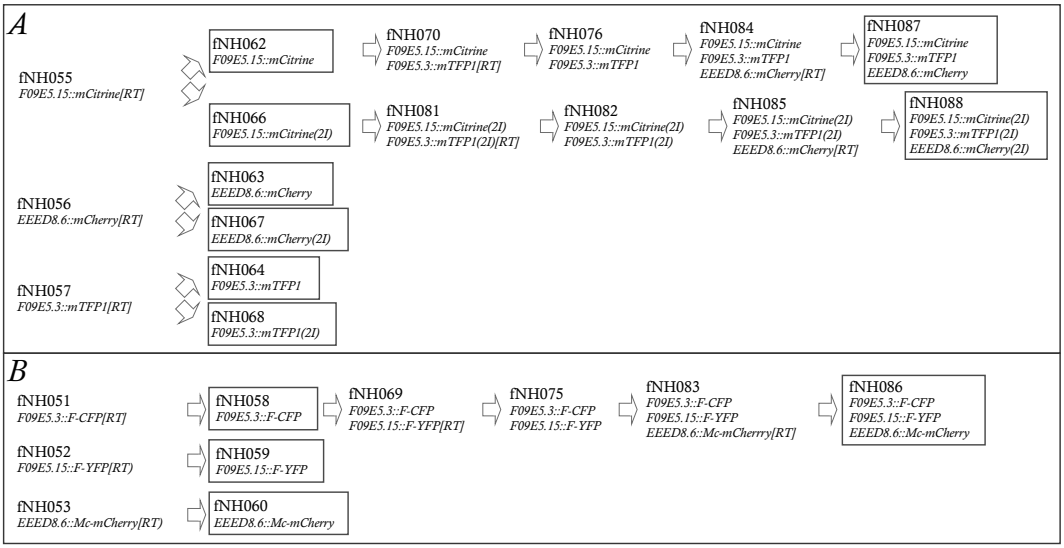

Figure 20. Counter-selection recombineering workflow for fosmid clone WRM069dD11. Codon-optimized FP CDSs, with and without two artificial introns (panel A) and also the F-CFP, F-YFP and Mc-mCherry FP CDSs (panel B) were introduced into fosmid clone WRM069dD11 via iterative rounds of counter-selection recombineering. The boxed fosmid constructs were microinjected to generate transgenic strains from which expression patterns were observed.

Briefly, *Not*I template fragments, isolated from their respective pCC1FOS-based construct, were PCR-amplified (Tables 3 and 4) to generate FP-RT-FP Cassette inserts. The homology arms were designed to seamlessly introduce the modified FP-RT-FP Cassette, and subsequent FP CDS, 5 codons upstream of the native genes stop codon. To ensure the final fusion gene terminated at the stop codon of the native gene, the stop codon of the FP gene was excluded from the ODN annealing sequence. Such ODNs were designed for the F-CFP, F-YFP and Mc-mCherry FP-RT-FP Cassette inserts however as the codon-optimized FP CDSs contained two successive stop codons the second stop codon was overlooked and one stop codon (TAA) of the synthetic codon-optimized FP gene was included in the reverse ODN annealing sequence and therefore the fosmid constructs containing the synthetic FP genes terminated prematurely and excluded the last five codons (Figure 19).

Table 3. RT-Cassette templates and ODNs used for creating fosmid-based constructs for clone WRM069dD11

Target gene	Insertion sequence	ODNs		Selection cassette	Replacement cassette
		Fwd	Rev		
F09E5.3	F-CFP	12074	12075	pNH050	pNH039
F09E5.15 (*prdx*-2)	F-YFP	12076	12077	pNH052	pNH041
EEED8.6 (*ccpp*-6)	Mc-mCherry	12078	12079	pNH053	pNH042
F09E5.3	mTFP1	12086	12087	pNH054	pNH013
F09E5.15 (*prdx*-2)	mCitrine	12082	12083	pNH055	pNH014
EEED8.6 (*ccpp*-6)	mCherry	12084	12085	pNH056	pNH015
F09E5.3	mTFP1(2I)	12086	12087	pNH054	pNH086
F09E5.15 (*prdx*-2)	mCitrine(2I)	12082	12083	pNH055	pNH087
EEED8.6 (*ccpp*-6)	mCherry(2I)	12084	12085	pNH056	pNH088

Table 4. Oligonucleotide sequences

No	sequence (5′-3′)[a]
12015	ccgggctgcatccgatgcaagtgtgtcgctgtcgacgagctcgcgagctcGTCTGACGCTCAGTGGAACGAAAACTCAC
12016	gtgaacaccgtcatatctcgacagcgaccatggAGTGCTGCCATAACCATGAGTGATAAC
12017	ccaattctcatgtttgacagcttatcatcccatggGCTGAGGGGCTGGCCGCTCACAATTC
12018	CGCGTTGGCCGATTCATTAATGCAGC
12024	accgctacacttgccagcccctcagcgcggccgctagcggatccttaaaaGCTGTCGAGATATGACGGTGTTCA
12025	atgttgtgtggaattgtgagcggccagcccctcagcgcggccgcggtaccTCTTGGAGTGGTGAATCCGTTAGC
12026	accgctacacttgccagcccctcagcgcggccgctagcggatccttaaaaATGGTCTCAAAGGGTGAAGAAGATA
12027	atgttgtgtggaattgtgagcggccagcccctcagcgcggccgcggtaccTTAGGATCCACTAGTCTTATACAAT
12028	accgctacacttgccagcccctcagcgcggccgctagcggatccttaaaaATGAGTAAAGGAGAAGAACTTTTCA
12029	atgttgtgtggaattgtgagcggccagcccctcagcgcggccgcggtaccCTATTTGTATAGTTCATCCATGCCA
12030	tactaactaaccctgattatttaaattttcagccaacacttgtcactactGCTGTCGAGATATGACGGTGTTCA
12031	acgcttccatcttcaatgttgtgtctaattttgaagttctgaaaatttaaTCTTGGAGTGGTGAATCCGTTAGC
12032	attatttaaattttcaggtaactaaaggcggaccattaccattcgcctggGCTGTCGAGATATGACGGTGTTCA
12033	gattagttagttagtaaaatcatgtttaaacttacctcagcatcgtaatgTCTTGGAGTGGTGAATCCGTTAGC
12036	GTATGTTGTGTGGAATTGTGAGCGG
12037	GCAGCTAATACGACTCACTATAGGGTGTGGTGGTTACGCGCAGCGTGA
12038	GCAGCTAATACGACTCACTATAGGTTACCGGCCGCCACCATGGTGAGCAAGGGCGAGGAGCTGT
12039	CTCTACAAATGTGGTATGGCTGATT
12042	tactacagccttcgcatatggaaatcgagcattcacaaaatacccagacgGCTGTCGAGATATGACGGTGTTCA
12043	ccaacaacagtttatgcttcacgtcacctttcaaaactccatctctgacaTCTTGGAGTGGTGAATCCGTTAGC
12044	atatggcttgatgtgtttcgcaagatatccagatcacatgaagcaacacgGCTGTCGAGATATGACGGTGTTCA
12045	cttccatcctcgatgttgtgacggattttgaaattaactttaattccgttTCTTGGAGTGGTGAATCCGTTAGC
12046	gtctccacagtttatgtacggttcgaaggcatatgttaagcatccggcggGCTGTCGAGATATGACGGTGTTCA
12047	tgagcttaagtcgttgctttatttctcccttaagagcaccgtcctctgggTCTTGGAGTGGTGAATCCGTTAGC
12048	atggggtgtccaatgctttgctagatatccagaccatatgaaacagcatgGCTGTCGAGATATGACGGTGTTCA
12049	gaaccgtcttcaatattatgtctaatcttaaagttggccttgataccattTCTTGGAGTGGTGAATCCGTTAGC
12074	ggctcaatcctcacttgttccgcatcggagctagcagtcttcttgatgatATGAGTAAAGGAGAAGAACTTTTC
12075	aacaaaaaaaaaagatgataacaaaaactgatctaaagacccttcagacaTTTGTATAGTTCATCCATGCCATG
12076	ggactccaggatccgacaccatcaagccaggagtcaaagaaagccaagagATGAGTAAAGGAGAAGAACTTTTC
12077	gggataccgggggaaattagagatgtaagacatttagtgcttcttgaagtaTTTGTATAGTTCATCCATGCCATG
12078	ccgcccgcgcccagaaaactttgaaaaagcaagcgaggaaggtgattcagATGGTCTCAAAGGGTGAAGAAGAT
12079	tggtctttatcagttctaatgtgggcgccaattcatttcatcatcacagtGGATCCACTAGTCTTATACAATTC
12082	ggactccaggatccgacaccatcaagccaggagtcaaagaaagccaagagATGGCCGCCAGTAAGGGTGAGGAG
12083	gggataccgggggaaattagagatgtaagacatttagtgcttcttgaagtaTTAAGCGCTTACGTAGAGCTCATC
12084	ccgcccgcgcccagaaaactttgaaaaagcaagcgaggaaggtgattcagATGGCCGCCAGTAAGGGTGAGGAG
12085	tggtctttatcagttctaatgtgggcgccaattcatttcatcatcacagtTTAAGCGCTTACGTAGAGCTCATC
12086	ggctcaatcctcacttgttccgcatcggagctagcagtcttcttgatgatATGGCCGCCTCAAAAGGAGAAGAA
12087	aacaaaaaaaaaagatgataacaaaaactgatctaaagacccttcagacaTTAAGCGCTTACGTAGAGCTCGTC
12094	catctcaccatatcgtcgaaggcgcgccattgtcctactcgtcgaataatATGGCCGCCTCAAAAGGAGAAGAA
12095	ggtatgatttcaacaggatttagcaaaaatatttagatttggggagttccTTAAGCGCTTACGTAGAGCTCGTC
12098	ttcaaaaacaaaaaaaaggattaaataagttacttgtcgagtctacagatAGCGCTTACGTAGAGCTCATCCAT
12113	tattgtcactatcatcaccattaggtgttctaatatcaatacttatgtttGCTGTCGAGATATGACGGTGTTCA
12114	gaacagcggtgatgtgacagagacgggtaaacaaaaaaaatcgacttgagTCTTGGAGTGGTGAATCCGTTAGC
20014	aattgaaggataaactcaatgagcgatatgccaataggaacagacgtgttATGAGTAAAGGAGAAGAACTTTTC
20015	aaaaacacaaatatacagaagatttcaaattaagcgccaaggaattgtccTTTGTATAGTTCATCCATGCCATG
20016	gtgatcttttcggaagcaaaatcgatttgaatgtggtctacaataatccaATGGTCTCAAAGGGTGAAGAAGAT
20017	ttcaaaaacaaaaaaaaggattaaataagttacttgtcgagtctacagatGGATCCACTAGTCTTATACAATTC
20024	caccatctcaccatatcgtcgaaggcgcgccattgtcctactcgtcgaatATGAGTAAAGGAGAAGAACTTTTC
20025	atgatttcaacaggatttagcaaaaatatttagatttggggagttccattTTTGTATAGTTCATCCATGCCATG
20026	acgcagccatggtaaagccgacaaaaagaagcttcataaaaaagcgaaaaATGGTCTCAAAGGGTGAAGAAGAT
20027	aatgcatttttaaatggtaagtgatataatcattgataaacgatcgaagaGGATCCACTAGTCTTATACAATTC
20030	acgcagccatggtaaagccgacaaaaagaagcttcataaaaaagcgaaaaATGGCCGCCAGTAAGGGTGAGGAG

[a]For ODNs used in recombineering procedures the homology arm and template annealing sequences are given in lower and upper case, respectively. Premature stop codons in the reverse ODNs 12083, 12085 and 12087 are underlined.

2.6 Identification of operon targets

To generate a list of target operons from the >1500 operon structures annotated in wormbase (WS232) the following criteria were used. The operon should be centrally located on a single fosmid clone to ensure maximum upstream and downstream sequence of the operon target is included and, due to the number of fluorescent reporters which could be spectrally distinguished, the operons should contain no more than 3 genes. The genes contained within these operons should have structures confirmed via ESTs (expressed sequence tags) with at least one of the genes having

published expression pattern data. Finally, the operon-genes should contain confirmed SL1 and SL2 spliced-leader sequences in downstream genes with the first gene confirmed for SL1. With the above criteria, the *C. elegans* genome in wormbase was scanned by eye, and a list of possible operon targets was compiled.

2.6.1 Identification of gene models

Upon identification of the operon targets the next step was to select the gene models to investigate. As the FPs will be recombineered 6 codons before the stop codon of the operon genes, the 3′ end of each gene model was investigated. Gene models supported by evidence in the form of EST alignments and ORFeome sequence tags were selected for further study.

2.6.2 Investigating expression patterns of operon-genes *via* fosmid-based translational reporters

The genes within operons CEOP1312 (Figure 42) and CEOP1358 (Figure 43) were iteratively tagged via counter-selection recombineering, utilizing the recombineering constructs generated in chapter 4, with F-CFP, F-YFP and Mc-mCherry CDSs. To test the synthesized codon-optimized FPs and investigate whether the lack of fluorescent signal observed (chapter 4) was due nonsense mediated mRNA decay (premature stop codon sourced from the codon-optimized FP CDSs was included in the ODN sequences by error), they were also introduced into CEOP1358 (Figure 21). The ODNs were designed to introduce the FP reporters 6 codons upstream from the stop codon of each gene (Table 5).

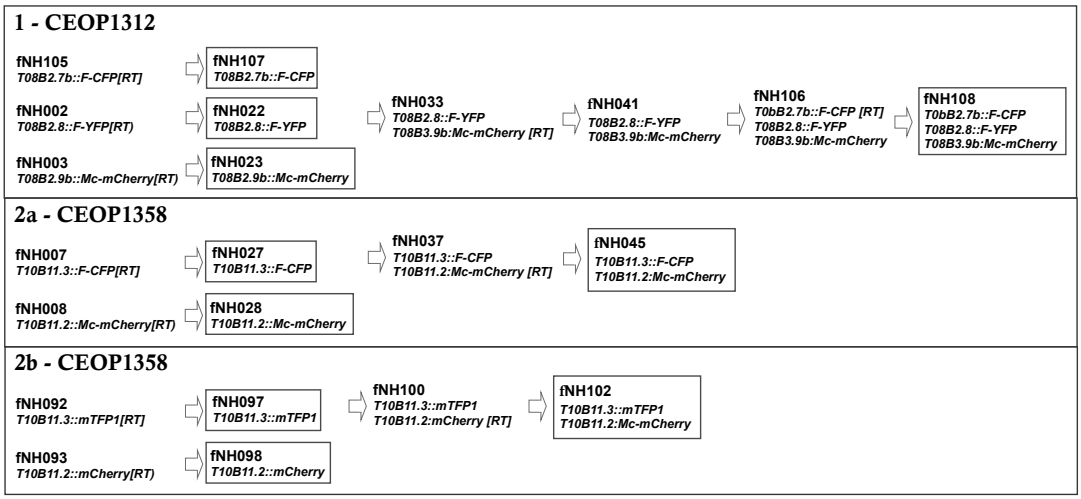

Figure 21. Counter-selection recombineering workflow for both operon targets. F-CFP, F-YFP and Mc-mCherry FP CDSs were introduced into CEOP1312 (panel 1), and F-CFP and Mc-mCherry FP CDSs (panel 2a) and codon-optimized mTFP1 and mCherry CDSs (panel 2b) were introduced into CEOP1358 via iterative rounds of counter-selection recombineering. The boxed fosmid constructs were microinjected to create transgenic strains from which fosmid-based expression patterns were obtained.

Table 5. RT-Cassettes and ODNs used for creating fosmid-based operon constructs

Target gene	Insertion sequence	ODNs Fwd	ODNs Rev	Selection cassette	Replacement cassette
T08B2.7b	F-CFP	12074	12075	pNH050	pNH039
T08B2.8	F-YFP	20014	20015	pNH052	pNH041
T08B2.9b	Mc-mCherry	20016	20017	pNH053	pNH042
T10B11.3	F-CFP	20024	20025	pNH050	pNH039
T10B11.2	Mc-mCherry	20026	20027	pNH053	pNH042
T10B11.3	mTFP1	12094	12095	pNH054	pNH013
T10B11.2	mCherry	20030	12098	pNH056	pNH015

2.7 Identification and subsequent modification of potential internal regulatory elements

Operons CEOP1312 and CEOP1358 were checked for synteny in the *Caenorhabditis* genus using ensembl. Pustell DNA matrix analysis (Pustell and Kafatos, 1984) was used to align CEOP1312 in *C. elegans*, *C. briggsae*, *C. brenneri* and *C. remanei* and a small

48 bp non-coding conserved sequence (NCS) was identified. As conserved sequences imply functional importance and significance, this sequence will be modified and its relevance to CEOP1312 dissected out.

The RT-Cassette (*Nco*I gel purified frag from pNH034) was PCR-amplified with ODNs 12113/12114 (Table 4), column purified and recombineered into fNH108, triple-tagged fosmid containing CEOP1312 tagged with F-CFP, F-YFP and Mc-mCherry CDSs, to create fNH109. The RT-Cassette in this construct was replaced with a gBlock (IDT, Belgium) centrally-containing the modified replacement sequence, and 200 bp of flanking sequence, to create fNH112 (NCS deleted), fNH113 (NCS reverse and complemented) and finally fNH114 (NCS replaced with a scrambled sequence of an equivalent length). These three fosmid-based constructs were microinjected and the independently generated lines were imaged.

3

Codon-optimized fluorescent proteins as reporters of gene expression in *C. elegans*

3.1 Introduction

To select FPs for multiple tagging, the spectral properties of the FPs must be investigated to minimize cross-talk and a list of currently available FPs was investigated (Shaner et al., 2007). The FPs for multiple-tagging were selected in the cyan, cyan/green, yellow and red wavelength regions to allow maximal spectral separation between the FPs. To introduce new FPs into the collection of reporters utilized in *C. elegans* research the FPs selected were mCerulean, mTFP1, mCitrine and mCherry and as the sequences were to be commercially synthesized, they were codon-optimized to maximize transcription and translation of the FP CDSs. To expand the utility of these FP constructs TAP tags, which could affix at both the N- and C-terminus of the FP, were also designed and commercially synthesized.

3.2 Methods

See sections 2.4.1 (design considerations), 2.4.2 (commercial synthesis), 2.4.3 (derivation of pGOv5-based FP-encoding sub-clones), 2.4.4 (introduction of artificial introns), 2.4.5 (building *myo-3*^{PROM} driven transcriptional reporters), 2.4.6 (immunocapture of *in vitro* transcription-translation products), 2.4.7 (affinity-capture: S-tag) and 2.4.8 (affinity-capture: PreScission and Strep-Tag II).

3.3 Results

3.3.1 FP reporter genes: choice, design and synthesis

After considering the biophysical and spectral properties for selecting FPs (Shaner et al., 2005, 2007) to act as reporters for multiple tagging, the four FPs chosen for synthesis, and subsequent validation, were mCerulean (Rizzo et al., 2004), mTFP1 (Ai et al., 2006), mCitrine (Griesbeck et al., 2001) and mCherry (Shaner et al., 2004) (Figure 22). The FP genes were designed using the software package Gene Designer (DNA 2.0, Villalobos et al., 2006 and Welch et al., 2011). Gene Designer enables the user to construct an amino acid sequence and then generate the corresponding

encoding DNA sequence by reverse translation. By specifying sequence requirements and/or restrictions the encoding sequence can be forced to include, or exclude, specific features such as, for example, specific restriction enzyme recognition sites. In addition, codon usage can be optimized for a particular species – in this case *C. elegans*. In this manner, each of the four FP amino acid sequences were compiled and CDSs generated. All such coding sequences were designed to be flanked by, and contain, unique restriction sites resulting in a modular assembly designed to facilitate all of the envisaged downstream sequence manipulations. Flanking *Not*I sites were included to allow facile excision of the complete coding sequence. A number of blunt-cutting restriction enzyme sites were included in the coding sequences to allow future insertion of artificial introns. All coding sequences, including those regions encoding the FPs and TAP-tag sections, were optimized for codon usage in *C. elegans* using a default threshold codon frequency of 12%.

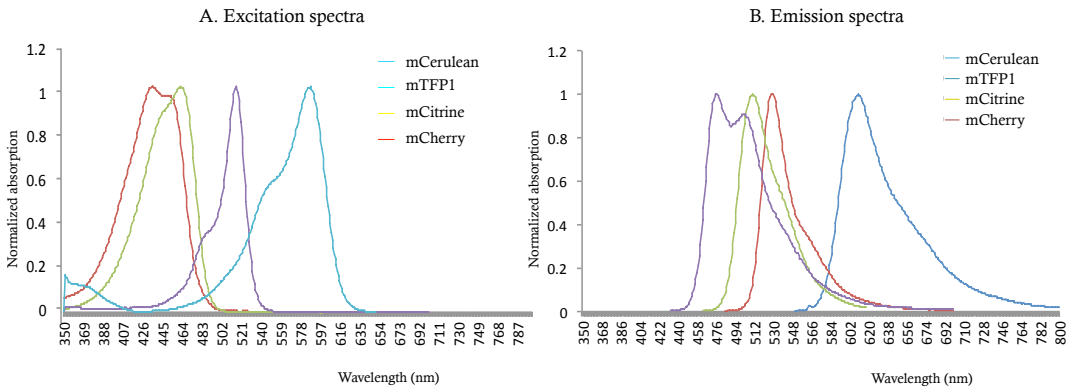

Figure 22. Excitation (A) and emission (B) spectra are displayed of the four chosen fluorescent proteins mCerulean, mTFP1, mCitrine and mCherry.

The primary amino acid sequence of each FP was as published apart from the following differences. In both mCerulean and mCitrine the A206K mutation, demonstrated to promote monomerization (Zacharias et al., 2002), was included. As regions of these proteins shared identical, or highly similar, amino acid sequences the reverse translation stage included a requirement that the four corresponding coding sequences should, within the constraints of the other sequence specifications, also be different from each other. This design step was included to minimize any potential intra-molecular recombination from occurring when two or more FP CDSs might be present within the same fosmid-based construct. In addition, an N-TAP tag and C-TAP tag were also designed for either the N- or C- termini of each FP. Both versions of the TAP tags were designed to fuse to the FP via a flexible linker [$(G_4S)_3$] followed by the internal tag Strep Tag II [WSHPQFEK] (Schmidt and Skerra, 2007), a short

linker [G_4S], a human rhinovirus 3C protease cleavage site [LEVLFQGP] (Walker et al., 1994), a second short linker [G_4S] and the external S-Tag [KETAAAKFERQHMDS] (Raines et al., 2000). Furthermore, two sequences of SV40 nuclear localization signal (2x NLS) were also designed for the C-TAP tag variant and all coding sequences contained two stop codons (TAA-TAG) to ensure translation termination.

Following their design, the coding sequences were synthesized commercially and provided in proprietary plasmid vectors. Sequences encoding C-TAP tag::2x NLS and N-TAP tag::mTFP1 were synthesized (Gene Oracle) and supplied as the respective plasmids pNH001 and pNH002 based in the common restriction enzyme site-depleted vector pGOv5. mCitrine, mCherry and mCerulean coding sequences were synthesized (Mr Gene) and supplied as the respective plasmids pNH003, pNH004 and pNH005 based in vector pMK.

A total of 52 constructs were created during the work described here (Table 6). Constructs pNH001 - pNH005 were commercially synthesized from which a series of sub-clones were generated with some containing either a N-/C-TAP tag and/or 2x NLS. Two different artificial introns were sequentially introduced into each contiguous codon-optimized FP CDS and the dual intron-containing codon-optimized FP CDSs were validated via the *myo-3* promoter constructs (Chapter 2.4.5). The F-CFP, F-YFP and Mc-mCherry CDSs were also placed under the control of the *myo-3* promoter to control for the methodology and imaging system employed. To test if the codon-optimized synthesized FP CDSs would successfully translate into functional FPs, the *in vitro* validation, via immuno-capture of the FP via appropriate antibodies onto agarose beads, was performed first and then followed by the *in vivo* validation, the generation of strains transgenic for the various FP CDSs with their expression driven by *myo-3*^{PROM}.

Table 6. A list of constructs generated with either the pGOv5 or pPD95.86 backbones

Construct	Insert	Vector	Promoter	NLS	Intron 1	Intron 2
pNH001	$[(G_4S)_3]$::C-TAP-tag::2xNLS	pGOv5	-	-	-	-
pNH002	N-TAP-tag::$[(G_4S)_3]$::mTFP1	pGOv5	-	-	-	-
pNH003	mCitrine	pGOv5	-	-	-	-
pNH004	mCherry	pGOv5	-	-	-	-
pNH005	mCerulean	pGOv5	-	-	-	-
pNH006	N-TAP-tag::$[(G_4S)_3]$::mCitrine	pGOv5	-	-	-	-
pNH007	N-TAP-tag::$[(G_4S)_3]$::mCherry	pGOv5	-	-	-	-
pNH008	N-TAP-tag::$[(G_4S)_3]$::mCerulean	pGOv5	-	-	-	-
pNH009	N-TAP-tag::mTFP1	pGOv5	-	-	-	-
pNH010	N-TAP-tag::mCitrine	pGOv5	-	-	-	-
pNH011	N-TAP-tag::mCherry	pGOv5	-	-	-	-
pNH012	N-TAP-tag::mCerulean	pGOv5	-	-	-	-
pNH013	mTFP1	pGOv5	-	-	-	-
pNH014	mCitrine	pGOv5	-	-	-	-
pNH015	mCherry	pGOv5	-	-	-	-
pNH016	mCerulean	pGOv5	-	-	-	-
pNH017	mTFP1::$[(G_4S)_3]$::C-TAP-tag::2xNLS	pGOv5	-	Y	-	-
pNH018	mCitrine::$[(G_4S)_3]$::C-TAP-tag::2xNLS	pGOv5	-	Y	-	-
pNH019	mCherry::$[(G_4S)_3]$::C-TAP-tag::2xNLS	pGOv5	-	Y	-	-
pNH020	mCerulean::$[(G_4S)_3]$::C-TAP-tag::2xNLS	pGOv5	-	Y	-	-
pNH022	mTFP1::$[(G_4S)_3]$::C-TAP-tag::2xNLS	pPD95.86	*myo*-3	Y	-	-
pNH023	mCitrine::$[(G_4S)_3]$::C-TAP-tag::2xNLS	pPD95.86	*myo*-3	Y	-	-
pNH024	mCherry::$[(G_4S)_3]$::C-TAP-tag::2xNLS	pPD95.86	*myo*-3	Y	-	-
pNH025	mCerulean::$[(G_4S)_3]$::C-TAP-tag::2xNLS	pPD95.86	*myo*-3	Y	-	-
pNH026	mTFP1::$[(G_4S)_3]$::C-TAP-tag	pGOv5	-	-	-	-
pNH027	mCitrine::$[(G_4S)_3]$::C-TAP-tag	pGOv5	-	-	-	-
pNH028	mCherry::$[(G_4S)_3]$::C-TAP-tag	pGOv5	-	-	-	-
pNH029	mCerulean::$[(G_4S)_3]$::C-TAP-tag	pGOv5	-	-	-	-
pNH030	mTFP1::$[(G_4S)_3]$::2xNLS	pGOv5	-	Y	-	-
pNH031	mCitrine::$[(G_4S)_3]$::2xNLS	pGOv5	-	Y	-	-
pNH032	mCherry::$[(G_4S)_3]$::2xNLS	pGOv5	-	Y	-	-
pNH033	mCerulean::$[(G_4S)_3]$::2xNLS	pGOv5	-	Y	-	-
pNH035	mCherry	pPD95.86	*myo*-3	-	-	-
pNH047	mTFP1	pPD95.86	*myo*-3	-	-	-
pNH048	mCitrine	pPD95.86	*myo*-3	-	-	-
pNH049	mCerulean	pPD95.86	*myo*-3	-	-	-
pNH074	F-CFP	pPD95.86	*myo*-3	-	n/a	n/a
pNH075	F-GFP	pPD95.86	*myo*-3	-	n/a	n/a
pNH076	F-YFP	pPD95.86	*myo*-3	-	n/a	n/a
pNH077	Mc-mCherry	pPD95.86	*myo*-3	-	n/a	n/a
pNH078	mTFP1[1I]	pGOv5	-	-	B	-
pNH079	mCitrine[1I]	pGOv5	-	-	α	-
pNH080	mCherry[1I]	pGOv5	-	-	β	-
pNH081	mCerulean[1I]	pGOv5	-	-	γ	-
pNH082	mTFP1[2I]	pGOv5	-	-	B	C
pNH083	mCitrine[2I]	pGOv5	-	-	α	D
pNH084	mCherry[2I]	pGOv5	-	-	β	E
pNH085	mCerulean[2I]	pGOv5	-	-	γ	F
pNH086	mTFP1[2I]	pPD95.86	*myo*-3	-	B	C
pNH087	mCitrine[2I]	pPD95.86	*myo*-3	-	α	D
pNH088	mCherry[2I]	pPD95.86	*myo*-3	-	β	E
pNH089	mCerulean[2I]	pPD95.86	*myo*-3	-	γ	F

3.3.2 *In vitro* validation: immuno-capture of *in vitro* expressed FPs

The concentration of the FPs to beads, via immuno-capture, and subsequent visualization of the expected fluorescence demonstrated the synthesized codon-optimized, for *C. elegans* expression, FP CDSs do fold correctly and the protein can be expressed *in vitro* using a commercial TNT kit and the characteristics of the fluorescence observed is as expected (Figure 23).

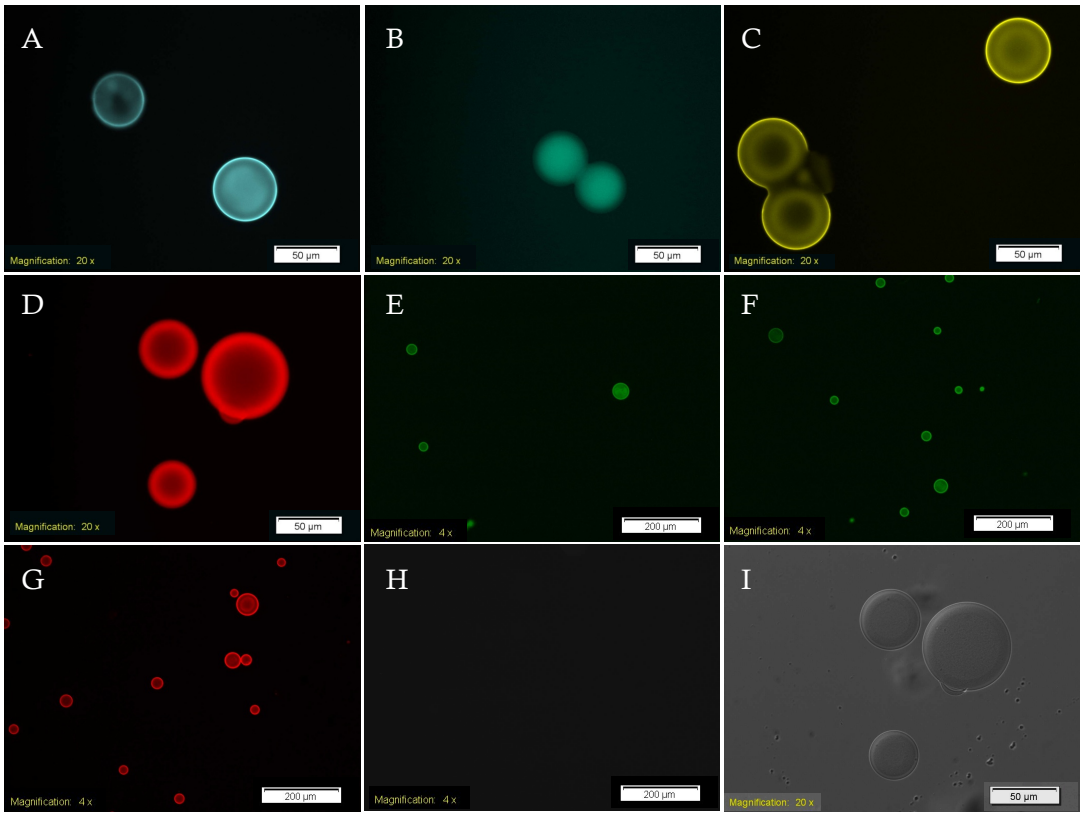

Figure 23. *In vitro* validation of synthetic FPs. $T7^{PROM}$-containing FP amplicons were translated (T7 TNT kit, Promega) and the FPs captured on protein G-agarose beads via appropriate anti-FP antibodies. A) mCerulean, B) mTFP1, C) mCitrine, D) mCherry, and controls, E) eGFP (Clontech), F) eGFP (codon-optimized for *C. elegans* expression, this lab), G) TagRFP-T (codon-optimized for expression in Sf9 cells, this lab), H) TNT kit control and I) beads alone (DIC). Scale bars in panels A – D and I represents 50 µm and scale bars in panels F – H illustrates 200 µm.

3.3.3 *In vivo* validation: *myo-3*^{PROM}-driven FP expression

In vivo FP validation was performed to investigate the FP CDS containing the C-TAP tag and 2x NLS sequences. These particular constructs were chosen to ensure all

synthesized sequences, that could be present in the construct used, were functional and not toxic to the nematode.

The CDSs to be investigated for expression were cloned behind the *myo-3* promoter sequence contained within vector pPD95.86 (Fire vector) and microinjected. The resulting transgenic lines were imaged. As demonstrated the FP translated from the synthesized codon-optimized FP CDSs folded correctly and when expressed *in vivo* no toxicity or side effects to the transgenic worms were detected (Figure 24). The 2x NLS sequence directs the FPs to the muscle cell nuclei (Figure 24, panels B and C). In addition to the punctate nuclear expression, there was evidence of more diffuse extra-nuclear expression most likely due to simple diffusion of the relatively small FP out of the nucleus.

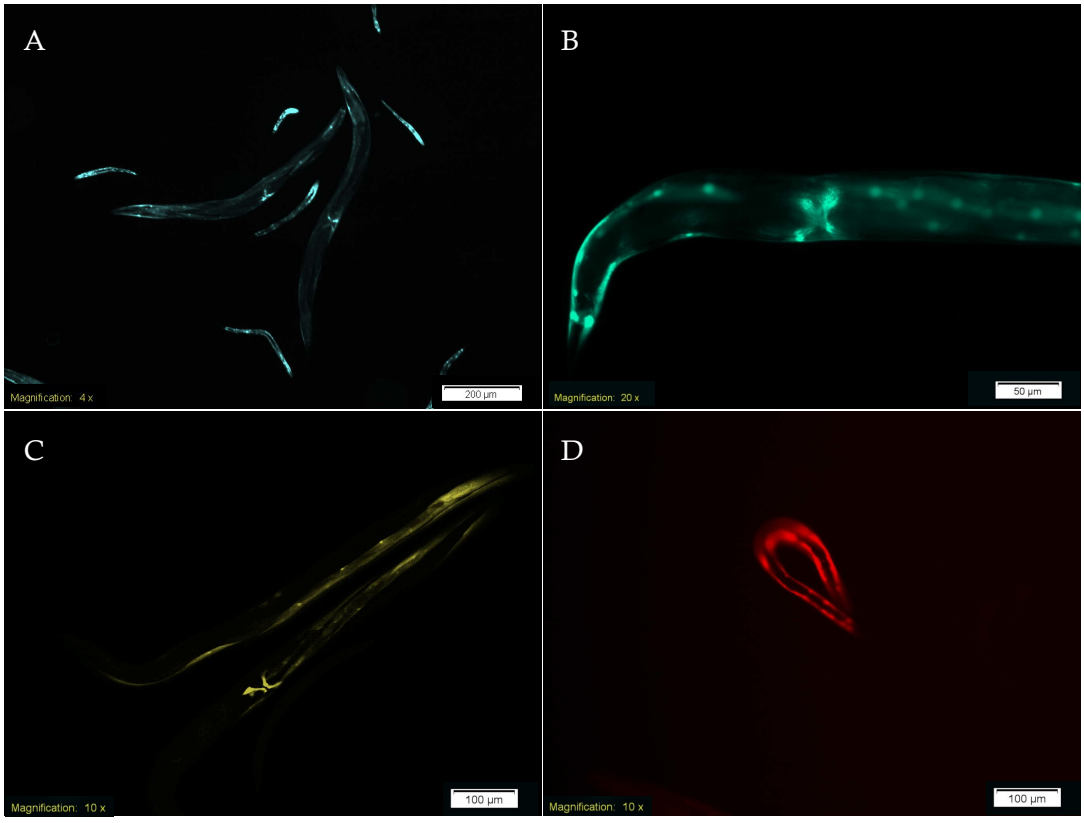

Figure 24. *in vivo* validation FP::C-TAP tag::2x NLS muscle expression. Myo-3^{PROM} driven FP::C-TAP tag::2x NLS, cloned into pPD95.86, were microinjected with pRF4 and independent lines generated and imaged. A) mCerulean, B) mTFP1, C) mCitrine and D) mCherry. Scale bar in panel A represents 200 µm, in panels C-D 100 µm and in panel B illustrates 50 µm.

The next *in vivo* validation step was to examine the FP alone to determine the photostability and brightness of the synthesized constructs and also to test if the

translated protein formed aggregates. Upon the generation of the independent transgenic lines, by microinjecting constructs pNH035, pNH047, pNH048 and pNH049, the results obtained (Figure 25) indicated that the FP expressed in the muscle cells of the worm with no indication of aggregation formation.

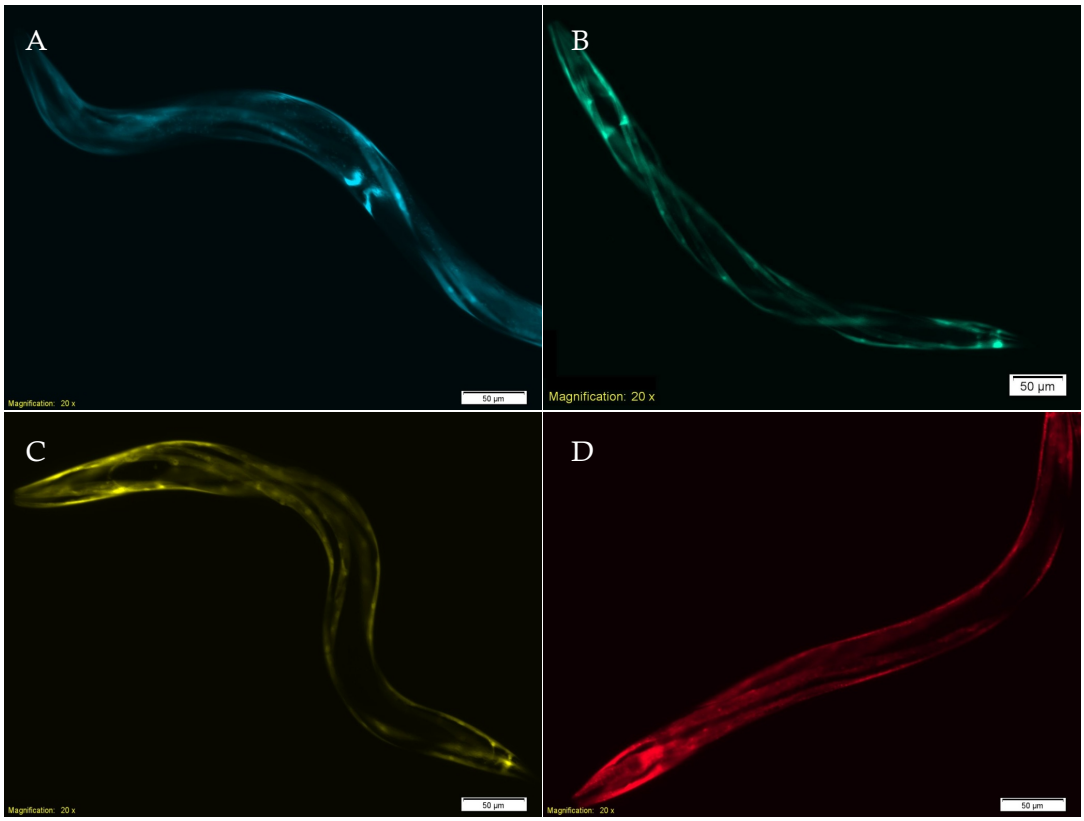

Figure 25. ***in vivo*** **validation: FP muscle expression.** Myo-3^{PROM} driven contiguous FP constructs were microinjected with pRF4 into gravid adult N2s and independent lines were created and imaged. A) mCerulean, B) mTFP1, C) mCitrine and D) mCherry. Scale bar illustrates 50 µm.

The final set of *in vivo* validation investigated if the insertion of 2 artificial introns in each FP CDS altered the rate of its transcription and translation and resulted in a visible increase of the brightness of the FP. The results obtained when imaging the nematodes transgenic for pNH086, pNH087, pNH088 and pNH089 indicated that the presence of the two artificial introns in each FP sequence does increase the signal of the resulting FP as the exposure times were reduced by at least five-fold (exposure times of 2 – 100 msec vs. exposure times of 50 – 500 msec for their respective non-intron containing counterpart FPs) in order to capture an equivalent image of the FP muscle expression where the FP gene was a contiguous CDS with no artificial introns present (Figure 26).

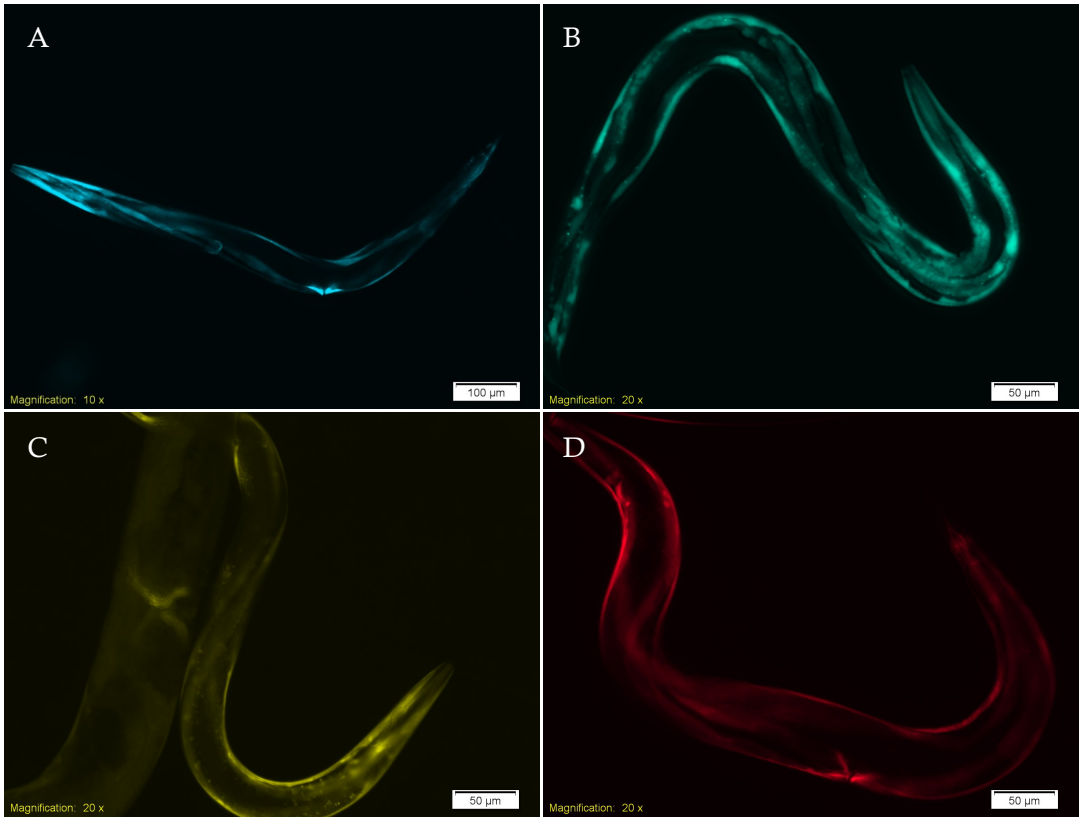

Figure 26. *in vivo* validation: FP (2 artificial intron) muscle expression. Myo3^{PROM} driven FPs (containing 2 artificial introns) plasmid constructs were microinjected with co-marker pRF4 and independent lines generated and finally imaged. A) mCerulean, B) mTFP1, C) mCitrine and D) mCherry. Scale bar in panel A indicates 100 µm and in the remaining panels 50 µm.

As a set of *in vivo* controls and also to allow comparison of the expression pattern data for the synthetic codon-optimized fluorescent proteins, the F-CFP, F-YFP and Mc-mCherry CDSs were also cloned into pPD95.86 to create constructs pNH074, pNH076 and pNH077 respectively. When the transgenic lines created from these constructs were imaged, (Figure 27) results demonstrated fluorescence in the expected tissue.

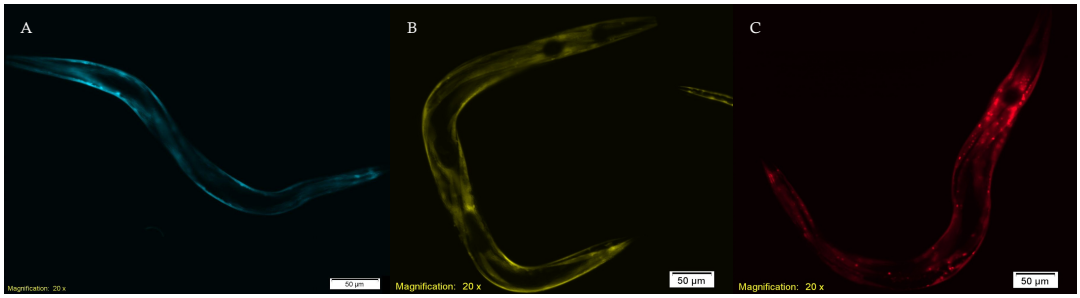

Figure 27. *in vivo* control FP muscle expression. Myo-3^{PROM} driven F-CFP, F-YFP and Mc-mCherry transcriptional reporter constructs were microinjected with pRF4 and independent lines imaged. A) Fire CFP, B) Fire YFP and C) McNally mCherry. Scale bar indicates 50 µm.

3.3.4 TAP-tag structure: affinity capture and protease release

The first step in validating the TAP tag was to investigate the functionality of the S-tag. Results illustrates that when the entire *FP::C-TAP tag* CDS was translated the S-tag was available for binding as it was used to anchor the FP to the S-protein coated agarose beads and the resulting protein fluoresced (Figure 28).

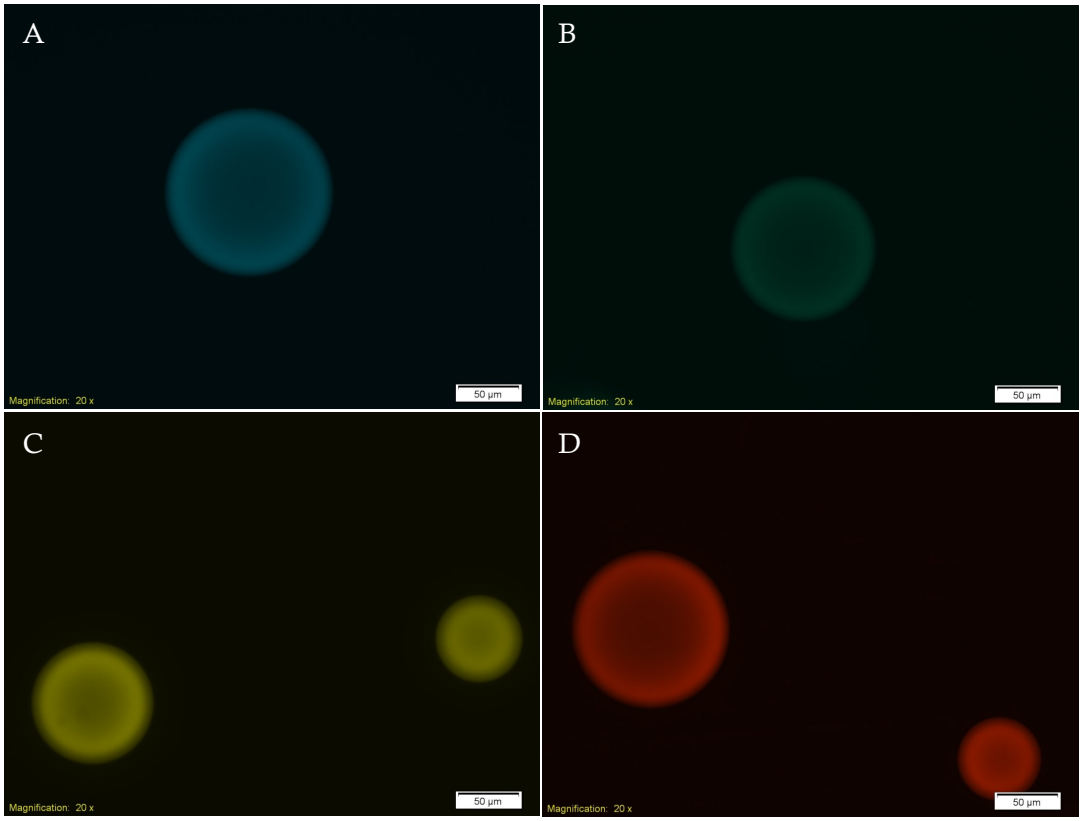

Figure 28. *in vitro* validation: FP::C-TAP tag. T7^{PROM}-containing FP::C-TAP tag amplicons were translated *in vitro* (TNT, Promega) and FPs captured onto S-protein beads via the S-tag. A) mCerulean, B) mTFP1, C) mCitrine and D) mCherry. Scale bar indicates 50 µm.

Next was to examine if the protease, human rhinovirus 3C protease, cleavage site could be recognized and cleaved, which should result in the S-tag being cleaved off the protein, and the remaining protein product only being identified via the Strep-Tag II using a monoclonal antibody. The use of this protease site will enable the utility of commercially available human rhinovirus 3C protease fused with, for example, GST- (PreScission, GE Healthcare) or His-tags (HRV 3C Protease, Novagen) that facilitate protease removal from column eluate after target protein cleavage and therefore HRV 3C protease was also used for identifying the cleavage site.

The FP mTFP1, which has a C-terminal fusion to TAP tag, was incubated with equal units of PreScission and HRV 3C Protease under identical conditions. These protein samples were separated by denaturing gel electrophoresis and visualized via HRP and ECL, using an anti-S-tag monoclonal antibody (Figure 29A) and subsequently using an anti-Strep-Tag II monoclonal antibody (Figure 29B). If the proteases cleaved at the protease site, in the TAP tag, the S-tag would no longer be attached to the FP and therefore should not be detected by the anti-S-tag antibody. However as the Strep-Tag II is still present within the FP, the protein should be identifiable by the anti-Strep-Tag II antibody. The results obtained suggests that neither protease successfully cleaved all the S-tag from the protein as in the presence of both proteases the S-tag monoclonal antibody could still recognize the protein (Figure 29A, lanes 3 and 6). Unexpectedly both monoclonal antibodies also recognized the protease controls (Figure 29, lanes 4 and 7).

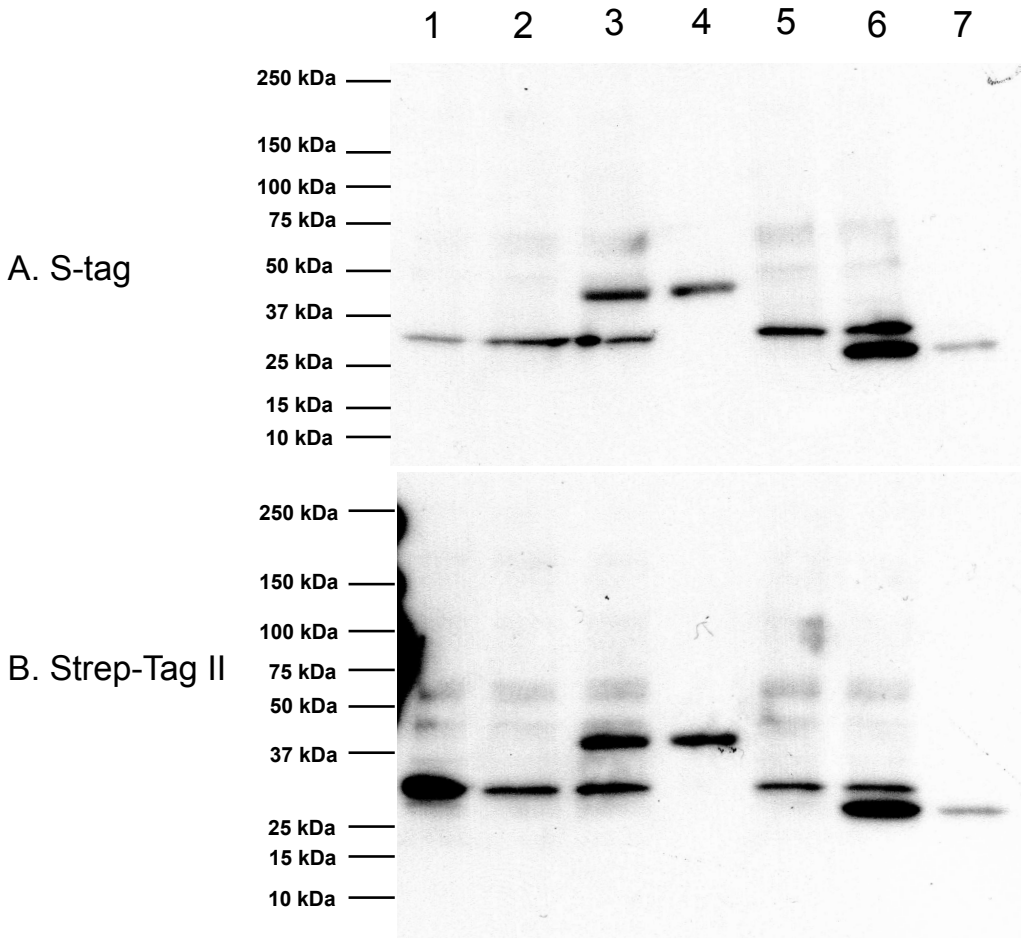

Figure 29. Immunoblot of protease-mediated cleavage products. mTFP1::C-TAP tag FP, translated from a $T7^{PROM}$-containing amplicon of the FP CDS in pNH026, was incubated with 10 units of either PreScission protease or HRV 3C protease , for 16 hr at 4 °C, to remove the S-tag and expose the Strep-Tag II. Blot A was probed using an anti-S-tag monoclonal antibody and protein on blot B with the anti-Strep-Tag II monoclonal antibody as the primary antibody. Protein samples were loaded on each gel in the same order. FP (lane 1), FP incubated with the cleavage buffer for PreScission and HRV 3C protease (lanes 2 and 5 respectively), PreScission and HRV 3C protease only (lanes 4 and 7 respectively), FP incubated with PreScission protease in the presence of the recommended buffer (lane 3) and FP incubated with HRV 3C protease with the recommended cleavage buffer (lane 6).

3.3.5 Epifluorescent filter sets: choice and properties

The filter sets selected for visualizing mCerulean, CFP and mTFP1, and YFP and mCitrine, and mCherry were all wide filter sets to ensure maximum signal from the emission of each FP was captured. However, due to the wide emission wavelength of the FPs and wide emission filters selected, bleed through (originating from the FP signal which overlaps that of the FP being imaged) of fluorescent signal into the other

filter sets presented a problem. To determine the degree of bleed through worms transgenic for constructs pNH086 (mTFP1 (2I)), pNH087 (mCitrine (2I)), pNH088 (mCherry (2I)) and pNH089 (mCerulean (2I)) were mounted on the same slide in mTFP1/mCitrine/mCherry and mCerulean/mCitrine/mCherry FP transgenic worm combinations to check for bleed through (Figure 30).

The combination of mCerulean, mCitrine and mCherry does not require spectral unmixing as there is no bleedthrough of the different FPs with the combination of filters used (Figure 30, Panels I - L). The combination of mTFP1, mCitrine and mCherry (Figure 30, Panel A), requires spectral unmixing (Figure 30, Panel C). Due to the significant spectral overlap of mTFP1 and mCitrine the FP emission of mTFP1 could be visualized using the wide YFP filter set (Figure 30, Panel C). To spectrally distinguish the two, reference images of all the FPs to be images were captured (images captured were unsaturated and only of one FP) and from the spectral profiles of each FP in these images, the signals of mTFP1 and mCitrine could successfully be spectrally unmixed in images which contained both mTFP1 and mCitrine (Figure 30, Panel G).

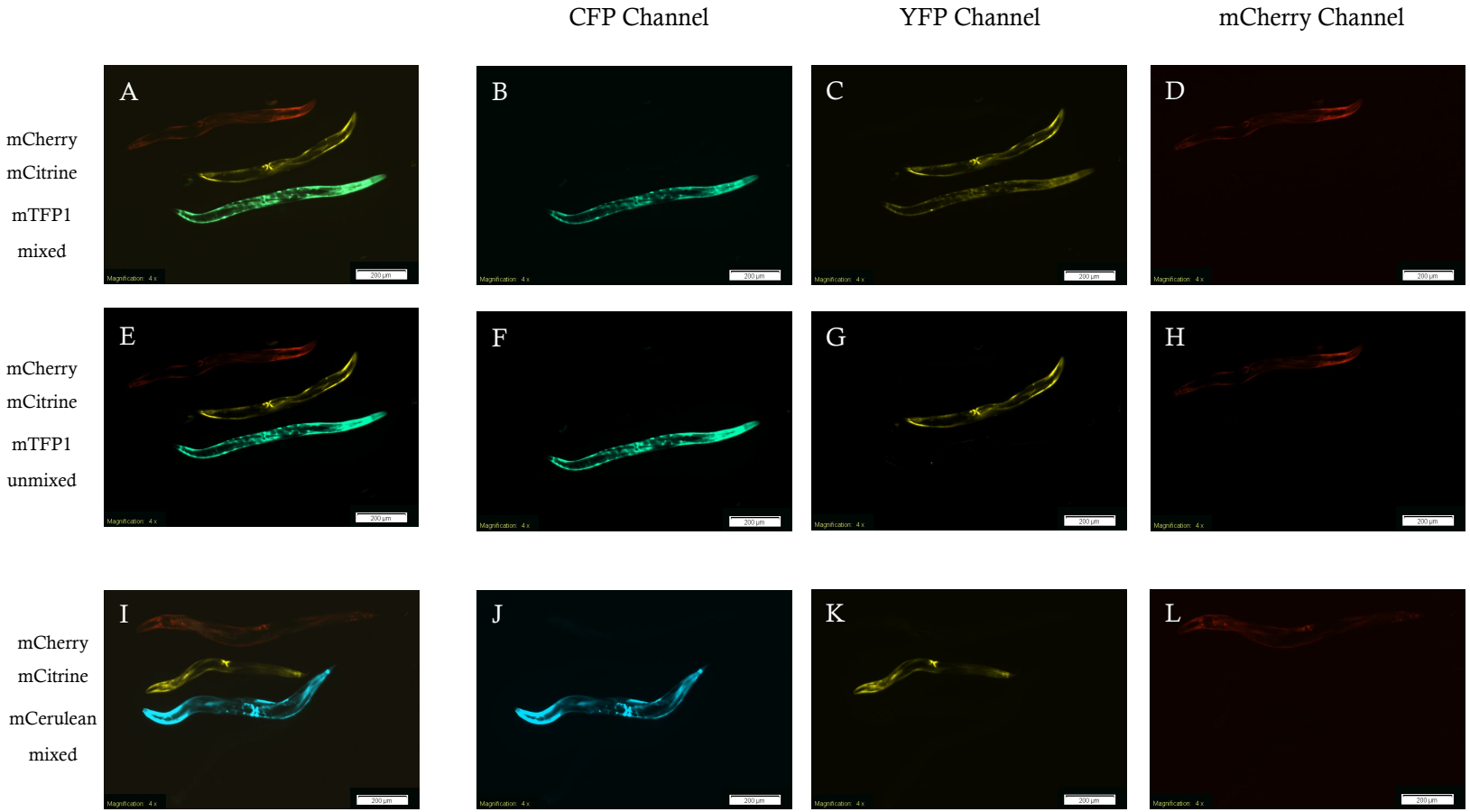

Figure 30. Spectral unmixing of mCerulean, mTFP1, mCitrine and mCherry. The images illustrate the spectral bleed through when imaging FP combinations of mCerulean, mCitrine, mCherry and mTFP1, mCitrine and mCherry. Transgenic strains expressing the respective FPs were mounted closely together on an agarose pad and imaged via the single channels, CFP, YFP and mCherry (panels B – D and J - L) and resulting images superimposed (panels A and I). mTFP1 bleeds through into the YFP channel (Panel C) however upon spectral unmixing, the mTFP1 signal can be distinguished from the mCitrine (Panel G). There is no cross talk between mCitrine and mCherry (Panel D). There is also no bleed through when using the FP combination of mCerulean, mCitrine and mCherry (Panels J – L). Scale bar indicates 200 μm.

3.4 Discussion

Epifluorescent visualization of each immunoaffinity captured FP, synthesized via an *in vitro* coupled transcription-translation system, confirmed that each respective de novo designed, commercially synthesized CDS successfully encoded a functional FP (Figure 23). Additional results in the form of the *myo-3*^{PROM} driven FP expression illustrated that transcription and translation from the codon-optimized *FP* CDSs resulted in functional protein which fluoresced and appeared to have no toxic affect to the nematodes. The *FP::C-TAP tag::2xNLS* CDSs successfully encoded for FPs that were visualized in the nuclei of muscle cells in transgenic worms when expression was driven by *myo-3*^{PROM} (Figure 24). The introduction of the two artificial introns in the FP CDS resulted in a FP with an increase in brightness of mTFP1, mCitrine and mCherry by five fold (Figures 25 and 26). This may have been caused by more stable mRNA molecules with a greater half-life or it could be due to an increase in the efficiency of the translation machinery producing the protein.

Disappointingly, and irrespective of which encoding construct was examined or whether expressed as a native protein or fused to additional sequences e.g., a C-TAP tag, mCerulean demonstrated rapid unexpected photobleaching under similar excitation conditions used for the other FPs. In the time required to locate and focus a transgenic worm, the FP signal had effectively depleted rendering it impractical to use as a fluorescent reporter. Unfortunately the introduction of the two artificial introns did not increase the signal of our codon-optimized mCerulean as it still rapidly bleached when imaged with our microscope setup. The reason(s) for the rapid fluorescence decay seen is not understood as the amino acid sequence translated from our synthetic codon-optimized mCerulean CDS is the same as the published mCerulean sequence. Further versions of cyan-shifted FPs have been generated, such as mTurquoise2 (Goedhart et al., 2012) and mCerulean3 (Markwardt et al., 2011), which are less likely to suffer with fluorescence instability as in both studies targeted residues were mutated and the resulting FPs had higher quantum yields. As the quantum yield of a FP is representative of the efficiency of the fluorescence process the inherent "brightness" of the FP is higher and the rate of decay of the fluorescent signal lower.

Another observation was that the signal from the codon-optimized synthesized mCitrine and mCherry FP reporters was not as high as the F-YFP (Fire et al., 1990) and Mc-mCherry (McNally et al., 2006), as judged by the exposure time required to obtain images with equivalent fluorescence, variants examined as part of this thesis and already in common use in the worm research community. The results observed were the opposite of what was expected, as the synthesized constructs had been codon optimized for *in vivo* expression in the worm, which lead to the assumption that their expression would be enhanced as transcription would be from an optimally encoded DNA sequence. However from results obtained the lack of brightness may be due to a decrease in the expression of the codon-optimized CDSs resulting in fewer FPs produced per mRNA unit. Even though the synthesized FPs are not the same as the Fire and McNally variants when comparing the exposure times unexpectedly the codon-optimized synthesized FPs required approximately. 5 to 10 times longer exposure in comparison to F-YFP and Mc-mCherry. This was an unexpected result as the FPs selected in the yellow and red wavelength ranges are reported to have equivalent (in the case of mCherry) or higher brightness and photostability (YFP vs. mCitrine and CFP vs. mCerulean). The F-YFP and Mc-mCherry FPs are each encoded by four exons separated by three short equally spaced introns. As use of "split" CDSs has been demonstrated to generally improve heterologous gene expression (Buchman and Berg, 1988) some synthetic codon-optimized FP variants encoded by three exons that are divided by 2 short artificial intron sequences were also created. The addition of the two intron sequences in the synthetic codon-optimized FPs did increase the level of protein expression in the transgenic nematode as judged by the decrease in exposure times required for imaging the strains transgenic for Myo-3 ^{PROM} driven codon-optimized synthesized FP constructs with two artificial introns against the non-intron containing equivalent. The exposure times needed to achieve a similar fluorescent signal from the two artificial-intron containing mCitrine and mCherry to F-YFP and McNally mCherry were still approximately. five times longer (50 – 100 msec). Out of the four synthesized FPs, mTFP1 proved to be bright and photostable requiring short exposure times (for the two intron containing variant 2 – 5 msec), which are comparable with the exposure times used for imaging F-CFP, F-YFP and Mc-mCherry transgene containing nematodes.

In reference to the results obtained with mCitrine, we had expected the two intron-containing variant to possess brightness equivalent to F-YFP as mCitrine was primarily developed to be both brighter and more photostable in comparison to the properties of standard YFP. The fluorescent signal observed from our synthetic codon-

optimized mCherry was another unexpected result. The Myo-3^{PROM} driven expression of the synthetic variant containing the two artificial intron sequences did not yield a fluorescent signal that was comparable, using the same exposure time and imaging settings, to that which was produced by the muscle driven expression of Mc-mCherry. The two mCherry FPs have the same amino acid sequence and both have been codon optimized for *C. elegans* expression. Interestingly, the design of the mCherry CDS described by McNally et al. (2006) involved a certain amount of manual intervention in codon choice (J. Audhya, pers. comm.) reinforcing the concept that the design of 'optimal' coding sequences depends upon theoretical and empirical observations.

It is difficult to define the exact reasons why the fluorescent signal from the synthetic codon-optimized FPs are lower than the signals from the reporters already available as numerous factors play a role in determining the brightness of a FP reporter. These factors include the efficiency of the transcription and translation of the DNA sequences encoding the FPs and also different copy numbers in the transgenic array. One possible way to improve the transcription and translation of the FPs could be to introduce a third artificial intron into the coding sequences as there is still an additional blunt in-frame restriction site in the coding sequences for each FP. The distribution of the introns may also have played a role, as the introns in the synthetic FPs are not equally spaced out. An additional point is that, when the coding sequences for the synthetic codon-optimized FPs were designed, we should not have assumed that codon optimizing our sequences for expression in *C. elegans* would result in optimized protein expression as it has been not always been found to do so (Gustafsson et al., 2012). A possible explanation for this is tRNA over-expression which results from a limiting supply of other cellular components when the tRNA alone is over-produced. An alternative explanation for the lack of expression can be that the designed genes contain rarely used codons, in *C. elegans*, and hence the heterologous protein is expressed at low levels. This low level of expression can also be exacerbated of the rare codons are clustered or at the N-terminal part of the protein (Gustafsson et al., 2004).

As FPs mTFP1, mCitrine and mCherry can be spectrally distinguished with the aid of software spectral unmixing (Figure 30) this would enable us to use a combination of three out of the four codon-optimized synthetic FPs to further investigate the selected operon structures within the *C. elegans* genome.

4

Generation and validation of counter-selection recombineering resources

4.1 Introduction

The counter-selection recombineering technique allows precise modifications to be made to constructs maintained at a low copy number. The *C. elegans* fosmid library is comprised of clones containing approximately. 40 kb of genomic sequence and covers approximately. 80% of the genome. Fosmid clones are ideal genomic clones for creating translational reporters, as they allow the inclusion of any regulatory elements in the 5′, 3’ and intronic sequences of a gene of interest. To facilitate the generation of multiple fosmid-based reporters in parallel via counter-selection recombineering, using the RT-Cassette, and introduce FP CDSs at specific locations within the fosmid clone a second resource set is created to complement a more streamlined recombineering protocol.

4.2 Methods

See sections 2.5.2 (generation of pNH034: a pCC1FOS-based RT-Cassette containing construct), 2.5.3 (generation of pNH034-based sub-clones), 2.5.4 (introducing F-CFP, F-GFP, F-YFP and Mc-mCherry CDSs into pNH034), 2.5.5 (identification of fosmid target) and 2.5.6 (construction of single- and multiple-tagged fosmid-based reporters via iterative counter-selection recombineering).

4.3 Results

4.3.1 Fosmid-based reporter construction

Here I describe a streamlined approach, and generated the resources to accompany this approach, for counter-selection recombineering utilizing the RT-Cassette to create FP fusion gene reporter constructs. By using smaller cultures, and shorter incubation times the method is simpler, faster and easier and also allows multiple recombineering reactions to be prepared for and carried out in parallel. As a subset of these resources were to include the RT-Cassette, and as we have data indicating that the RT-Cassette is unstable when propagated in a high copy number vector (I. A. Hope, pers comm), we introduced the RT-Cassette into the low copy number maintained vector

pCC1FOS. This particular backbone was selected as it contains the OriV-inducible origin of replication and thus allows the constructs with this backbone to be modified via counter-selection recombineering when it is present in an *E. coli* cell at single copy number. Once all modifications have been made, the OriV can be activated resulting in the copy number of the pCC1FOS construct increasing to approximately. 50 copies per cell. The inducible property of the backbone allows users to increase the yield and purity of plasmid DNA from which the required RT-Cassette PCR templates and replacement FP cassettes can be isolated and purified. *Nco*I sites were also included and flank the RT-Cassette to allow easy identification and excision of the *Nco*I fragment, containing the RT-Cassette, from the pCC1FOS backbone. The constructs which make up the recombineering resources were created in pairs with the first providing the PCR template for the FP-RT-FP Cassette and the second would provide the replacement FP CDS, excised as a single restriction fragment, to replace the FP-RT-FP Cassette.

The introduction of the synthesized codon-optimized FP CDSs and the F-CFP, F-GFP, F-YFP and Mc-mCherry CDSs, and the subsequent introduction of the RT-Cassette into the central region, flanked by approximately. 200 bp of FP gene sequence, of each FP was designed to streamline the protocol and negate the requirement for a second PCR step generating the replacement sequence. The replacement sequence is a gel-purified *Not*I restriction fragment containing the FP CDS eliminating the possibility of introducing PCR-generated errors into the final construct and hence also would reduce the need to sequence the final construct.

For the recombineering resource set a total of 42 constructs were generated (Table 7). These recombineering resources utilize the pCC1FOS backbone and were generated by introducing the RT-Cassette into this backbone. The RT-Cassette was subsequently replaced with the FP CDSs, including the Fire FPs (CFP, GFP and YFP), Mc-mCherry and the synthesized codon-optimized FPs and their variants including TAP tags and nuclear localization signals. Finally the RT-Cassette was introduced centrally into each FP CDS, in the pCC1FOS vector, leaving in most instances approximately. 200 bp of FP CDS flanking the RT-Cassette (FP-RT-FP Cassette).

The codon-optimized synthesized FP CDSs were introduced into pNH034 as a restriction fragment (*Fsp*I-*Lgu*I gel purified fragment) from the respective pGOv5 construct containing the required CDS (Table 8). The F- and Mc-mCherry FP CDSs

were PCR-amplified, from the original plasmids available as part of the Fire vector kit, and the purified PCR products were recombineered into pNH038 (Table 8).

Table 7. pCC1FOS based resources for counter-selection recombineering.

	RT-cassette				Replacement cassette		
Sequence to be inserted	RT	kb	Fwd (5′-3′)	Rev (5′-3′)	pCC1Fos-based	pGOv5-based	kb
RT-cassette	pNH034	2.0	GCTGTCGAGATATGACGGTGTTCA	TCTTGGAGTGGTGAATCCGTTAGC	-	-	-
F-CFP	pNH050	2.4	ATGAGTAAAGGAGAAGAACTTTTC	[*]TTTGTATAGTTCATCCATGCCATG	pNH039	n/a	0.9
F-GFP	pNH051	2.4	ATGAGTAAAGGAGAAGAACTTTTC	[*]TTTGTATAGTTCATCCATGCCATG	pNH040	n/a	0.9
F-YFP	pNH052	2.4	ATGAGTAAAGGAGAAGAACTTTTC	[*]TTTGTATAGTTCATCCATGCCATG	pNH041	n/a	0.9
Mc-mCherry	pNH053	2.4	ATGGTCTCAAAGGGTGAAGAAGAT	[*]GGATCCACTAGTCTTATACAATTC	pNH042	n/a	0.9
mTFP1	pNH054	2.4	ATGGCCGCCTCAAAAGGAGAAGAA	[*]AGCGCTTACGTAGAGCTCGTCCAT	pNH043	pNH013	0.7
mCitrine	pNH055	2.4	ATGGCCGCCAGTAAGGGTGAGGAG	[*]AGCGCTTACGTAGAGCTCATCCAT	pNH044	pNH014	0.7
mCherry	pNH056	2.4	ATGGCCGCCAGTAAGGGTGAGGAG	[*]AGCGCTTACGTAGAGCTCATCCAT	pNH045	pNH015	0.7
mCerulean	pNH057	2.4	ATGGCCGCCTCAAAAGGAGAAGAA	[*]AGCGCTTACGTAGAGCTCGTCCAT	pNH046	pNH016	0.7
N-TAP-tag::[$(G_4S)_3$]::mTFP1	pNH066	2.6	ATGGTTAAAGAAACAGCAGCAGCG	[*]AGCGCTTACGTAGAGCTCGTCCAT	pNH058	pNH002	0.9
N-TAP-tag::[$(G_4S)_3$]::mCitrine	pNH067	2.6	ATGGTTAAAGAAACAGCAGCAGCG	[*]AGCGCTTACGTAGAGCTCATCCAT	pNH059	pNH006	0.9
N-TAP-tag::[$(G_4S)_3$]::mCherry	pNH068	2.6	ATGGTTAAAGAAACAGCAGCAGCG	[*]AGCGCTTACGTAGAGCTCATCCAT	pNH060	pNH007	0.9
N-TAP-tag::[$(G_4S)_3$]::mCerulean	pNH069	2.6	ATGGTTAAAGAAACAGCAGCAGCG	[*]AGCGCTTACGTAGAGCTCGTCCAT	pNH061	pNH008	0.9
mTFP1::[$(G_4S)_3$]::C-TAP-tag::2xNLS	pNH070	2.6	ATGGCCGCCTCAAAAGGAGAAGAA	[*]AGCGCTAACTTTTCGCTTCTTCTT	pNH062	pNH017	0.95
mCitrine::[$(G_4S)_3$]::C-TAP-tag::2xNLS	pNH071	2.6	ATGGCCGCCAGTAAGGGTGAGGAG	[*]AGCGCTAACTTTTCGCTTCTTCTT	pNH063	pNH018	0.95
mCherry::[$(G_4S)_3$]::C-TAP-tag::2xNLS	pNH072	2.6	ATGGCCGCCAGTAAGGGTGAGGAG	[*]AGCGCTAACTTTTCGCTTCTTCTT	pNH064	pNH019	0.95
mCerulean::[$(G_4S)_3$]::C-TAP-tag::2xNLS	pNH073	2.6	ATGGCCGCCTCAAAAGGAGAAGAA	[*]AGCGCTAACTTTTCGCTTCTTCTT	pNH065	pNH020	0.95
mTFP1::[$(G_4S)_3$]::C-TAP-tag	pNH094	2.3	ATGGCCGCCTCAAAAGGAGAAGAA	[*]AGCCCATGAGTCCATATGCTGTCT	pNH090	pNH026	0.9
mCitrine::[$(G_4S)_3$]::C-TAP-tag	pNH095	2.3	ATGGCCGCCAGTAAGGGTGAGGAG	[*]AGCCCATGAGTCCATATGCTGTCT	pNH091	pNH027	0.9
mCherry::[$(G_4S)_3$]::C-TAP-tag	pNH096	2.3	ATGGCCGCCAGTAAGGGTGAGGAG	[*]AGCCCATGAGTCCATATGCTGTCT	pNH092	pNH028	0.9
mCerulean::[$(G_4S)_3$]::C-TAP-tag	pNH097	2.3	ATGGCCGCCTCAAAAGGAGAAGAA	[*]AGCCCATGAGTCCATATGCTGTCT	pNH093	pNH029	0.9
mTFP1::[$(G_4S)_3$]::2xNLS	pNH070	2.6	ATGGCCGCCTCAAAAGGAGAAGAA	[*]AGCGCTAACTTTTCGCTTCTTCTT	n/a	pNH030	0.8
mCitrine::[$(G_4S)_3$]::2xNLS	pNH071	2.6	ATGGCCGCCAGTAAGGGTGAGGAG	[*]AGCGCTAACTTTTCGCTTCTTCTT	n/a	pNH031	0.8
mCherry::[$(G_4S)_3$]::2xNLS	pNH072	2.6	ATGGCCGCCAGTAAGGGTGAGGAG	[*]AGCGCTAACTTTTCGCTTCTTCTT	n/a	pNH032	0.8
mCerulean::[$(G_4S)_3$]::2xNLS	pNH073	2.6	ATGGCCGCCTCAAAAGGAGAAGAA	[*]AGCGCTAACTTTTCGCTTCTTCTT	n/a	pNH033	0.8
mTFP1[B]	pNH054	2.4	ATGGCCGCCTCAAAAGGAGAAGAA	[*]AGCGCTTACGTAGAGCTCGTCCAT	n/a	pNH078	0.8
mCitrine[α]	pNH055	2.4	ATGGCCGCCAGTAAGGGTGAGGAG	[*]AGCGCTTACGTAGAGCTCATCCAT	n/a	pNH079	0.8
mCherry[β]	pNH056	2.4	ATGGCCGCCAGTAAGGGTGAGGAG	[*]AGCGCTTACGTAGAGCTCATCCAT	n/a	pNH080	0.8
mCerulean[γ]	pNH057	2.4	ATGGCCGCCTCAAAAGGAGAAGAA	[*]AGCGCTTACGTAGAGCTCGTCCAT	n/a	pNH081	0.8
mTFP1[BC]	pNH054	2.4	ATGGCCGCCTCAAAAGGAGAAGAA	[*]AGCGCTTACGTAGAGCTCGTCCAT	n/a	pNH082	0.8
mCitrine[αD]	pNH055	2.4	ATGGCCGCCAGTAAGGGTGAGGAG	[*]AGCGCTTACGTAGAGCTCATCCAT	n/a	pNH083	0.8
mCherry[βE]	pNH056	2.4	ATGGCCGCCAGTAAGGGTGAGGAG	[*]AGCGCTTACGTAGAGCTCATCCAT	n/a	pNH084	0.8
mCerulean[γF]	pNH057	2.4	ATGGCCGCCTCAAAAGGAGAAGAA	[*]AGCGCTTACGTAGAGCTCGTCCAT	n/a	pNH085	0.8

Table 8. pCC1FOS based constructs

Construct	Insert	Source	RT-cassette	ODNs
pNH043	mTFP1	pNH013	pNH054	12042,12043
pNH044	mCitrine	pNH014	pNH055	12044,12045
pNH045	mCherry	pNH015	pNH056	12046,12047
pNH046	mCerulean	pNH016	pNH057	12048,12049
pNH058	N-TAP-tag::mTFP1	pNH002	pNH066	12042,12043
pNH059	N-TAP-tag::mCitrine	pNH006	pNH067	12044,12045
pNH060	N-TAP-tag::mCherry	pNH007	pNH068	12046,12047
pNH061	N-TAP-tag::mCerulean	pNH008	pNH069	12048,12049
pNH062	mTFP1::C-TAP-tag::2xNLS	pNH017	pNH070	12042,12043
pNH063	mCitrine::C-TAP-tag::2xNLS	pNH018	pNH071	12044,12045
pNH064	mCherry::C-TAP-tag::2xNLS	pNH019	pNH072	12046,12047
pNH065	mCerulean::C-TAP-tag::2xNLS	pNH020	pNH073	12048,12049
pNH090	mTFP1::$[(G_4S)_3]$::C-TAP-tag	pNH026	pNH094	12042,12043
pNH091	mCitrine::$[(G_4S)_3]$::C-TAP-tag	pNH027	pNH095	12044,12045
pNH092	mCherry::$[(G_4S)_3]$::C-TAP-tag	pNH028	pNH096	12046,12047
pNH093	mCerulean::$[(G_4S)_3]$::C-TAP-tag	pNH029	pNH097	12048,12049
pNH039	F-CFP	12028/12029	pNH050	12030/12031
pNH040	F-GFP	12028/12029	pNH051	12030/12031
pNH041	F-YFP	12028/12029	pNH052	12030/12031
pNH042	Mc-mCherry	12026/12027	pNH053	12032/12033

To validate the recombineering resources generated and the variants of codon-optimized FP CDSs in the context of being utilized in translational reporters, the contiguous FP CDS and two artificial intron containing FP CDS were seamlessly recombineered into a single fosmid clone containing three centrally-located genes with previously published expression pattern data. The fosmid clone, WRM069dD11, was identified and the FP CDSs of mTFP1, mCitrine and mCherry were introduced respectively into *F09E5.3*, *F09E5.15* and *EEED8.6* (Figure 31). To compare the codon-optimized synthesized FPs, the readily used, within the worm community, F-CFP, F-YFP and Mc-mCherry CDSs were recombineered into *F09E5.3*, *F09E5.15* and *EEED8.6* five codons upstream of the native genes stop codon. Due to an oversight when designing the reverse recombineering ODNs to introduce the synthesized codon-optimized FP CDSs, one of the stop codons of the FP was included (two stop codons were included in the design of the codon-optimized FP CDSs) and therefore the genes tagged with the synthesized codon-optimized FP CDSs were missing the last 5 codons of the endogenous gene.

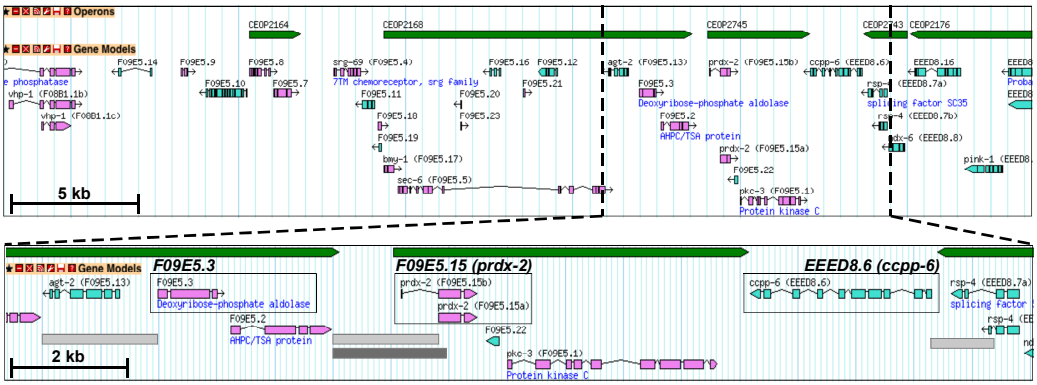

Figure 31. Genomic location of fosmid clone WRM069dD11. The upper panel illustrates the genomic region contained on the fosmid and the lower panel highlights the position of the three genes. The solid green arrows at the top depict operon structures, forward genes (purple) and genes on the reverse strand (green). The lower panel is an expansion of the region, which contains the three genes of interest (the three genes are boxed in grey). The light grey bars represent the location used as "promoter" region in the transcriptional expression patterns (BC14910, BC13145 and BC11803 respectively) obtained by (McKay et al., 2003; Hunt-Newbury et al., 2007). The dark grey bar under *F09E5.15a* represents the promoter region used by Isermann et al. (2004).

The FP-RT-FP Cassette of the corresponding reporter sequence was introduced into the respective target gene in fosmid clone WRM063dD11 via positive selection recombineering (Figure 31 and Table 3). Subsequently the FP-RT-FP Cassette was replaced with the reporter sequence via negative selection recombineering. Construction integrity of a fosmid clone containing the desired FP reporter sequence, identified initially by colony-PCR and RE incubation, was confirmed by DNA sequencing.

The efficiency of the positive selection step, when utilizing the RT-FP-RT Cassettes created as part of the recombineering resource generated, is the same as previously published (Dolphin and Hope, 2006). However the efficiency of the negative selection step, when using a *Not*I restriction fragment containing the required FP CDS, is 100-fold more efficient compared to performing the negative selection step as previously described (Dolphin and Hope, 2006).

Once all the fosmid-based translational reporters had been generated and sequence confirmed (more than 98 % of all constructs sequenced were correct), all the single and triple-tagged fosmid-based constructs were microinjected into N2s (dual-tagged fosmid constructs were not used as all expression patterns for the genes within the operon could be obtained from the single and triple-tagged fosmid constructs). A

single gravid worm from each resulting rolling transgenic line was lysed and PCR screened to confirm the presence, within the heritable array, of the reporter fusion gene within each of these lines by PCR, using ODNs that flank the insertion point (Figures 32 and 33). Transgenic strains found to contain the reporter fusion gene will have two PCR products (Figure 33, lane 1 and Figure 34, lanes 1, 6 and 11), one from the reporter fusion gene (which should be the biggest) and one from the native gene within the genome (the smaller fragment). Lines found to contain the modified fosmid insert were imaged.

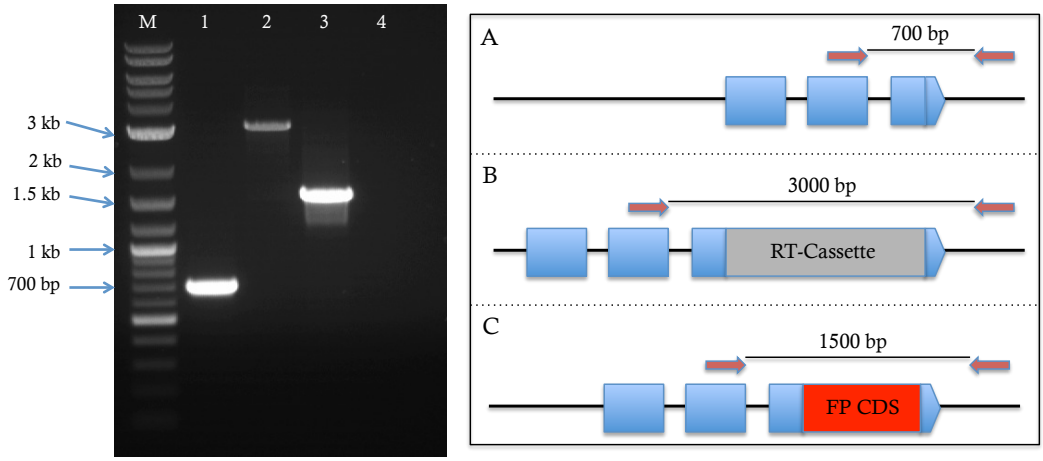

Figure 32. PCR-based identification of intermediate and final products during construction of recombineering-mediated fosmid-based reporters. An agarose gel illustrating the typical PCR fragments obtained when creating a fosmid-based reporter construct and a schematic (not to scale) depicting the annealing location of the flanking ODNs (red arrows above genes (blue boxes)) when amplifying unmodified fosmid (Panel A), RT-Cassette-containing fosmids (Panel B) and FP CDS-containing fosmid (Panel C). The agarose gel illustrates PCR products generated using flanking ODNs from fosmid DNA that was unmodified (lane 1 and Panel A), or contained either a RT-cassette-based counter-selection sequence (lane 2 and Panel B) or FP-encoding replacement sequence (lane 3 and Panel C). DNA ladder (2-log ladder, NEB, lane M) and negative (no template) control (lane 4).

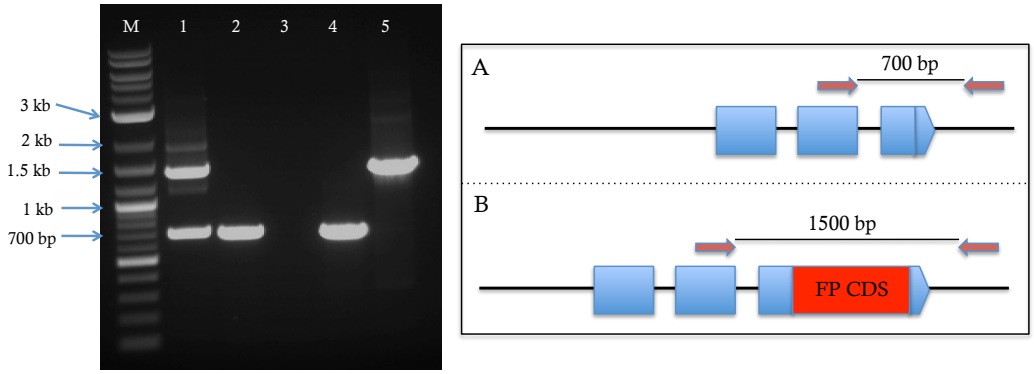

Figure 33. Agarose gel image and schematic exhibiting the PCR products obtained when screening transgenic lines for the presence of a single-tagged gene fosmid construct. A typical agarose gel image of the PCR products visualized when screening a single gravid worm from each transgenic line to confirm the presence of a single fusion gene and schematic. PCR products generated using flanking ODNs from a transgenic strain containing the single fusion gene and the native gene within the genome (lane 1) and from wild-type worms (lane 2). DNA ladder (2-log ladder, NEB, lane M), negative (no template) control (lane 3), unmodified fosmid DNA (lane 4) and fosmid containing the fusion gene, which was microinjected to create the transgenic strain (lane 5). The schematic (not to scale) illustrates the position of the annealing ODNs (red arrows above the gene sequence) when screening the native gene (Panel A) and a fusion gene (Panel B).

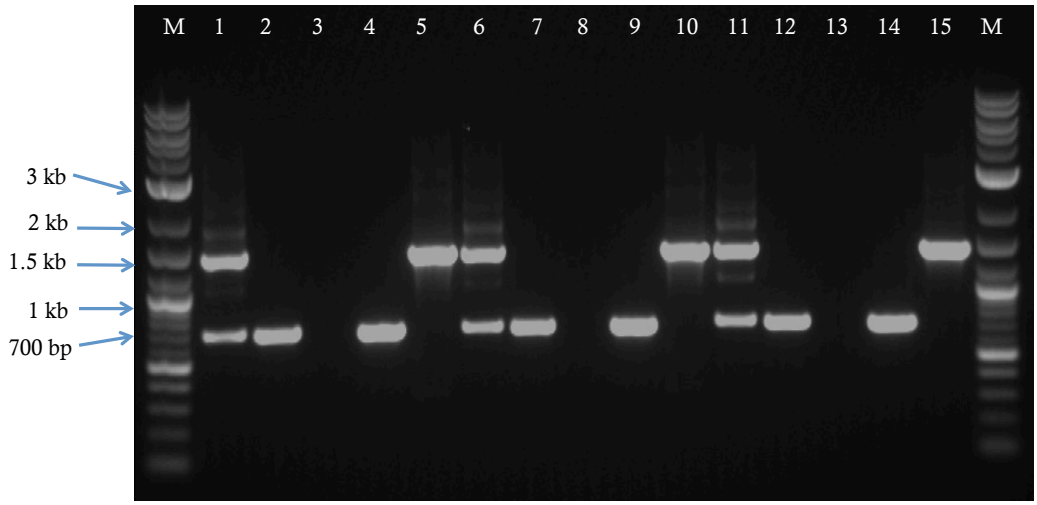

Figure 34. Agarose gel image of the PCR fragments obtained when screening transgenic lines for the presence of a triple gene tagged fosmid-based construct. A typical agarose gel picture of the PCR products obtained when screening single gravid worms from each transgenic line to confirm the presence of the multiple fusion genes. PCR products generated using flanking ODNs, to each fusion gene, from a transgenic strain containing the multiple fusion genes (lanes 1, 6 and 11) and wild type worms amplified with the same ODNs (lanes 2, 7 and 12). DNA ladder (2-log ladder, NEB, lane M), negative (no template) control (lanes 3, 8 and 13), unmodified fosmid DNA (lanes 4, 9 and 14) and fosmid containing the fusion gene, which was microinjected to create the transgenic strain (lanes 5, 10 and 15).

The three genes, on the single fosmid, were selected for the validation of the recombineering resources as each gene has published expression pattern data and as a comparison the corresponding transcriptional reporter transgenic strains for each previously published expression patterm for each gene were also imaged in house. Each of the three transgenic strains (transgenic for transcriptional reporters) were generated during the high-throughput expression pattern studies (McKay et al., 2003; Hunt-Newbury et al., 2007) and were transformed with plasmid constructs containing the promoter region (Figure 31) of each gene driving GFP expression. Strains BC14910, BC13145 and BC11803, transgenic for the promoters of *F09E5.3*, *F09E5.15* and *EEED8.6*, respectively, were obtained from CGC. Additional expression patterns for *F09E5.15* (*prdx-2*) were also available (Isermann et al. 2004 and Olahova et al. 2008).

Images obtained when visualizing strains BC14910, BC13145 and BC11803 (Figure 35) demonstrated via transcriptional reporters that *F09E5.3* was expressed in the hypodermis and intestine. *F09E5.15* (*prdx-2*) was expressed in the intestine, in head neurons and also in unidentified tail cells. Isermann et al. (2004) previously demonstrated PRDX-2 expression in two interneurons (I4, pharyngeal and I2 sensory). Olahova et al. (2008) demonstrated subsequently, via an immunohistochemical approach, that PRDX-2 was also expressed in gonad and intestine. GFP expression in strains transgenic for a transcriptional *EEED8.6*^{PROM}*::gfp* reporter construct (McKay et al., 2003; Hunt-Newbury et al., 2007) (Figure 31), was visualized in head neurons. This expression pattern was in agreement with that reported previously (Isermann et al. 2004).

4.3.2 Translational reporter expression patterns

Expression pattern data from the fosmid-based constructs with both contiguous codon-optimized synthesized FP CDS (Figure 36) and codon-optimized synthesized FP CDS containing two artificial introns (Figure 37), yielded disappointing results as brightness was "poor" which was assumed to be due to low expression levels of the FP (even with high exposure times, fluorescence could hardly be detected). The only expression pattern which could be deciphered was that for *F09E5.3*, which was tagged with mTFP1, irrespective of whether it was encoded by the contiguous (Figure 36) or 2-intron-containing CDS (Figure 37), was

bright and showed expression in the intestine and also the hypodermis. The expression patterns visualized from the fosmid constructs containing genes tagged with F-CFP, F-YFP and Mc-mCherry (Figure 38) suggested that F09E5.3 was expressed in the hypodermis (Figure 38, panels B and D) and also in the intestine and EEED8.6 was expressed in the CEP sensory (mechanosensory) neurons (Figure 38, panels J and L). However, the expression pattern obtained for F09E5.15 was unusual as it produced what appears to be an aggregation pattern (Figure 38, panels F, H and N). Upon investigation via confocal imaging, no structures could be visualized under the F-YFP aggregates.

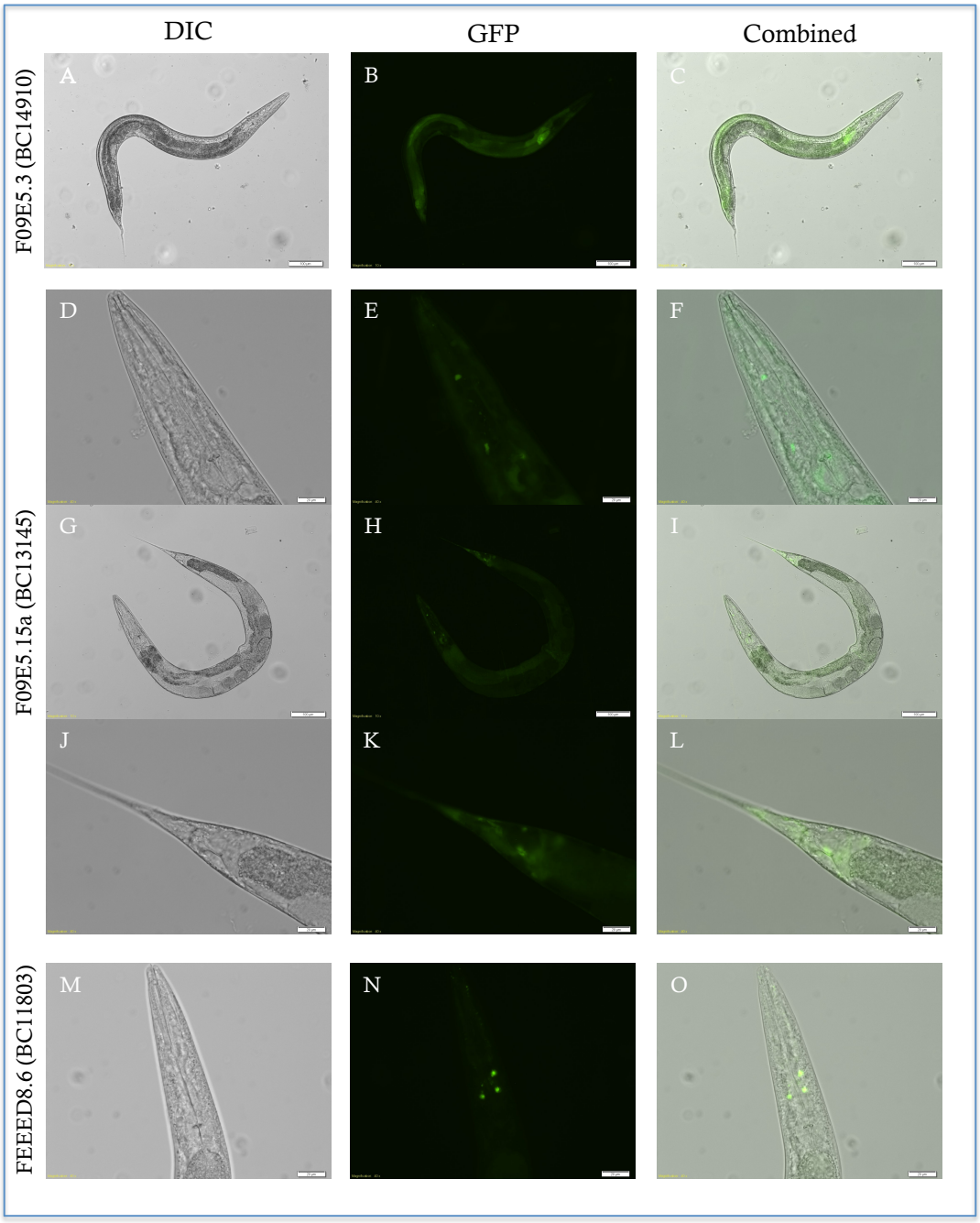

Figure 35. Expression patterns of F09E5.3, F09E5.15 and EEED8.6 derived from transcriptional reporters. Transgenic strains consisting of transcriptional GFP promoter constructs for genes *F09E5.3* (BC14910, panels A - C), *F09E5.15* (*prdx-2*, BC13145, panels D - L) and *EEED8.6* (*ccpp*-6, BC11803, panels M - O) were imaged. Expression of GFP driven by *F09E5.3*^{PROM} was seen in the hypodermis and intestinal cells. The promoter region of *F09E5.15* (*prdx-2*), utilized in the transcriptional reporter, directed expression of GFP into intestinal cells, head neurons and also in unidentified tail cells. Finally expression of *EEED8.6*^{PROM} fused to GFP was seen in discrete head neurons. Scale bar for panels A-C and G-I are at 100 μm and for panels D-F and J-O are at 20 μm.

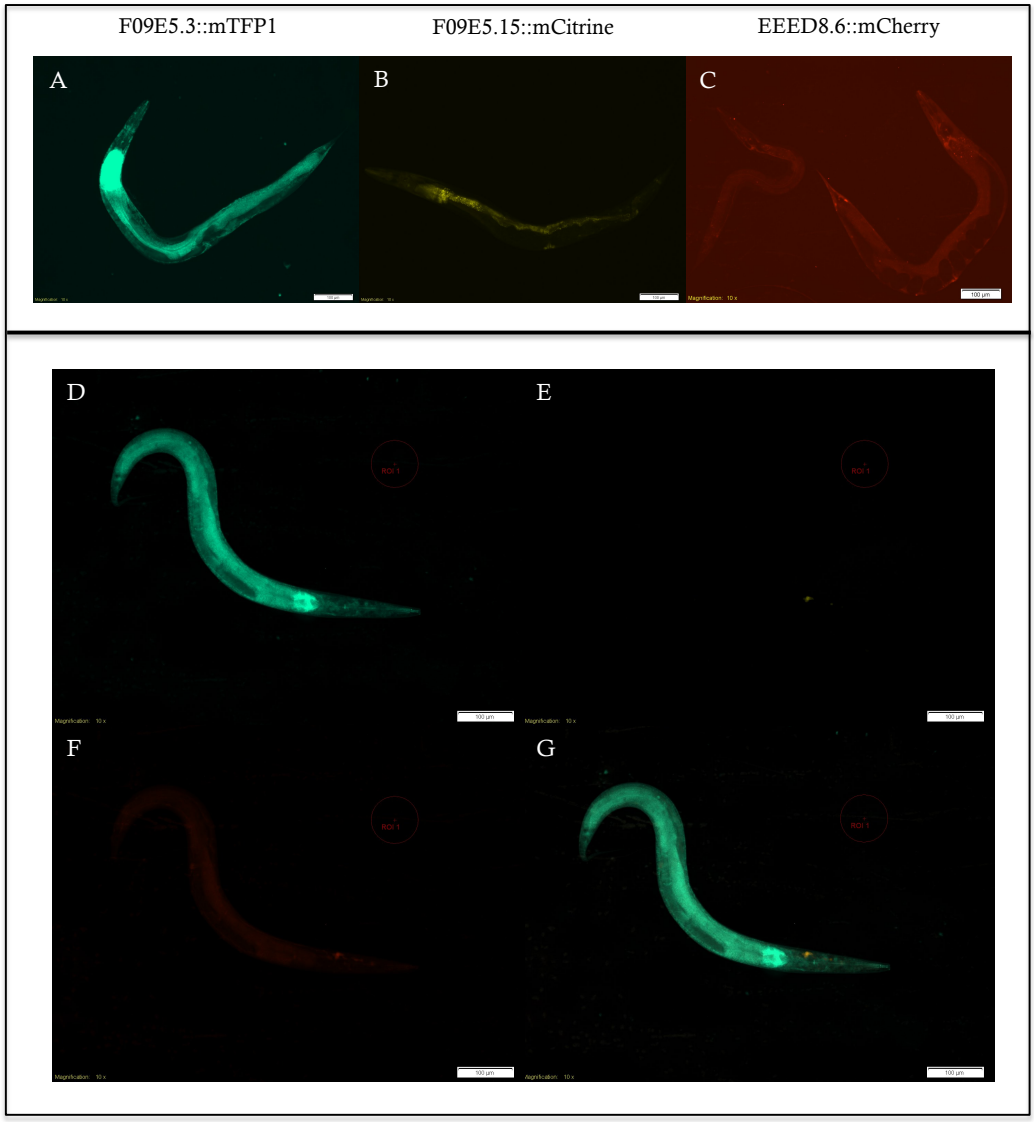

Figure 36. Fluorescence imaging depicting expression patterns of F09E5.3::mTFP1, F09E5.15::mCitrine and EEED8.6::mCherry from fosmid-based translational reporters. Fosmid constructs containing a singularly tagged (panels A – C) and multiple tagged genes (panels D – G) with the codon-optimized synthetic fluorescent reporters were microinjected into the gonad of gravid N2s. The resulting independently generated transgenic lines were imaged. *F09E5.3* (tagged with mTFP1) was visualized in the hypodermis and also in the intestine. Expression of F09E5.15 (tagged with mCitrine) and EEED8.6 (tagged with mCherry) showed very faint expression in the intestinal (granular in appearance) and faint expression in the vulva. The expression pattern obtained from a transgenic line for the multiple-tagged fosmid construct is illustrated in panels D – G. The first row of images is through each individual channel, after spectral unmixing, the composite of all three is shown underneath besides the DIC image. Scale bar for all images is 100 μm.

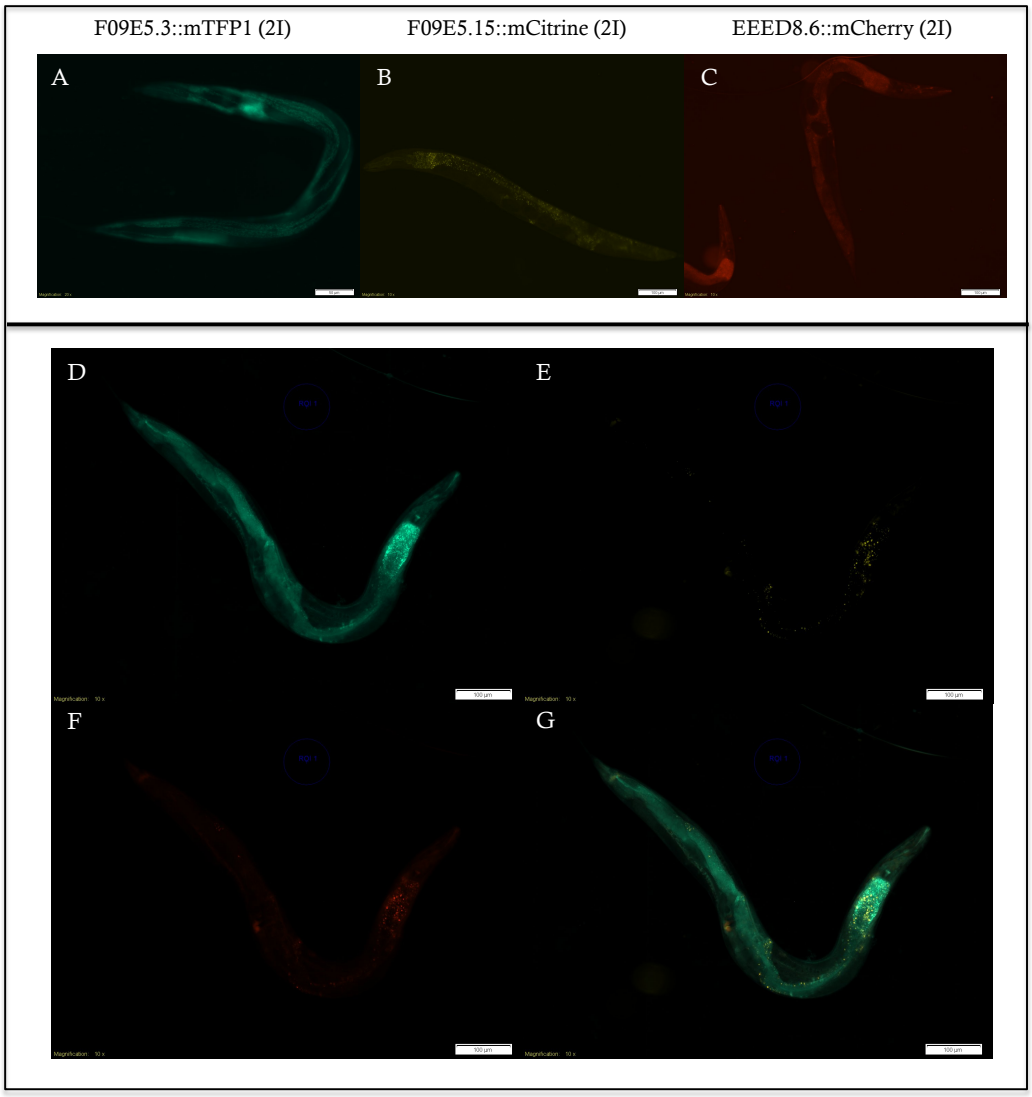

Figure 37. Fluorescence microscopy illustrating expression patterns for F09E5.3::mTFP1(2I), F09E5.15::mCitrine(2I) and EEED8.6::mCherry(2I) from fosmid-based translational reporters. Fosmid constructs containing either a single-tagged (panels A - C) or multiple-tagged genes (panels D - G) with the codon-optimized synthetic FPs containing two artificial introns were microinjected into the syncytium in the gonad of gravid wild type worms. The resulting independently-generated transgenic lines were imaged and revealed to show expression of mTFP1, fused in-frame to F09E5.3, in the hypodermis and also in intestinal cells. However, F09E5.15 fused to mCitrine, and EEED8.6 fused to mCherry showed no observable expression patterns. The typical expression patterns observed from a transgenic line for the multiply-tagged fosmid reporter are exhibited in panels D - G. The first row of images (panels D – F) is through each individual channel, after spectral unmixing and the composite of all three is shown (panel G). Scale bar for all images is 100 μm.

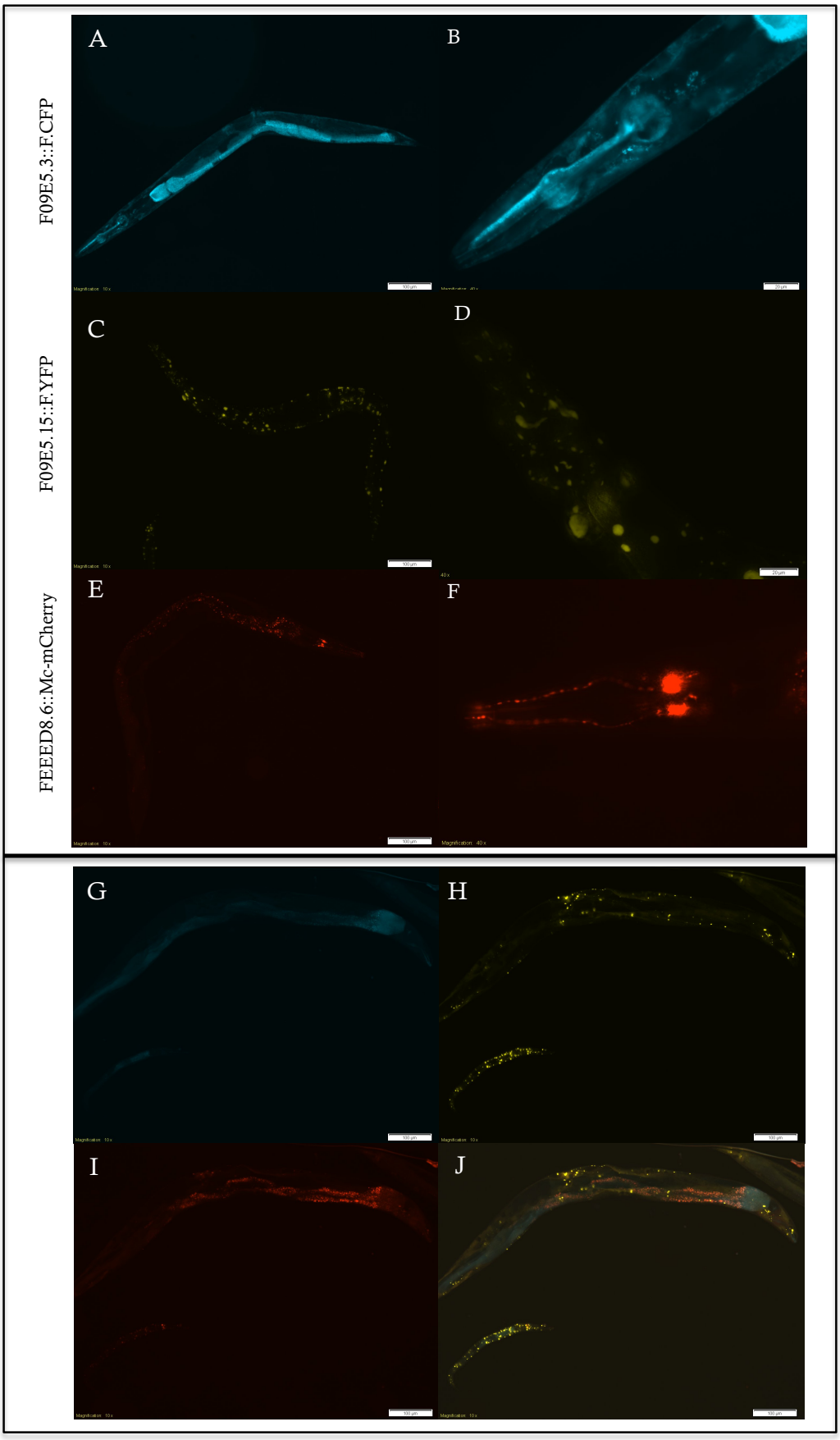

Figure 38. Fluorescence microscopy of expression patterns for F09E5.3::F-CFP, F09E5.15::F-YFP and EEED8.6::Mc-mCherry from fosmid-based translational

reporters. Adult hermaphrodites transgenic for fosmid constructs containing either single-tagged (panels A – F) or multiple-tagged (panels G – J) genes were imaged. Expression of F09E5.3, tagged with F-CFP, was observed in the hypodermis and intestine and expression of EEED8.6, tagged with Mc-mCherry, was seen in head neurons, specifically the four cephalic neurons (2 neurons can be seen in panel F). F09E5.15, tagged with F-YFP, gave an expression pattern of punctate, intense spots distributed through out the body. Scale bar for panels A - F are 100 μm and panels G - J is 20 μm.

4.4 Discussion

A second set of resources was designed in conjunction with the first set, detailed in Chapter 3, with the aim of speeding up and simplifying the counter-selection recombineering method that utilizes the RT-Cassette. As a number of constructs within the resource set created here contains the RT-Cassette, pCC1FOS was chosen as the backbone of this resource collection as it was discovered that the RT-Cassette was unstable when propagated in a high copy number vector (I. A. Hope, pers comm). pCC1FOS was thus selected as the backbone as it is an inducible backbone that is maintained at a low copy number allowing manipulations to constructs containing the RT-Cassette to be performed. Upon induction, via activation of the second origin of replication, OriV ('Copy control', Epicentre), the copy number would increase from low (1 – 2 copies) to approximately. 50 copies per *E. coli* cell to improve DNA purity and yield of the resulting constructs. *Nco*I sites were also included and flanked the RT-cassette to allow easy identification and excision of the *Nco*I-*Nco*I fragment containing the counter-selection cassette from the rest of pNH034. The subsequent constructs, which were created from pNH034, were generated in pairs as the first would provide the template for PCR amplification of the FP-RT-FP Cassette and the second provides the replacement sequence which could be excised from its parent construct as a restriction fragment (FP CDSs are flanked by *Not*I). The RT-Cassette was introduced centrally into the FP CDSs, in pCC1FOS backbone, resulting in 200bp of the FP CDS flanking the RT-Cassette (FP-RT-FP Cassette). The addition of this flanking sequence in the FP-RT-FP Cassette template ensures that the negative-selection recombineering step is performed with high efficiency, due to the longer homology arms, generating fewer false positives which is consistent with the hypothesis that the longer the homology arms the higher the efficiency of homologous recombination (Shen and Huang, 1986; Watt et al., 1985). The increase in efficiency also results in the need for fewer

bacterial cells to be electroporated, reducing culture volumes and incubation times speeding up the process. The increase in efficiency is particularly important for the negative selection step as there is always a little background and the increase in efficiency will reduce the background in comparison to the number of colonies containing a successfully generated recombinant.

Having generated this pCC1FOS-based resource set, and streamlined the recombineering protocol utilizing the RT-Cassette, a selection of these resources were used to build a series of fosmid-based translational reporter constructs via iterative rounds of recombineering. The translational reporters were all built from the same fosmid clone, WRM069dD11, which included the genes *F09E5.3*, *F09E5.15* (*prdx-2*) and *EEED8.6* (*ccpp-6*) that encode, respectively, a putative deoxyribose-phosphate aldolase (DERA), a peroxiredoxin and a metallocarboxypeptidase. The iterative rounds of recombineering generated single, double and triple-tagged fosmid constructs for mTFP1(2I)/mCitrine(2I)/mCherry(2I), mTFP1/mCitrine/mCherry and F-CFP/F-YFP/Mc-mCherry FP combinations. Transgenic lines were generated for the single- and triple- tagged constructs and the resulting expression patterns from the fosmid-based translational reporters investigated (Figures 36, 37 and 38). Examination of the resulting transgenic lines revealed that the codon-optimized synthesized FPs, with the exception of mTFP1, are not bright enough to be used in translational reporter constructs. In both cases where the codon-optimized synthesized FPs were seamlessly inserted, with and without the addition of the two artificial introns, the expression patterns as visualized with our imaging system, were not bright enough to determine an expression pattern for the gene that is was fused to. This could have been due to a number of reasons including reduced translation of FP from the codon-optimized FP CDSs designed. The decrease in fluorescence observed from the translational reporters was more pronounced in comparison to the signal visualized from the transcriptional *myo-3*^{PROM} reporters (Chapter 3). The premature stop signal in the translational reporters, due to the stop codon included in the reverse recombineering ODN for the codon-optimized synthesized FP CDS, may also have played a role in the low fluorescent signal as nonsense mediated mRNA decay of the resulting mRNA transcripts from these constructs may have occurred (reviewed by Mango, 2001). An alternative reason for the apparent reduction in fluorescent signal may have been due to the size of the fosmid DNA resulting in fewer copies of the fusion gene being incorporated into each extra-chromosomal array in the transgenic animals. Whatever the reason(s)

behind the low signal visualized for mCitrine and mCherry CDS (with and without the presence of the two artificial introns) the expression patterns from resulting nematodes transgenic for either *F09E5.15::mCitrine/mCitrine(2I)* and *EEED8.6::mCherry/mCherry(2I)* could not be deciphered and further examination of those lines was not pursued. Interestingly worms expressing mTFP1/mTFP1(2I) were consistently bright confirming that the mTFP1, encoded by the codon-optimized CDS designed, is a useful addition to the FP palette for analyzing gene expression in *C. elegans* utilizing FP reporters.

The recombineering resources generated utilizing the F-CFP, F-YFP and Mc-mCherry CDSs were successfully used to seamlessly tag *F09E5.3*, *F09E5.15* and *EEED8* with F-CFP, F-YFP and Mc-mCherry, respectively, 5 codons upstream of the stop codon. All transgenic lines derived from the single- and triple-tagged fosmid reporters exhibited bright expression patterns which could easily be interpreted. F-CFP and F-YFP are encoded by CDSs that had not been codon-optimized for expression in *C. elegans*, however these were significantly brighter than their spectral counterparts encoded by the synthetic codon-optimized FP CDSs. Only the Mc-mCherry used in these fosmid reporters is codon-optimized for *C. elegans* expression supporting the hypothesis that the design of gene sequences, containing introns, for optimal transcription and translation depends upon both examination of the theory and empirical observations.

Fosmid-based expression patterns of *F09E5.3*, *F09E5.15* and *EEED8.6* were determined by examination of worms transgenic for the single- and triple-tagged F-CFP/F-YFP/Mc-mCherry constructs. *F09E5.3* encodes DERA, part of the pentose phosphate shunt, and is responsible for catalyzing the reversible reaction of 2-deoxy-D- ribose 5-phosphate to D-glyceraldehyde 3-phosphate and acetaldehyde (Racker, 1952). One might expect it to have widespread expression which would be consistent with an enzyme that is involved in intermediate metabolism. Examination of strains transgenic with the single- and triple-tagged *F09E5.3::F-CFP* had expression of F-CFP through out the intestine, hypodermis and also in the head. Similarly worms transgenic for *F09E5.3::mTFP1/mTFP1(2I)* also exhibited an equivalent wide spread expression pattern of mTFP1 to that of F-CFP. The expression pattern visualized from the translational reporters is comparable to the previously published expression pattern of F09E5.3 via strain BC14910 which is

transgenic for transcriptional reporter *F09E5.3*^{PROM}*::GFP* (McKay et al., 2003; Hunt-Newbury et al., 2007).

F09E5.15 (*prdx-2*) encodes one of the two 2-Cys peroxiredoxins that is expressed in *C. elegans*. These conserved thioredoxin-coupled peroxidases are important in the overall oxidative-stress response employed by many multicellular organisms. There is also increasing evidence that they have additional complex roles in redox signaling as regulators of H_2O_2 (Rhee and Woo, 2011). A generalized anti-oxidant role may suggest a widespread distribution for PRDX-2 in *C. elegans*, however, previously reported expression has been more discreet suggesting that PRDX-2 may have a more specific role. Previous expression was reported, via transcriptional reporters, in pharyngeal interneurons I2 and I4 (Isermann et al. 2004) which was subsequently confirmed via immunohistochemistry which revealed additional expression in the gonad and intestine (Olahova et al. 2008). Expression patterns resulting from strain BC13145, transgenic for transcriptional reporter *F09E5.15*^{PROM}*::GFP* (McKay et al., 2003; Hunt-Newbury et al., 2007) revealed a more evenly distributed pattern including intestinal, head and tail neurons. In comparison to the expression patterns obtained for PRDX-2, via transcriptional reporters and immunohistochemistry, our transgenic strains, containing single-and triple-tagged constructs with *F09E5.15::F-YFP*, exhibited a very unusual expression that was not consistent with any of the previous reports. Expression of F-YFP was visualized as bright punctate spots distributed throughout the body and these spots did not co-localize to any specific tissue or cell type. On occasion the punctate spots appeared as paired foci (symmetrical) and if real, this could suggest an uncharacterized functional relevance. Alternatively, the expression visualized may be artifactual and represent aggregation of F-YFP, however aggregation like this from translational reporters is a rarely witnessed (I.A.H pers comm). The aggregation, visualized as 'blobs' of fluorescence, may be due to the mitochondrial targeting sequence in the 3' UTR of prdx-2 resulting in YFP aggregating in the transgenic animals.

The third gene under investigation *EEED8.6* (*ccpp-6*) encodes for one of two *C. elegans* cytosolic carboxypeptidases and has been identified as a deglutamase (Kimura et al., 2010) that mediates, co-operatively with tubulin tyrosine ligase like (TTLL) glutamylases, the deglutamylation and glutamylation of tubulin and hence regulating the microtubular network present in the neuronal sensory cilia in *C.*

elegans. A transcriptional reporter for *ccpp-6* suggested labial neuron expression in contrast to a translational reporter that suggested CCPP-6 is expressed in putative amphid cell bodies and the discrepancy between the two was explained by the possibility of including enhancer(s) in the translational construct (Kimura et al., 2010). Investigation of the nematodes transgenic for either the single or triple-tagged constructs containing *EEED8.6::Mc-mCherry* exhibited a clear and discreet expression pattern in head neurons and anterior dendrites. Further examination of the cell bodies of the head neurons identified these to likely be the putative mechanosensory CEP cephalic neurons.

In conclusion, the newly created recombineering resource, consisting of the counter-selection RT-Cassette located within the central section of a FP CDS plus corresponding intact FP CDS, built within the copy-number inducible pCC1FOS-based vector enabled the rapid and facile construction, via iterative rounds of counter-selection recombineering, of single, double and triple-tagged translational style-reporters in fosmid clones. This resource will enable others to build similar fosmid-based FP reporters in a quick and simple manner.

5

Investigating internal regulation of operon genes: use of recombineered fosmid-based reporters

5.1 Introduction

Operons within the *C. elegans* genome are a significant structure and over 17% of the genome are contained within them. These operons are gene-clusters containing between two and eight genes with short intergenic sequences (100 – 400 bp) and when transcribed produce a polycistronic mRNA unit. This mRNA then undergoes *trans*-splicing, into monocistronic mRNA units, and *cis*-splicing, to remove introns, to form mature mRNA. The *trans*-splicing of the polycistronic mRNA unit is performed by two distinct spliced-leader (SL) sequences. To investigate the expression patterns of genes contained within an operon, in the context of the actual operon structure, genomic clones can be utilized to create translational reporters. The technique of recombineering fosmid clones to introduce FP CDSs into target genes is an ideal way to investigate expression patterns of operon-genes.

5.2 Methods

See sections 2.6 (identification of operon targets), 2.6.1 (identification of gene models) and 2.6.2 (investigating expression patterns of operon-genes via fosmid-based translational reporters).

5.3 Results

5.3.1 Identification of operon targets

After taking into consideration the criteria (Method 2.6) to be used to identify potential operon targets the wormbase genome browser was scanned by eye and a list of possible operons compiled. From this list of potential operons a number were further investigated and two targets identified for dissection – CEOP1312, a three gene operon centrally located on fosmid clone WRM063bF06 (Figure 39, panel A), and CEOP1358 a two gene operon, located on fosmid clone WRM024aC07 (Figure 39, panel B).

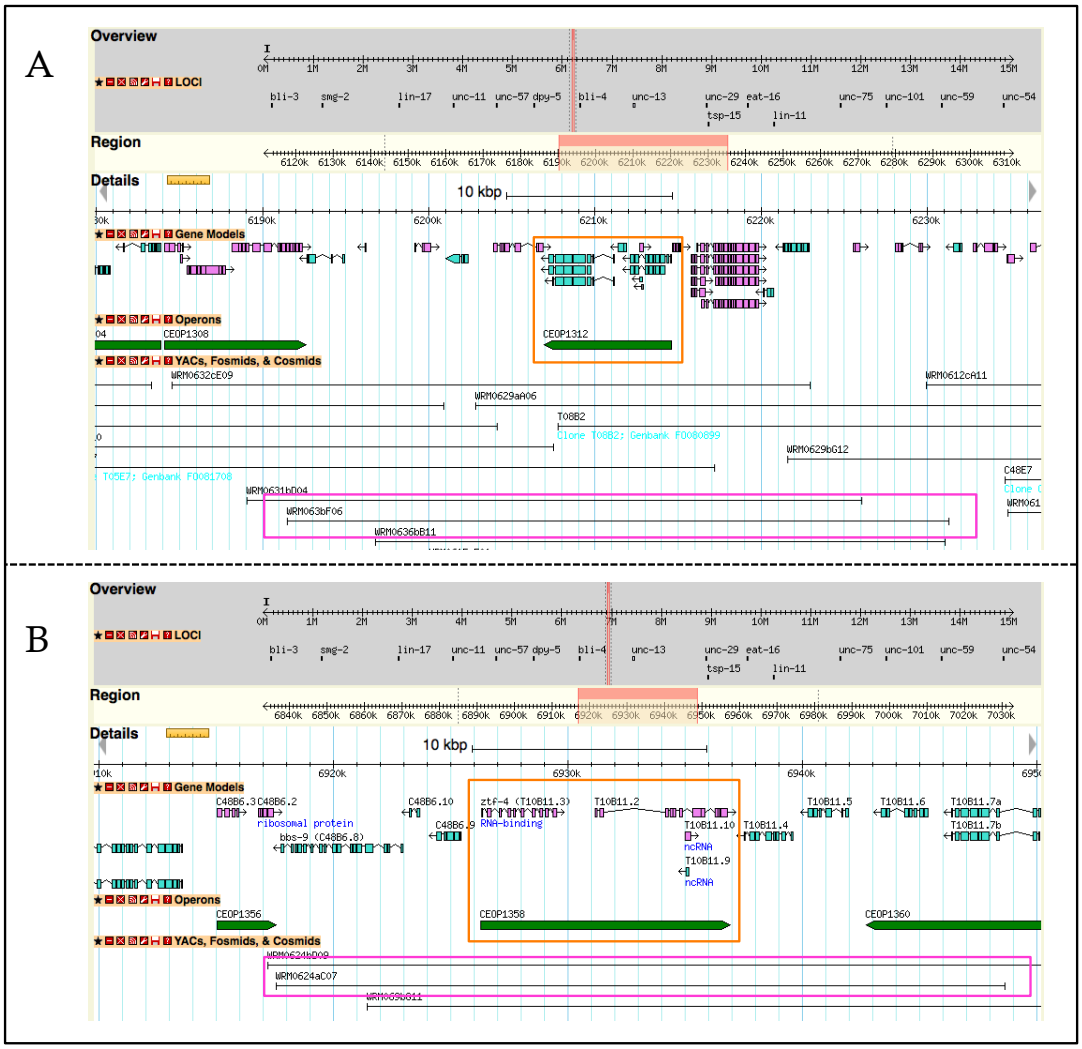

Figure 39. Wormbase snapshots illustrating the location of the two target operons on their respective fosmid clones. The fosmids WRM063bF06 (Panel A) and WRM024aC07 (Panel B) are highlighted by pink rectangles and the centrally-located, on their respective fosmid clones, operons CEOP1312 (Panel A) and CEOP1358 (Panel B) and enclosed in an orange box.

5.3.2 Selection of a gene model for each operon gene

5.3.2.1 CEOP1312

Operon CEOP1312 is a three-gene operon containing genes *T08B2.7*, *T08B2.8* and *T08B2.9*. *T08B2.7* has three gene models, *T08B2.8* has one and *T08B2.9* has two. As the FP CDS, excluding the stop codon, were to be introduced 6 codons upstream of the stop codon of the operon gene, the 3' ends of each gene model were

investigated. For *T08B2.7*, according to wormbase WS232, there are three gene models of which two, *T08B2.7a* and *T08B2.7b*, have the same 3' structure as the location of the stop codon in both models is the same. In addition, *T08B2.7a* uses a different transcription start site to *T08B2.7b*, supported by a single EST, and contains an additional small non-coding exon, the second exon in gene models *T08B2.7a* and *T08B2.7c*, (Figure 40) absent from *T08B2.7b*. Further examination of the single EST which supports the alternative transcription start site reveals no frame shifts resulting in stop codons, suggesting that it is not an artifact. Close examination of the third model, *T08B2.7c*, revealed that the final exon contained an apparent frame-shift and, as such, this model was excluded. Apart from a few exceptions all EST alignments and the ORFeome sequence tags (sequence reads from the ORFeome project) support gene model *T08B2.7b*, this model was presumed to represent the correct 3' end and chosen for reporter fusion generation.

For the terminal gene in operon CEOP1312, *T08B2.9*, the 3' ends of both gene models, *T08B2.9a* and *T08B2.9b*, aligned to each other i.e. were the same, and as the FP CDS was to be recombineered into the 3' end of the gene the EST and ORFeome data corresponding to each alternative model were closely examined. The ORFeome sequence tags and EST alignments support both gene models.

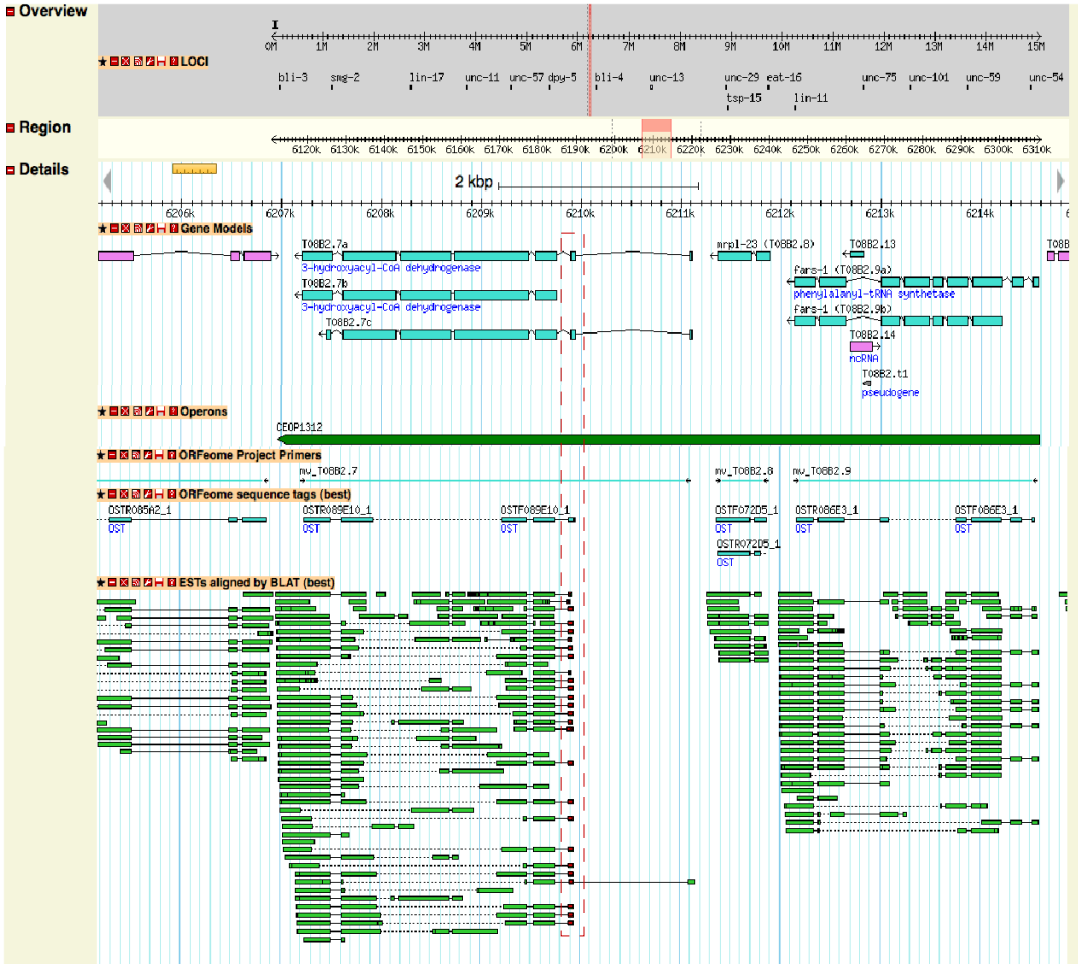

Figure 40. EST alignment and ORF sequences for CEOP1312. A snapshot of wormbase, WS232, illustrating operon CEOP1312. The gene models are the boxes and pointed arrows in blue under the genome and the dark green arrow depicts the operon structure. The thin blue boxes, under the gene models and labeled OST, are the ORFeome sequence tags and finally the green boxes under the OSTs are the ESTs which have been aligned to the gene model. The red dashed line encloses the non-coding exon identified in gene *T08B2.7*.

5.3.2.2 CEOP1358

Operon CEOP1358 contains two genes *T10B11.3* and *T10B11.2*. Each of the two genes within operon CEOP1358 are annotated within WormBase with a single gene model the structures of which are both confirmed by EST alignments and ORFeome sequence tags (Figure 41). As such, ODNs were designed to introduce reporter sequences into each gene.

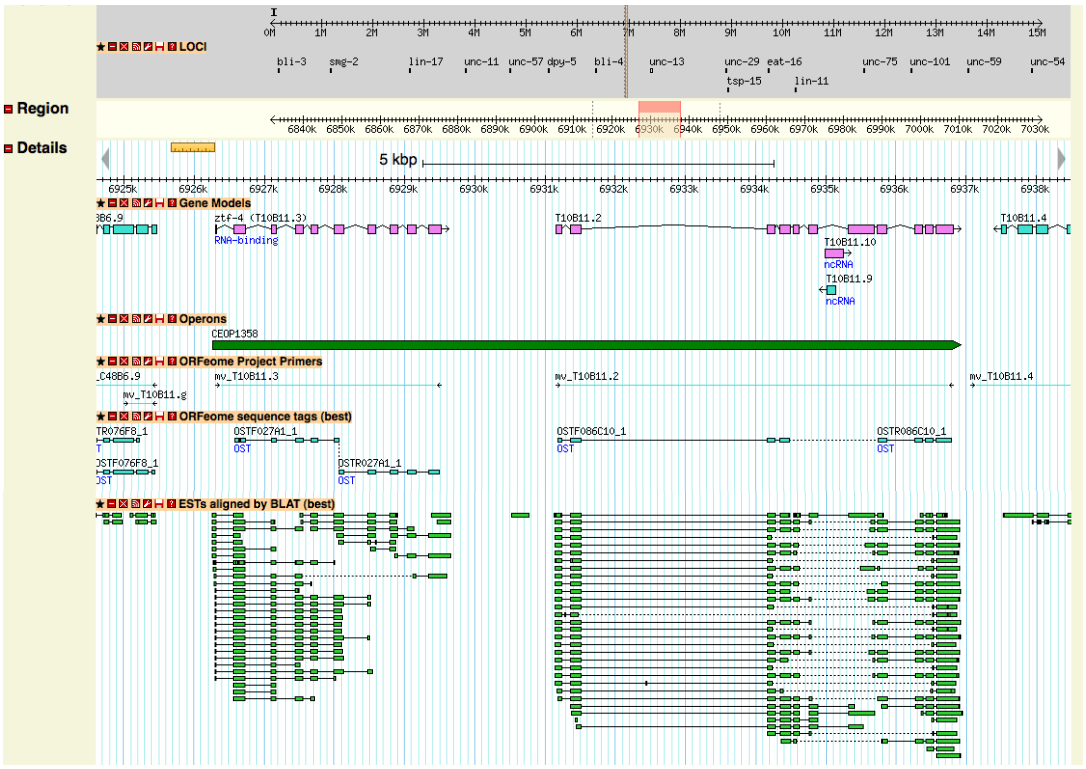

Figure 41. EST alignment and ORF sequences for CEOP1358. A snapshot of wormbase, WS232, illustrating operon CEOP1358. The gene models, *T10B11.3* and *T10B11.2*, are the boxes and pointed arrows in blue under the genome and the dark green arrow depicts the operon structure. The thin blue boxes, under the gene models and labeled OST, are the ORFeome sequence tags and finally the green boxes under the OSTs are the ESTs which have been aligned to the gene model.

5.3.3 Recombineering reporter constructs as multiple-tagged operon genes

The genes within operons CEOP1312 and CEOP1358 were iteratively tagged via counter-selection recombineering, utilizing the recombineering tools and resources described in Chapter 4, containing F-CFP, F-YFP and Mc-mCherry FP reporters (Figure 21). To indirectly examine whether the lack of fluorescent signal previously illustrated from the codon-optimized synthesized FPs (Chapter 4) was due to nonsense mediated mRNA decay, as a premature stop codon (sourced from the codon-optimized FP CDS) was included by error, mTFP1 and mCherry were also seamlessly recombineered into the gene coding sequences in CEOP1358 (Figure 21).

Pairs of ODNs were designed to introduce the FP-RT-FP Cassette 6 codons upstream from the stop codon of each gene (Tables 4, page 69, and 5, page 71) within operons CEOP1312 and CEOP1358.

FP-RT-FP Cassette inserts, for the respective FP reporter to be introduced, were generated by PCR (Table 5, page 71) and recombineered into the respective fosmid clone. The FP-RT-FP Cassette were subsequently replaced with reporter sequences excised from the appropriate constructs as *Not*I restriction fragments. The single- and multiple-tagged fosmid constructs (Figure 21) were microinjected into gravid wild type worms. A minimum of three independent transgenic lines were created for each construct and animals from each line imaged.

5.3.4 CEOP1312

The three genes within operon CEOP1312 were iteratively tagged with *F-CFP* (*T08B2.7b*), *F-YFP* (*T08B2.8*) and *Mc-mCherry* (*T08B2.9b*) (Figure 42).

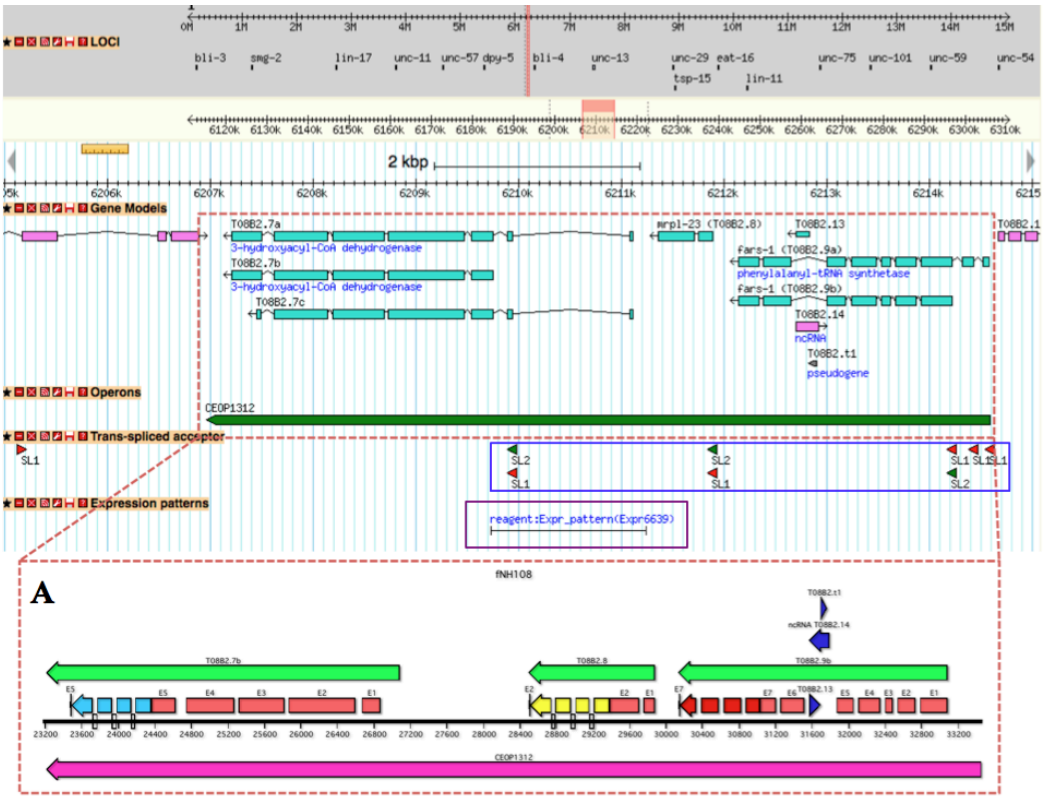

Figure 42. Snapshot of CEOP1312 from wormbase and schematic of final fosmid-based translational reporter. Above is a snapshot from wormbase WS232 of the three gene

operon CEOP1312 with the *trans*-splice acceptor sites (enclosed by a blue rectangle) and the region utilized in the transcriptional reporter (highlighted by the purple box) that was used to generate transgenic strain BC15507. Schematic illustrating the structure of the triple-tagged reporter fusion construct after all the three operon genes had been iteratively tagged with a FP CDS (Panel A).

To allow comparison of the expression pattern for *T08B2.7b::F-CFP* obtained from the fosmid-based translational reporter, strain BC15507, transgenic for a transcriptional reporter containing the promoter region of *T08B2.7b* driving expression of *gfp* (McKay et al., 2003; Hunt-Newbury et al., 2007), was imaged (Figure 43). Expression of GFP from the transcriptional reporter was visualized in body wall muscle (Figure 43, panels H and I), hypodermis, head (Figure 43, panels A – F) and tail neurons as previously reported (Hunt-Newbury et al., 2007). In contrast imaging of strains transgenic for the fosmid-based translational reporter *T08B2.7b::F-CFP*, revealed CFP expression in intestine and also as foci in body wall cells (Figure 44, panels A – D). Expression for *T08B2.8::F-YFP* was seen in the intestine and also faintly in muscle cells (Figure 44, panels E – H) and the expression pattern for *T08B2.9b::Mc-mCherry* was visualized as faint expression in the hypodermis (Figure 44, panels I – L). The results obtained illustrated that the expression of three genes visualized simultaneously in operon CEOP1312 showed expression in different tissue.

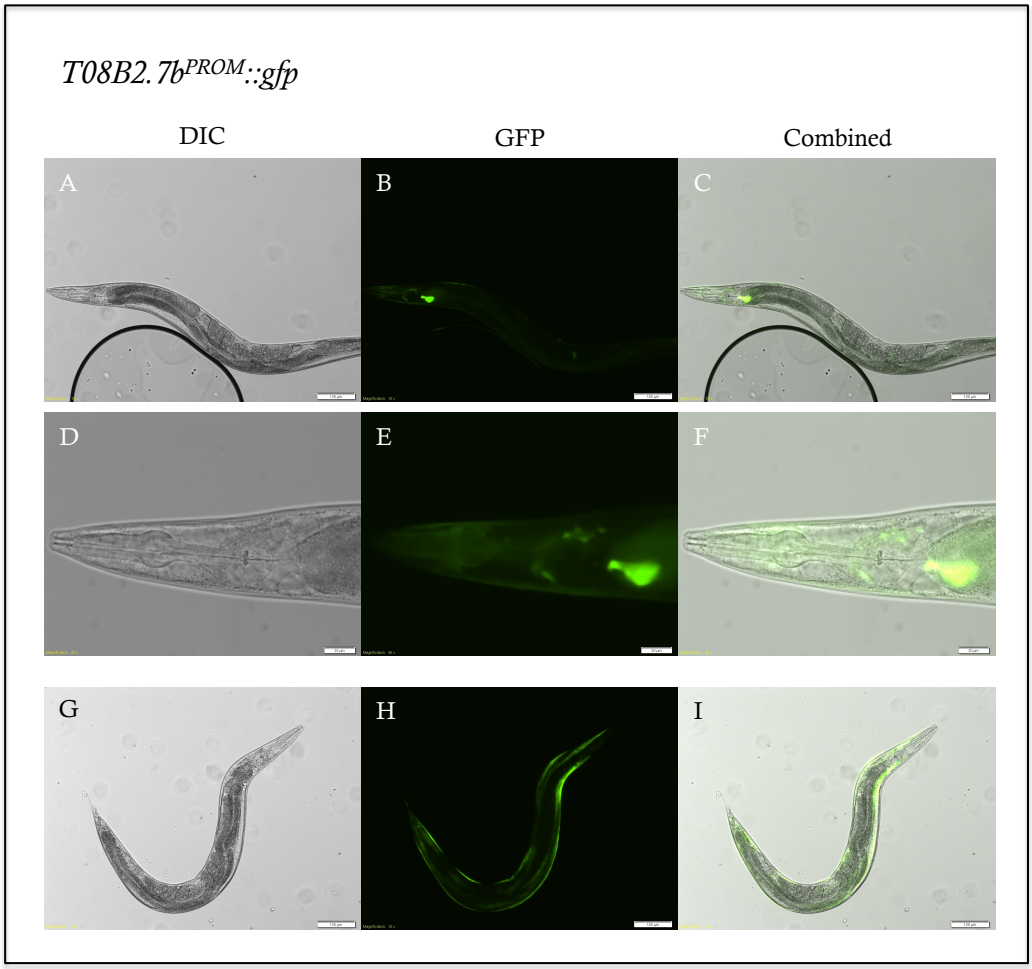

Figure 43. GFP expression pattern from a transcriptional reporter for T08B2.7b. Strain BC15507 (*T08B2.7b*^{PROM}*::GFP*, Hunt-Newbury et al., 2007), was imaged and GFP expression was seen in head (Panels A – F) and tail neurons an also in the body wall muscle and hypodermis (Panels G – I). Scale bar 100 μm (panels A – C and G – I) and 20 μm (panels D – F).

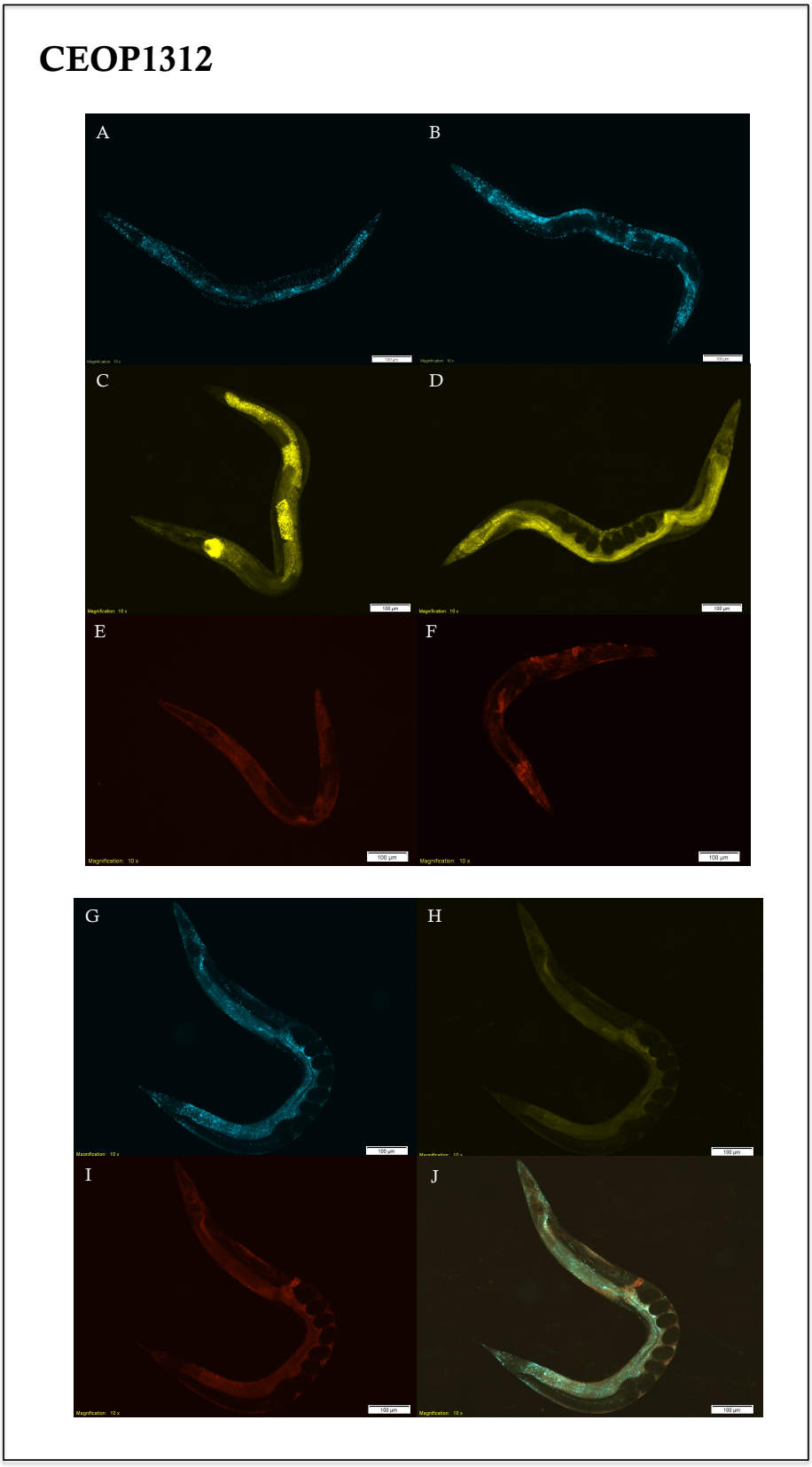

Figure 44. Translational reporter expression patterns of operon CEOP1312. Independently generated transgenic strains for either the single- (Panels A – F) or triple-tagged (Panels G – J) fosmid reporters (four in total) were imaged. Translational reporter expression patterns for *T08B2.7b::F-CFP* (Panels A - B) was observed in the intestine and also as foci in body wall cells and *T08B2.8* (F-YFP, Panels C - D) was visualized faintly in

muscle and also in the intestine of imaged animals. Expression of *T08B2.9b* (Mc-mCherry, Panels E - F) was seen in the hypodermis. Expression patterns, imaged from strains transgenic for the triple-tagged reporter, mimic expression visualized from the single-tagged reporters (Panels G - I). Scale bar represents 100 μm.

5.3.5 CEOP1358

The expression of the operonic genes, within CEOP1358, was interrogated via the use of recombineered fosmid-based translational reporter constructs. *T10B11.2* was tagged with *F-CFP*, and *T10B11.3* with *Mc-mCherry* CDSs (Figure 45, Panel A). The two genes were also tagged with the codon-optimized CDSs for contiguous *mTFP1* (*T10B11.2*) and contiguous *mCherry* (*T10B11.3*) (Figure 45, Panel B) to compare the brightness of the codon-optimized FP CDSs designed with the expression F-CFP and Mc-mCherry. The single and multiply tagged genes, six fosmid constructs in total (Figure 21) were microinjected and the independently generated lines imaged.

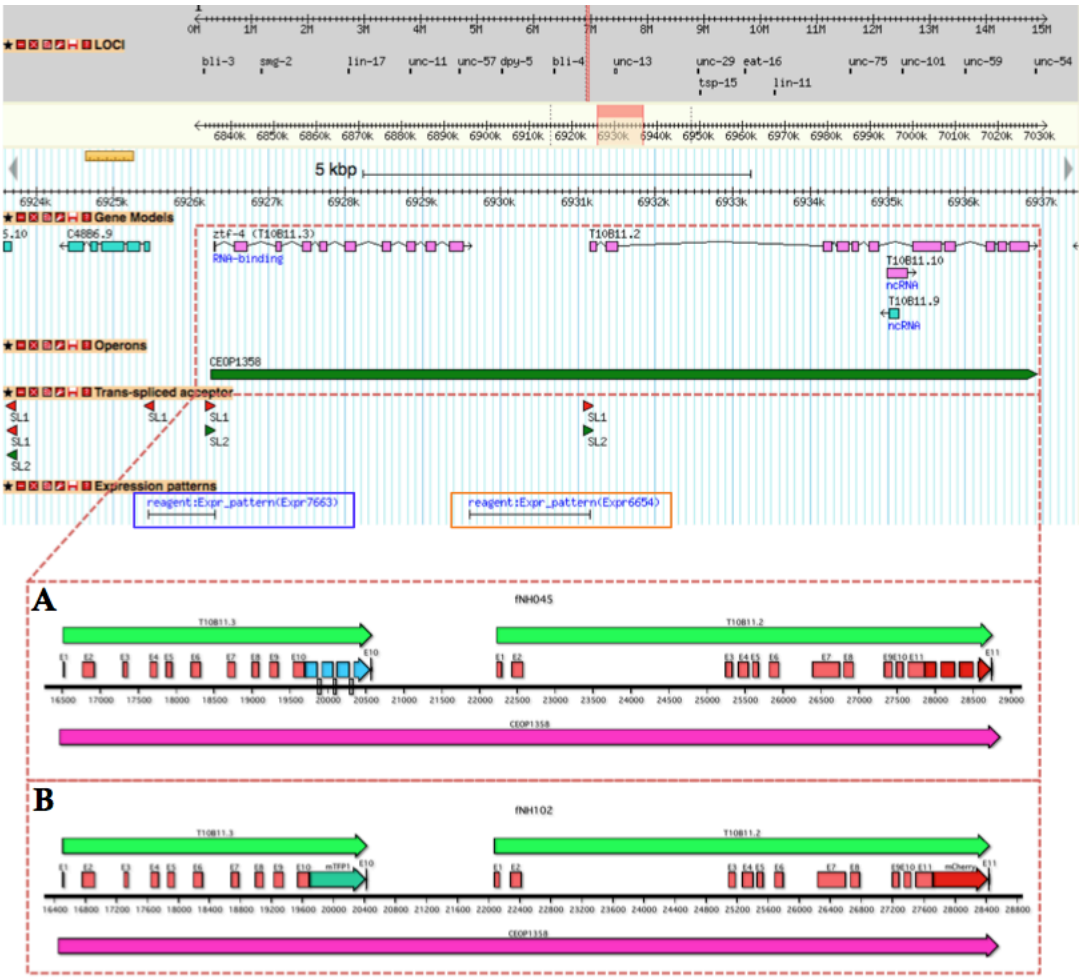

Figure 45. Wormbase snapshot of CEOP1358 and schematics of the final fosmid-based translational reporters. A snapshot of wormbase, WS232, of the two gene operon CEOP1358 with the *trans*-splice acceptor sites and also the promoter regions utilized in the two transcriptional reporters used to generate transgenic strains BC11336 (Highlighted by orange box, Hunt-Newbury et al., 2007) and UL1027 and UL2136 (enclosed by purple box, Reece-Hoyes et al., 2007). Panel A shows a schematic, illustrating the structures of the tagged operon genes after both genes had been iteratively tagged with F-CFP and Mc-mCherry whereas panel B shows the schematic of the final fosmid construct after introducing both codon-optimized mTFP1 and mCherry into the operon structure.

Both genes in this operon have previously published expression patterns generated from transcriptional style reporter constructs. Images of the expression patterns for T10B11.3 via two transgenic strains, UL1027 and UL2136 which are transgenic for *T10B11.3*^{PROM}*::GFP*, were available from 'The Hope Laboratory Expression Pattern Database' and were generated as part of a project to provide insight into transcription factor expression patterns (Reece-Hoyes et al., 2007). Transgenic strain UL1027 contains approximately. 4 kb promoter sequence for T10B11.3, in comparison to strain UL2136 which contains an approximately. 850 bp

promoterome fragment driving *gfp* expression (I. A. Hope, pers comm). GFP expression was visualized in muscle (Figure 46, panels A-B), the vulva (Figure 46, panels C-D) and pharyngeal nerves (Figure 46, panels G-H) (Reece-Hoyes et al., 2007). Transgenic strain BC11336 (*T10B11.2*^{PROM}*::GFP*; Hunt-Newbury et al., 2007) was obtained from CGC (Caenorhabditis Genetics Center) and imaged in house. GFP expression was observed in the spermatheca, faintly in the intestine and also seen in some head (Figure 47, Panels H-I) and tail neurons (Figure 47, Panels B - C).

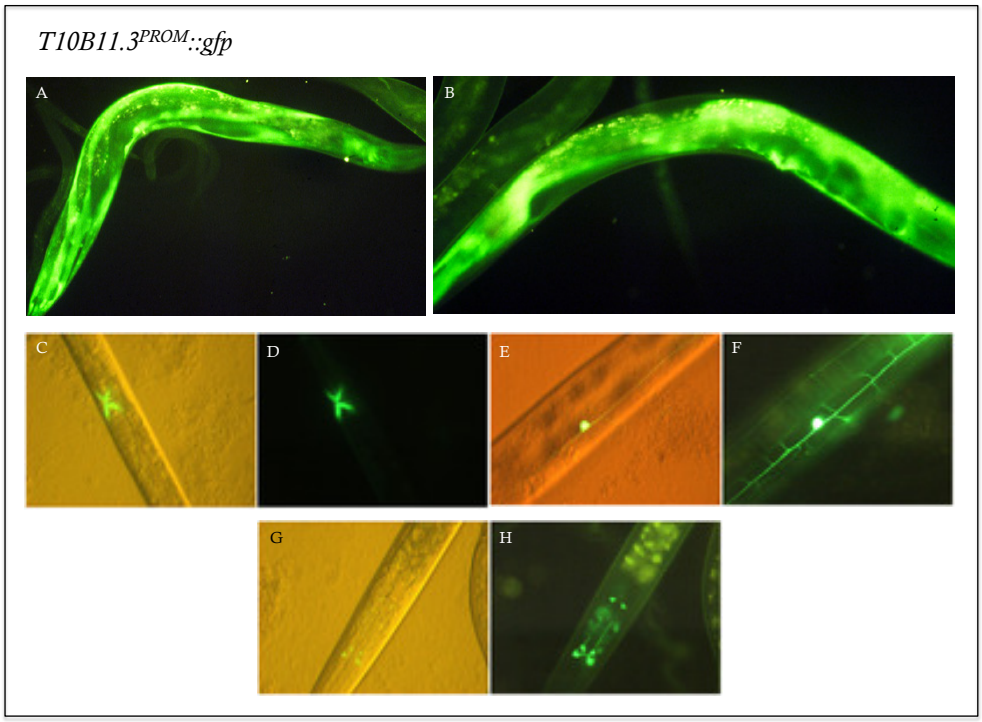

Figure 46. GFP expression patterns from strains UL1027 (panels A and B) and UL2136 (panels C – H). GFP expression for animals transgenic for GFP driven by *T10B11.3*^{PROM} demonstrated expression in bodywall muscle cells (Panel A and B). Expression can also be seen in the vulva (Panels C and D), a single nerve cell (Panels E and F) and pharyngeal nerves (Panels G and H). Images from (Reece-Hoyes et al., 2007).

Expression for *T10B11.3* was visualized via the F-CFP in the intestine, in head and tail neurons and also in somatic gonad, the non-germline component of the gonad arm (Figure 48). Expression of *T10B11.2* was detected via the Mc-mCherry in the spermatheca, throughout the nervous system and also faint expression was observed in muscle cells (Figure 48). Expression of both genes in operons CEOP1358 shows expression in different tissues.

Finally fosmid-based reporters containing *T10B11.3::mTFP1* and *T10B11.2::mCherry* were also constructed in order to be able to compare the brightness between FPs encoded by established CDSs and those expressed from the codon-optimized CDSs designed as part of this project. Fosmid constructs containing the operonic genes tagged with codon-optimized mTFP1 and mCherry CDSs were microinjected, both single- and dual-tagged genes, and independently generated lines were imaged. Expression for T10B11.3, which was tagged with mTFP1, was imaged in a few cases in the intestine and expression of the codon-optimized synthesized mCherry, fused to T10B11.2, was extremely faint as it could only be visualized in the muscle (Figure 49). The visualization of the two genes in CEOP1358 was difficult when using the synthetic codon-optimized mTFP1 and mCherry FPs suggesting that the codon-optimized FPs are not suitable when visualizing expression patterns via fosmid-based translational reporters.

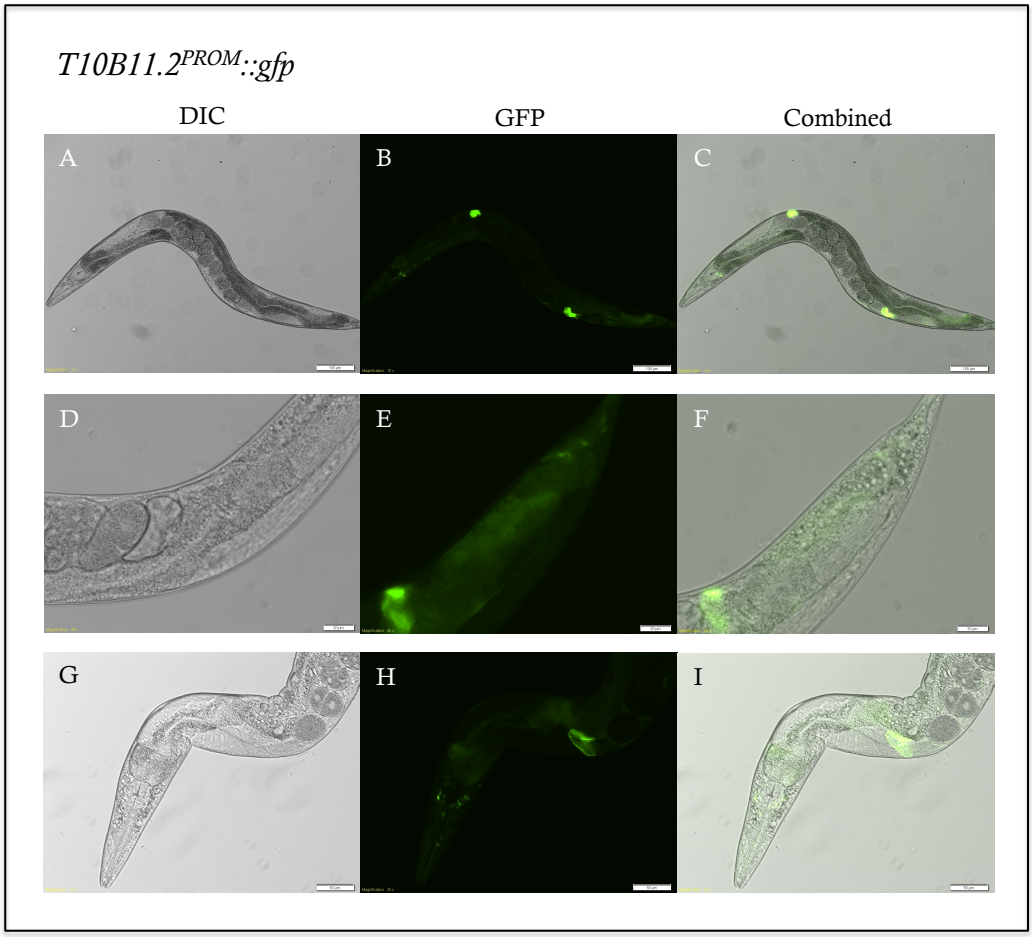

Figure 47. GFP expression pattern from transcriptional reporter for T10B11.2. Transgenic strain BC11336 (*T10B11.2*^{PROM}*::GFP*, Hunt-Newbury et al., 2007), obtained from CGC, was imaged and expression was visualized in some head and tail neurons (Panels D – F), the spermatheca (Panels A – C and G – I) and faintly in the intestine (Panels E – F). Scale bars in panels A – C indicates 100 μm, panels D – F 20 μm and in panels G – I 50 μm.

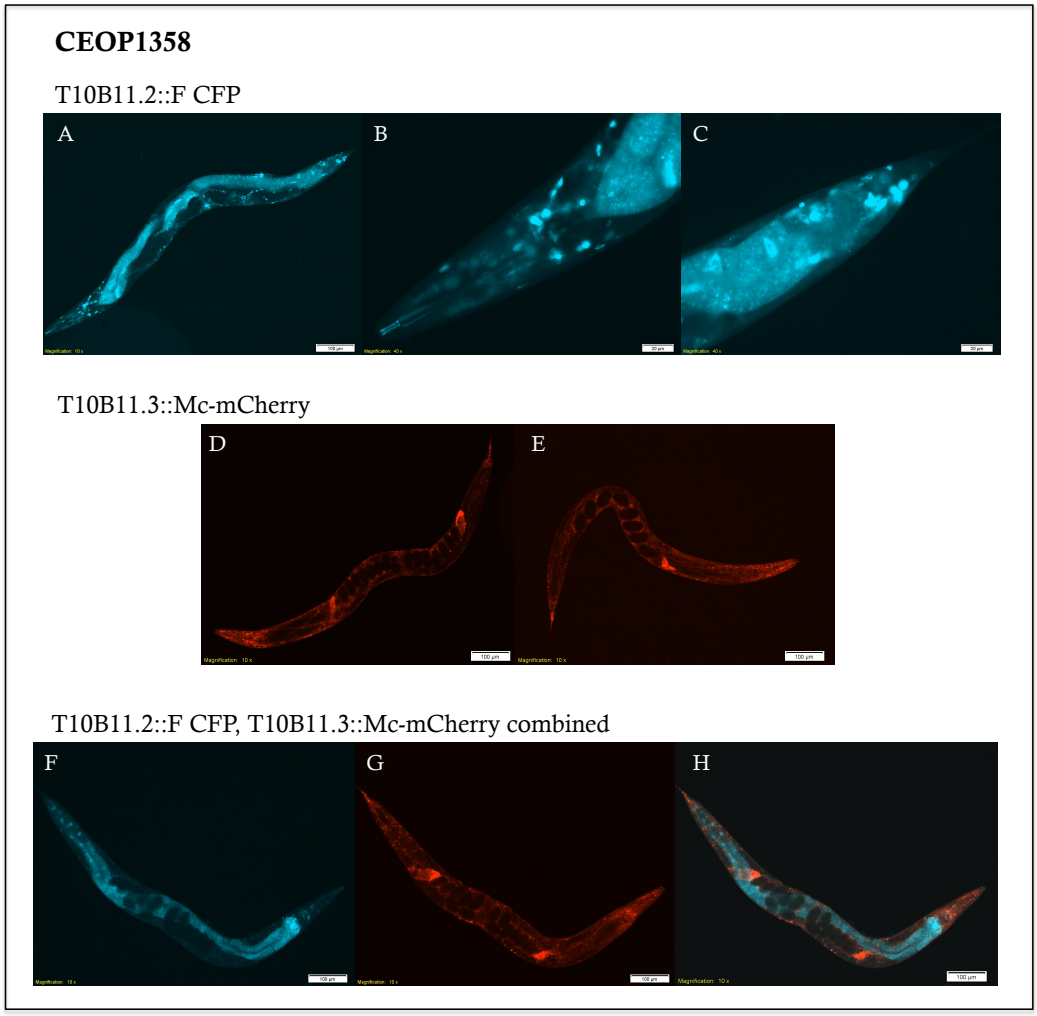

Figure 48. F-CFP and Mc-mCherry expression patterns for genes within operon CEOP1358. Fosmid constructs with single and multiply tagged genes, with F-CFP and Mc-mCherry, for operon CEOP1358 were microinjected into N2 adults and the resulting independent lines were generated. Expression for *T10B11.3* via a fosmid-based reporter was observed in head and tail neurons and also in the intestine. Fosmid reporter based expression for *T10B11.2* was visualized in the spermatheca, through out the nervous system and also faintly in muscle cells. Scale bar in panels B – C indicates 20 μm and in panels A, D – H indicates 100 μm.

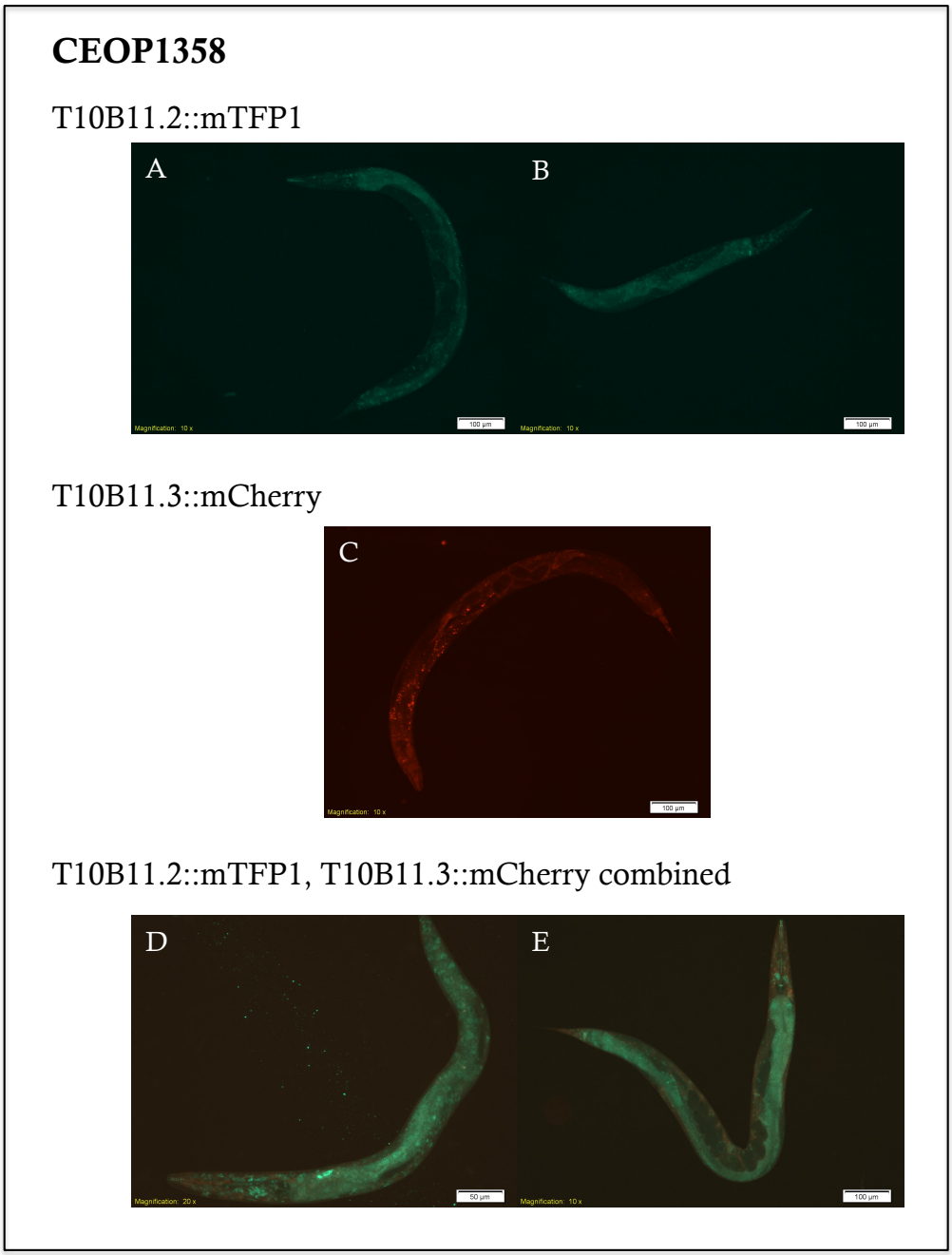

Figure 49. mTFP1 and mCherry expression patterns for genes within operon CEOP1358. Independently generated transgenic worms for fosmid constructs containing single and multiply tagged genes, with mTFP1 and mCherry, were imaged. Fosmid-based expression of *T10B11.3*, tagged with mTFP1, was only visualized in the intestine and the expression of *T10B11.2*, fused to mCherry, was only faintly visualized in the muscle. Scale bar represents 100 μm.

5.4 Discussion

A subset of constructs from the collection of recombineering resources utilizing the FP-RT-FP Cassette, generation detailed in Chapter 4, were utilized to seamlessly tag the genes contained within two different operons via iterative rounds of counter-selection recombineering. The first operon, CEOP1312, is a three-gene operon containing genes *T08B2.7*, *T08B2.8* and *T08B2.9* and is centrally located on fosmid clone WRM063bF06 from which the translational reporters were constructed. *T08B2.7*, *T08B2.8* and *T08B2.9* encode, respectively, hydroxyacyl-coenzyme A dehydrogenase orthologue, a large mitochondrial ribosome protein and a predicted phenylalanyl-t-RNA synthetase. The iterative rounds of recombineering-generated single, double and triple-tagged fosmid constructs for F-CFP/F-YFP/Mc-mCherry FP combinations with the FP CDS (excluding the FP stop codon) introduced 6 codons upstream of the stop codon of the tagged gene. Transgenic lines were generated for the single- and triple-tagged constructs and expression patterns were investigated (Figure 44). The second operon, CEOP1358, is a 2-gene operon, containing *T10B11.3* and *T10B11.2*, and is centrally located on fosmid clone WRM0624aC07. Fosmid-based translational reporters were built via iterative rounds of recombineering for F-CFP/Mc-mCherry and mTFP1/mCherry combinations with the FP CDSs introduced (excluding the stop codon of the FP) 6 codons upstream of the stop codon. Codon-optimized mTFP1 and mCherry CDSs were introduced into the two genes within operon CEOP1358 to investigate whether the presence of the premature stop codon in the previous translational reporter constructs (Chapter 4) had previously affected the expression level of the corresponding fusion protein, perhaps via a mechanism such as nonsense mediated mRNA decay (Mango, 2001) as a possible explanation for the lack of fluorescent signal. Transgenic lines were generated for all the fosmid-based translational reporters and resulting expression patterns examined.

Expression patterns of *T08B2.7*, *T08B2.8* and *T08B2.9*, from the fosmid-based translational reporters, were determined by examination of worms transgenic for the single- and triple-tagged F-CFP/F-YFP/Mc-mCherry constructs. *T08B2.7* encodes a gene orthologous to the human gene hydroxyacyl-coenzyme A dehydrogenase/ 3-ketoacyl-coenzyme A thiolase/enoyl-coenzyme A hydratase, a trifunctional protein. Mutation of the protein's alpha subunit leads to disease – familial hyperinsulinemic hypoglycemia (low blood glucose caused by excessive

insulin) which can be reversed by raising the blood glucose level (Ibdah et al., 1999; Ibdah et al., 2001). Hydroxyacyl-Coenzyme A dehydrogenase is an enzyme and a member of the 3-hydroxyacyl-CoA dehydrogenase gene family. The encoded protein functions in the mitochondrial matrix to catalyze an oxidation of a straight chain fatty acid as part of the beta-oxidation pathway to generate acetyl-CoA (for the citric acid cycle) and due to its involvement in metabolism one may expect widespread expression. Investigation of strains transgenic with the single- and triple- tagged *T08B2.7::F-CFP* revealed expression of F-CFP throughout the intestine and also as foci in body wall cells. Expression pattern resulting from strain BC15507 (Figure 43, McKay et al., 2003; Hunt-Newbury et al., 2007) revealed a similar expression pattern as wide spread expression of GFP was visualized in body wall muscle, albeit not as punctate spots, hypodermis and also in head neurons.

T08B2.8 (mrpl-23) encodes a large mitochondrial ribosome protein that is orthologous to human MRPL23 (Shaye and Greenwald, 2011). The mitochondrial ribosomal proteins are encoded by nuclear genes and aid protein synthesis within the mitochondrion (Tsang et al., 1995). Examination of strains transgenic with the single- and triple-tagged *T08B2.8::F-YFP* revealed expression in the intestine and also faint fluorescence expression in muscle cells. As the protein is mitochondrial, one might expect to see punctate expression consistent with expression in these organelles. However, such a pattern was not seen and this may have been due to the resolution of the imaging system being unable to capture such detail and may have been visualized if the strains were imaged on a confocal system.

T08B2.9 (fars-1) is predicted to encode a phenylalanyl-t-RNA synthetase, homologous to the human phenylalanyl-t-RNA synthetase, and participates in phenylalanine, tyrosine and tryptophan biosynthesis and aminoacyl-tRNA biosynthesis. The phenylalanyl t-RNA synthetase belongs to a family of ligases which specifically catalyze the formation of carbon-oxygen bonds in aminoacyl-tRNA and other related molecules. Imaging of the strains transgenic for *T08B2.9::Mc-mCherry* revealed faint expression in the hypodermis (Figure 44).

Expression patterns of the genes *T10B11.3* and *T10B11.2*, contained within the second operon CEOP1358, were determined by examination of worms transgenic for single- and double-tagged F-CFP/Mc-mCherry and mTFP1/mCherry fosmid constructs. *T10B11.3 (ztf-4)* encodes a transcription factor containing a zinc finger binding domain which initiates their modulatory effect in the region of any

sequence where the domain binds (Clarke and Berg, 1998). As *T10B11.3* encodes for a transcription factor, very localized expression, in time and space, might be expected. Expression patterns of T10B11.3::F-CFP from strains transgenic for single- and double- tagged *T10B11.3::F-CFP* (Figure 48) revealed expression in head and tail neurons, intestinal expression and also expression in the somatic gonad. However, expression of GFP when imaging strains transgenic for *T10B11.3^{PROM}::GFP* (UL1027 and UL2136) revealed a different expression pattern, compared to what was seen for F-CFP from the translational reporters, as GFP was observed in muscle, vulva and pharyngeal nerves (Figure 46).

T10B11.2, the second gene in this two-gene operon, was first identified as a sphingolipid kinase from a genome-wide microarray where the T10B11.2 mRNA was detected in oocytes (Kubagawa et al., 2006). T10B11.2 was later described as a ceramide kinase that participates in sphingolipid metabolism and regulates ceramide biosynthesis (Sugiura et al., 2002). The ceramide synthesis pathway is required for induced apoptosis in the *C. elegans* germline and also suggested to play a role in a parallel pathway that occurs in the mitochondrial step of the apoptosis process (Deng et al., 2008). Sphingolipids are a class of lipids that play important roles in signal transmission and cell recognition. Polyunsaturated fatty acids (PUFAs) are precursors of signaling molecules which function in oocytes to control sperm motility (Kubagawa et al., 2006). Imaging of strains transgenic for single- and double-tagged *T10B11.2::Mc-mCherry* revealed high expression in the spermathecae with additional expression throughout the nervous system and also faint expression in muscle cells.

Strains transgenic for single- and double-tagged *T10B11.3::mTFP1* and *T10B11.2::mCherry* were generated to investigate for two reasons, firstly to investigate if nonsense mediated mRNA decay could be a potential explanation for the lack of fluorescence signal previously visualized from the codon-optimized synthesized FPs (Chapter 4). Secondly to allow comparison of the mTFP1 and mCherry expression patterns against the F-CFP and Mc-mCherry patterns. Expression patterns from strains transgenic for fosmid-based translational reporters, tagged with the codon-optimized mTFP1 and mCherry CDSs, could not be interpreted due to the low fluorescence levels. From the lack of interpretable expression patterns it can be deduced that codon-optimization of the FP CDSs, in this instance, does not result in FPs with optimal fluorescence properties. As the translational reporters created for T10B11.3::mTFP1 and T10B11.2::mCherry do

not contain a premature stop codon, this would exclude the possibility of the low fluorescent signal resulting from nonsense mediated mRNA decay. This also corresponds to previous publications that state genes lacking 50 bp or more due to the presence of a premature stop codon undergo nonsense mediated mRNA decay (Mango, 2001). In conclusion, as both sets of fosmid constructs only differ by the FP CDSs they contain, the codon-optimized FP CDSs also cannot be utilized in fosmid-based reporters to provide expression pattern data for any *goi* tagged with codon-optimized FP CDS.

6

Dissection of potential internal regulatory elements within *C. elegans* operons

6.1 Introduction

Operon gene organization in the *C. elegans* genome is more complex than that in bacterial genome where a major upstream promoter controls the expression of the genes within the operon. In 2007, Huang et al established that 25 % of *C. elegans* operons contain internal promoter elements in the intergenic sequences of downstream genes in the operon and these operons were termed "hybrid-operons". Expression pattern data generated with transcriptional reporters that utilized either the "major upstream operon promoter" or the "internal promoter" driving *gfp* expression exhibited different expression patterns as in many cases expression was visualized in different tissues (Huang et al., 2007). To dissect out the function of an "internal promoter", it must be investigated in the context of the whole operon. Recombineering would allow precise and seamless modifications to be made to potential internal regulatory elements, on a fosmid-based translational reporter, and their regulatory role, within the operon, to be dissected out.

6.2 Methods

See section 2.7 (identification and subsequent modification of potential internal regulatory elements).

6.3 Results

The two operon structures were investigated for conservation in the *Caenorhabditis* genus as 96% of the operons in *C. elegans* are conserved in *C. briggsae* (Stein et al., 2003). Sequences containing operons CEOP1312 and CEOP1358 were visualized in Ensembl and aligned with their equivalent regions in *C. briggsae*, *C. brenneri* and *C. remanei*. In addition to performing these global alignments across the available *Caenorhabditis* genus, the sequences encompassing the three- and two-gene operons from *C. elegans* were exported for subsequent detailed analysis in MacVector (v 9.5).

C. elegans operon CEOP1358 is not conserved in the *Caenorhabditis* genus as when *T10B11.3* and *T10B11.2* were aligned to their respective homologs in *C. briggsae*, *C.*

brenneri and *C. remanei* it was found that the operon structure in *C. elegans* is not conserved as the genes are not syntenic i.e. do no possess common chromosome sequences (Figure 50). Equivalents of genes *T10B11.3* and *T10B11.2* are present in the *C. briggsae* genome and are located on the same chromosome. However, the *T10B11.3* equivalent, *CBG12729*, is in the reverse orientation and is separated from the *T10B11.2* equivalent, *CBG12724*, by four genes. This is suggestive that the operon in *C. elegans* formed after the *C. elegans* and *C. briggsae* lineages diverged.

In contrast CEOP1312 is a conserved operon in the *Caenorhabditis* genus as it is syntenic (the genes homologous to those in CEOP1312 are in the same relative positions in the additional genomes investigated) between the *Caenorhabditis* species investigated here (*C. elegans*, *C. briggsae*, *C. brenneri* and *C. remanei*) (Figure 51). Pustell DNA sequence matrix alignment (Pustell and Kafatos, 1984) of the operon was performed between *C. elegans*, *C. briggsae*, *C. brenneri* and *C. remanei* using MacVector (Figure 52). From these alignments of the operon sequences from the four species, a conserved non-coding sequence (NCS) was identified in the intergenic region (Figure 52), between *T08B2.8* and *T08B2.7* in CEOP1312. The non-coding conserved sequence was determined by manual alignment and were illustrated with plotsimilarity and the overall consensus sequence emphasized with the tool 'weblogo' (Schneider and Stephens, 1990).

ModEncode data was also scanned to identify any potential transcription factor binding sites in the region of the non-coding conserved sequence identified in CEOP1312. A subset of the ModEncode data, generated by the Snyder group illustrated in Wormbase (Figure 53) depicts the locations of the transcription factor binding sites that were identified by GFP-ChiP. From this data, the NCS appears to be within the binding site of some transcription factors.

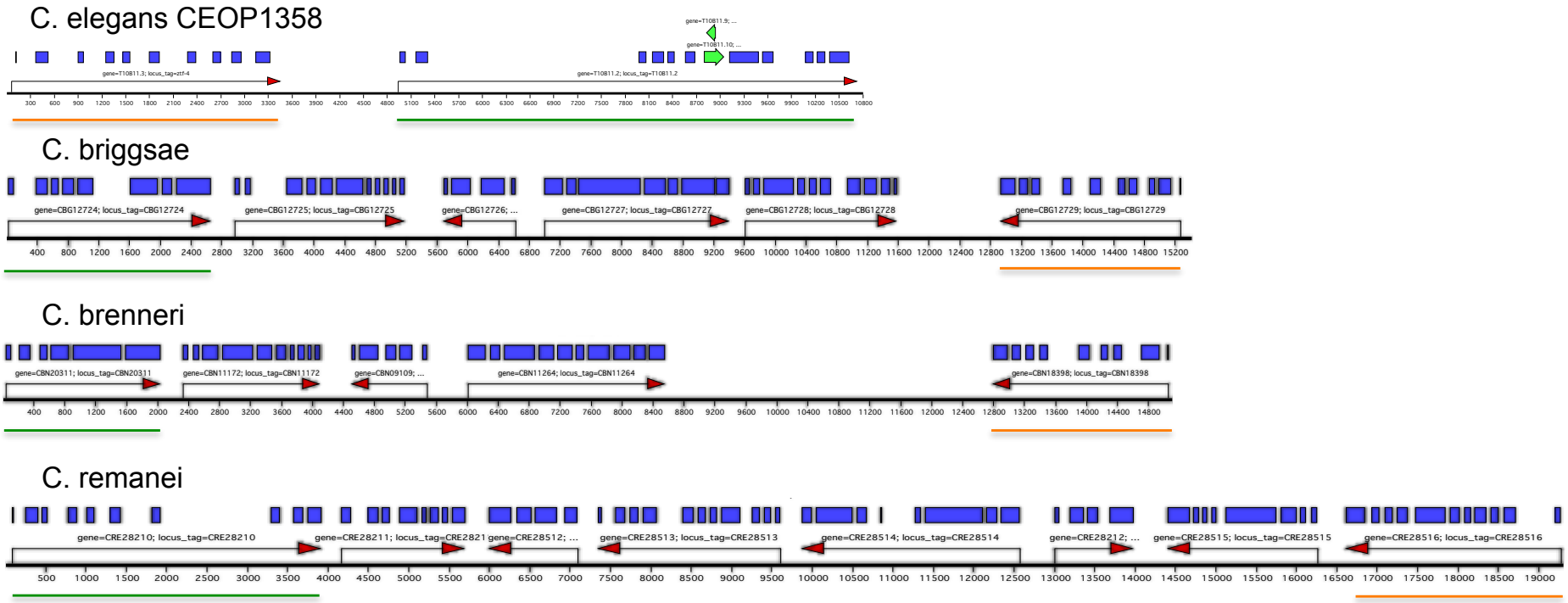

Figure 50. Alignment of genes in *C. elegans* operon CEOP1358 in *C. briggsae*, *C. brenneri* and *C. remanei*. Genes that are homologous to *T110B11.3*, genes underlined in orange, and *T110B11.2*, genes underlined in green, were identified in *C. briggsae*, *C. brenneri* and *C. remanei*. As illustrated above the operon structure is not conserved as the genes are not in the same order and also not on the same strand of DNA (the *T10B11.3* equivalent is on the opposite strand).

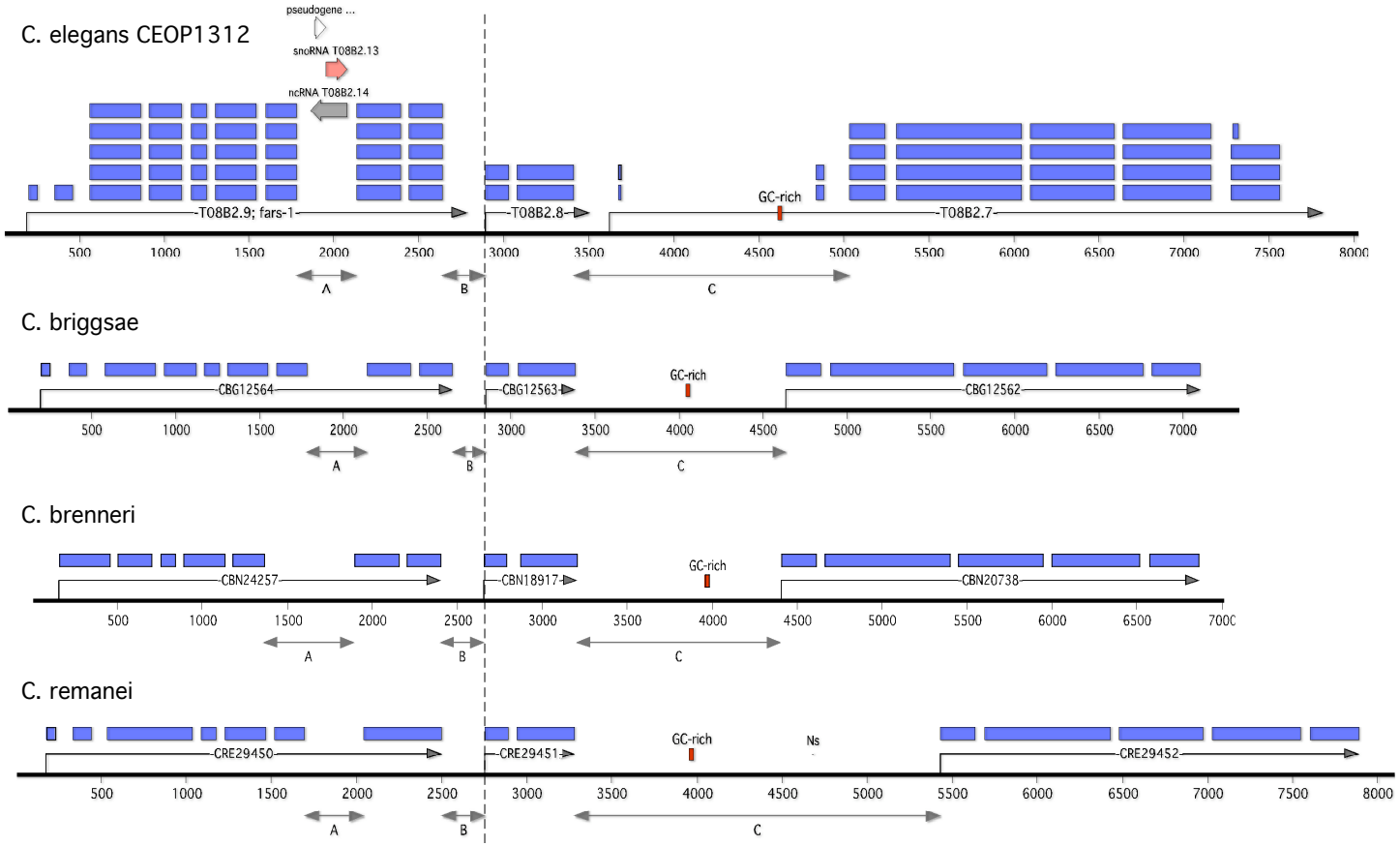

Figure 51. Alignment of CEOP1312 in *C. elegans*, *C. briggsae*, *C. brenneri* and *C. remanei*. Genes homologous to *T08B2.7*, *T08B2.8* and *T08B2.9* were identified in *C. briggsae*, *C. brenneri* and *C. remanei* and the genomic sequence of the genes in each species aligned to one another. As illustrated, the operon structure is conserved in the four species with the intergenic region lengths (doubled headed arrows A, B and C) and intron lengths also conserved. The vertical dashed line is the point from which the operon structure was aligned from and the genomic regions are to scale. The short red rectangles labeled 'GC Rich' above each genomic region indicates the location of the conserved NCS which was identified in each species.

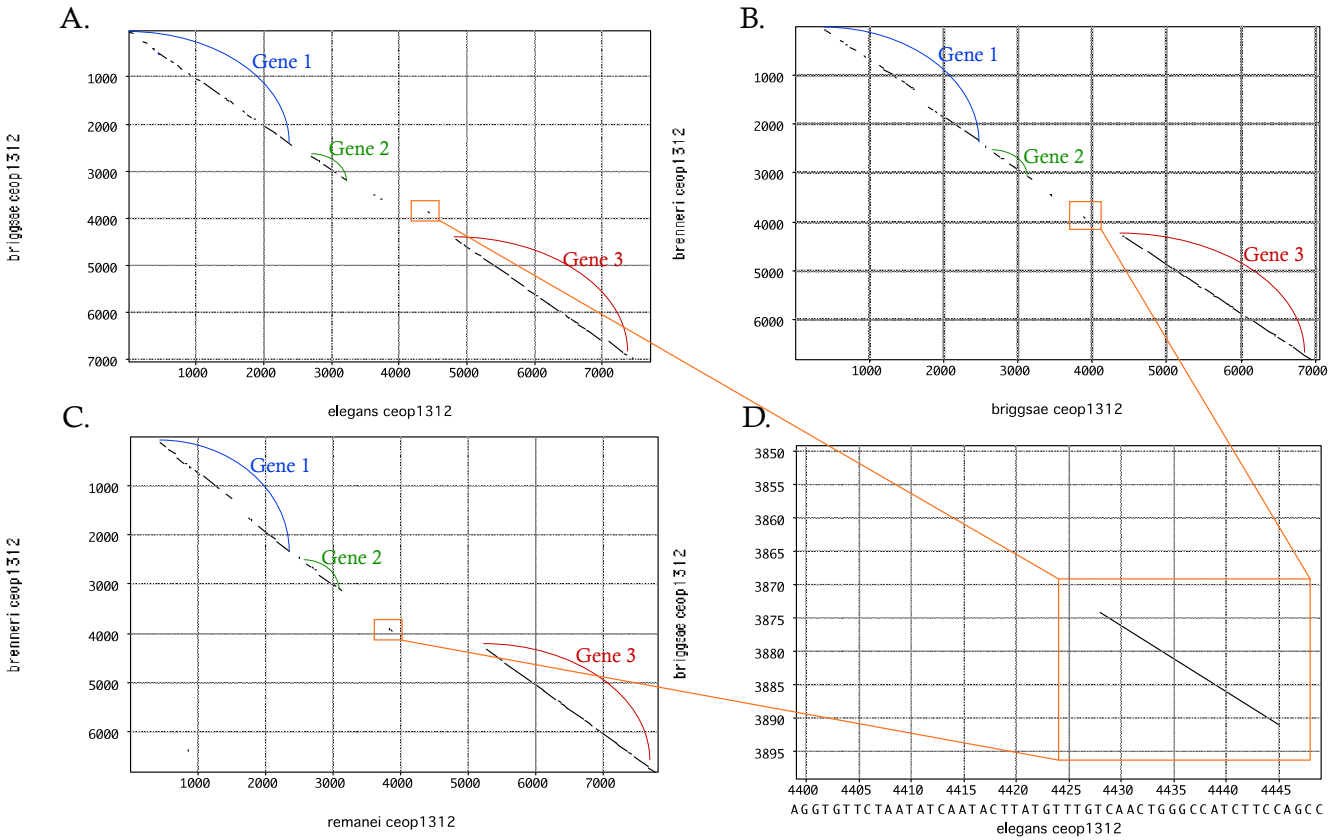

Figure 52. Pustell DNA matrix of CEOP1312 between *C. elegans*, *C. briggsae*, *C. brenneri* and *C. remanei*. A collection of three pair-wise alignments (pustell DNA matrices) comparing the sequence of operon structure CEOP1312 in *C. elegans* and *C. briggsae* (Panel A), *C. briggsae* and *C. brenneri* (Panel B), *C. brenneri* and *C. remanei* (Panel C) and a close up of the NCS (region enclosed by an orange box in each panel) identified (Panel D). The three genes that make up the operon structure are conserved as highlighted in blue for *T08B2.9* and its equivalents, in green for *T08B2.8* and equivalents and *T08B2.7* and equivalents in red. Between gene 2 and 3 in the matrices shown a short sequence in a non-coding region also appears to be conserved and is highlighted in orange. When this region is magnified (Panel D) a short stretch of sequence (approximately. 25 bp) in the non-coding region also appears to be conserved.

A Pustell DNA matrix, more simply known as a 'dot blot', was performed between two sequences. One sequence is represented on the x-axis and the other on the y-axis and if the x coordinate matches the y coordinate it is marked with a dot. This tool is useful for identifying regions of similarity between two sequences and will also illustrate insertions, deletions and repeats (Pustell and Kafatos, 1984).

Pair-wise alignments between *C. elegans* and *C. briggsae* (Figure 52A), *C. briggsae* and *C. brenneri* (Figure 52B) and *C. brenneri* and *C. remanei* (Figure 52C) were performed. The genes and their order in the operon was found to be conserved and a small stretch of non-coding sequence was also identified as being conserved (Figure 52D). The NCS also overlaps sites identified, via the GFP-ChIP ModEncode study, to be potential binding sites for the transcription factors (TFs) tested (Figure 53). The TFs identified in the NCS are ALR-1, BLMP-1, GEI-11, AMA-1, HLH-8, DAF-16, EGL-5, MAB-5 and ELT-3. The NCS identified (Figure 52) was then assessed to determine the actual length of the sequence. The non-coding conserved sequence identified was aligned by eye (Figure 54) and illustrated by Logo (Figure 55) and plotsimilarity (Figure 56).

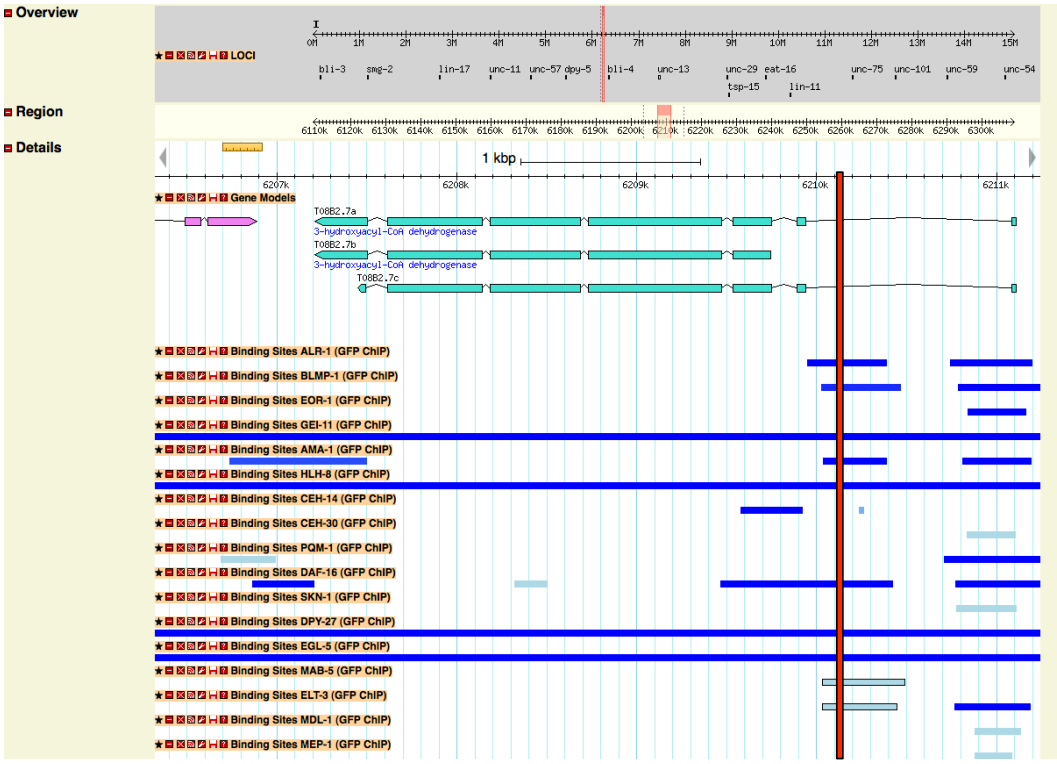

Figure 53. Wormbase snapshot of ModEncode data (GFP ChiP) identifying TF binding sites. ModEncode data (Gerstein et al., 2010) from the Snyder group has identified the potential location of TF binding sites in CEOP1312 and the blue blocks, under the respective

TF sub-heading, illustrate the binding sites. The bold vertical line represents the location of the NCS identified via bioinformatic analysis.

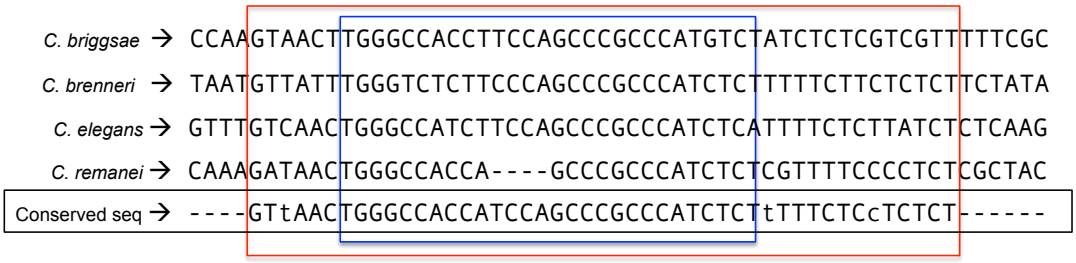

Figure 54. Alignment of conserved NCS identified in CEOP1312. The sequences of each non-coding conserved region, identified in Figure 52, were aligned by eye. A highly conserved GC rich region is enclosed by the blue box whilst the red box highlights the overall conserved sequence. The consensus-conserved sequence is below the four sequences and is encased by the black rectangle and the capital letters represent sequence with 100% homology and the lower case 50-75% homologous sequence.

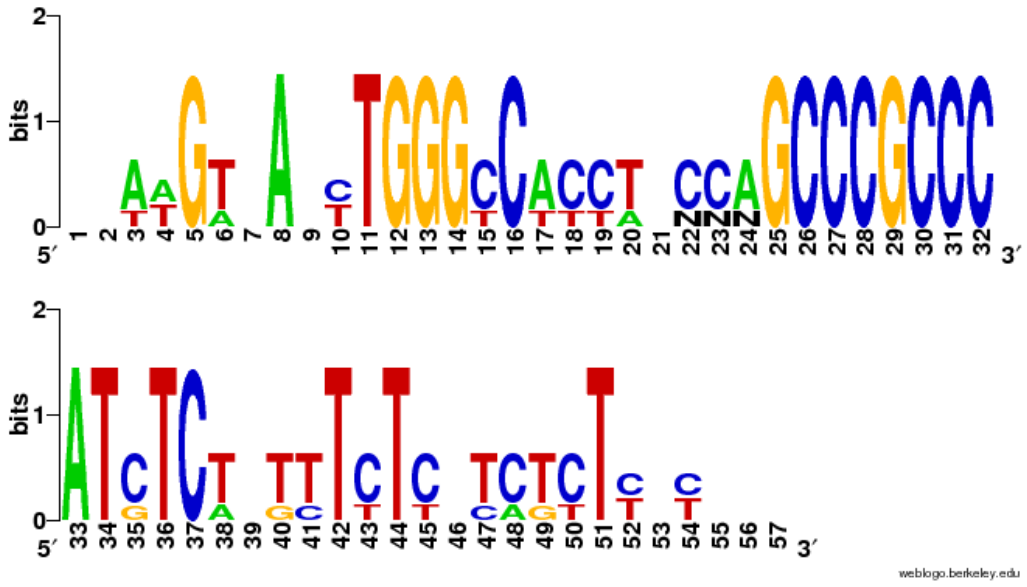

Figure 55. The multiple alignment of the conserved NCS in CEOP1312 represented as a sequence logo. A multiple sequence file (msf) of the NCS identified in each *Caenorhabditis* spp investigated, generated by P. Cunningham, was uploaded to Weblogo (Schneider and Stephens, 1990) and the pattern within the msf is pictorially illustrated. The Y-axis represents the position in the sequence.

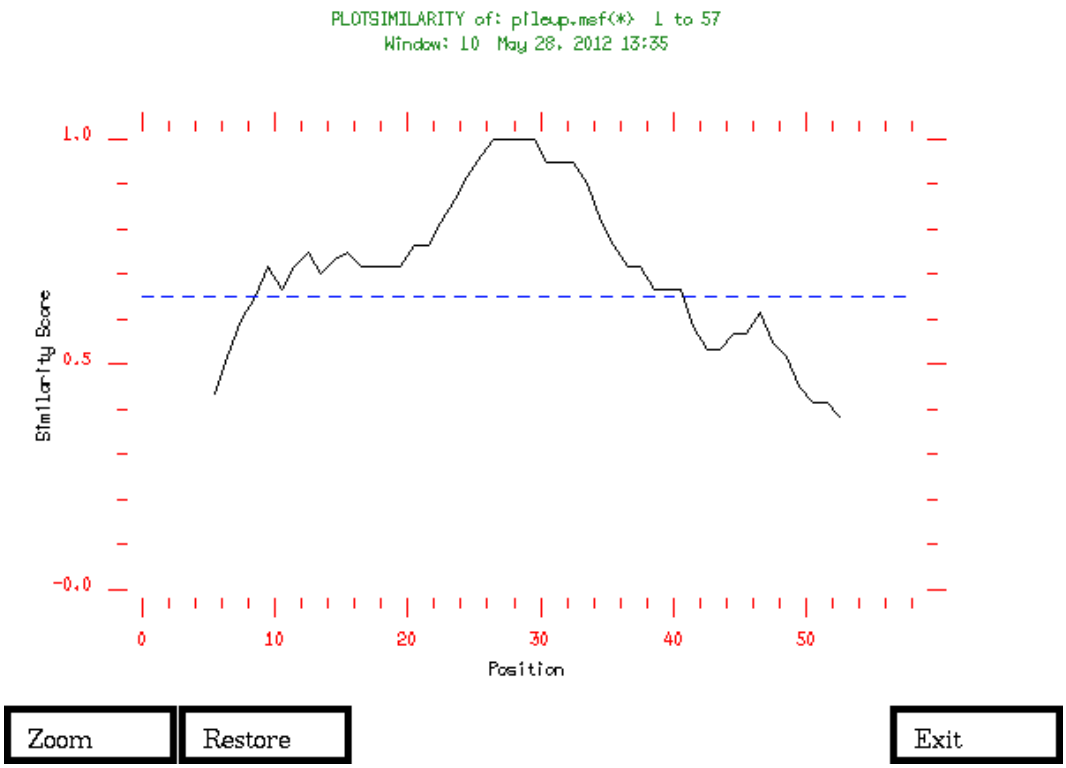

Figure 56. Graphical representation of the degree of similarity in the NCS in the intergenic region of *T08B2.7b*. Plot similarity of the conserved sequence illustrates the conserved sequence (above the blue dotted line) with a region of very high similarity, potentially the GC rich region, in the middle of the conserved sequence.

6.3.1 Modifications of the conserved non-coding sequence

To dissect out if the conserved NCS identified in CEOP1312 plays a role(s) in regulating possible transcriptional events from start points within the operon structure, fosmid-based translational reporter constructs were created where the NCS was deleted, reversed and complemented and finally replaced with a scrambled sequence of equivalent length.

The FP-RT-FP cassette was PCR-amplified (ODNs 12115/12116, Table 4) and introduced via positive-selection recombineering into fNH108 (triple-tagged CEOP1312 with F-CFP, F-YFP and Mc-mCherry FP CDSs) to create fNH109. The RT Cassette was replaced via negative-selection recombineering with an appropriate gBlock™ (IDT) creating fNH112, fNH113 and fNH114 in which the conserved NCS was, respectively, deleted, reversed and complemented and replaced with a

scrambled sequence of equivalent length. These three fosmid-based reporters were microinjected and the resulting independently generated lines imaged.

Investigation of the three modifications made to the conserved NCS revealed expression of *T08B2.7::F-CFP* in the intestine and as bright punctate foci in body wall cells. Expression of *T08B2.8* was observed in the intestine and faintly in the muscle and *T08B2.9* was visualized in muscle cells. Despite the conserved NCS being deleted (Figure 57), sequence reverse and complemented (Figure 58) and replaced with a scrambled sequence (Figure 59), no difference was visualized in the expression patterns of the three genes within the operon.

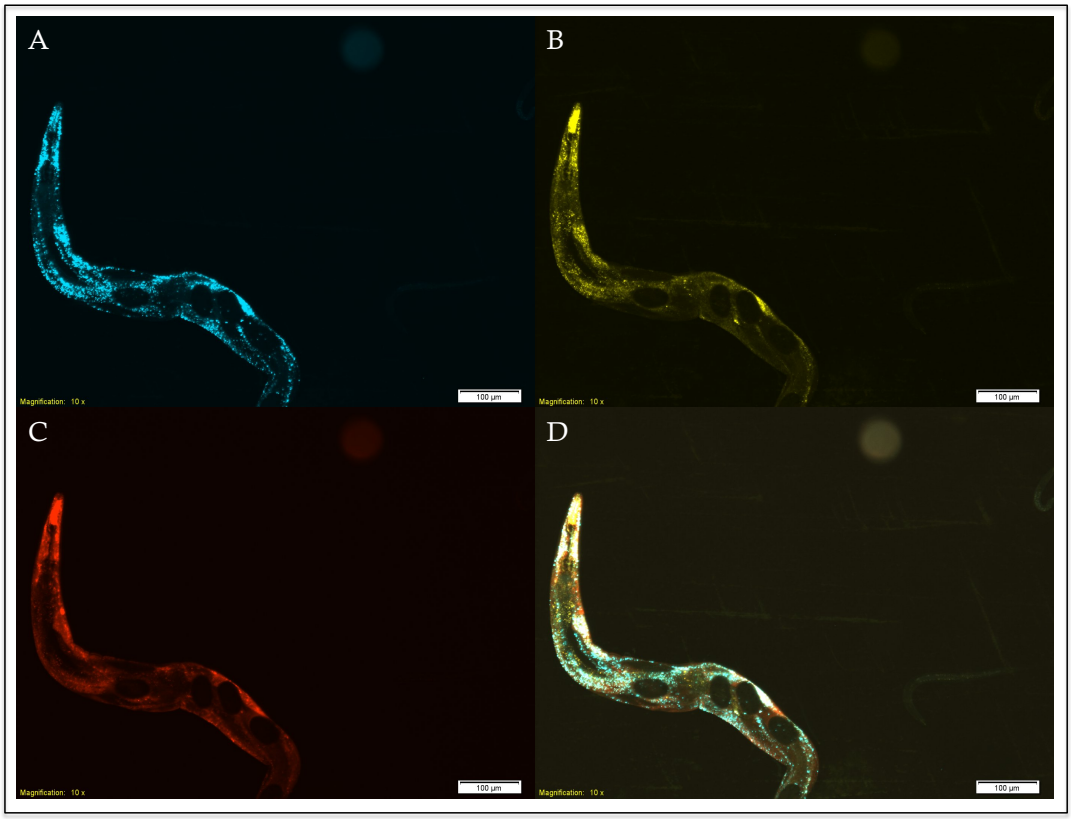

Figure 57. Expression patterns visualized when the NCS is deleted. Independently generated lines of animals, transgenic for the fosmid construct with the conserved non-coding region deleted, were imaged through filter sets respective to the FP being imaged before the images were superimposed and a composite formulated. Expression of *T08B2.7b::F-CFP*, (Panel A) was visualized as punctate spots in the body wall cells and expression of *T08B2.8::F-YFP* (Panel B) was visualized in the faintly in the intestine and muscle and *T08B2.9::Mc-mCherry* (Panel C) was observed in the hypodermis. Panel D is a composite of all the expression patterns imaged and the scale bar indicates 100 μm.

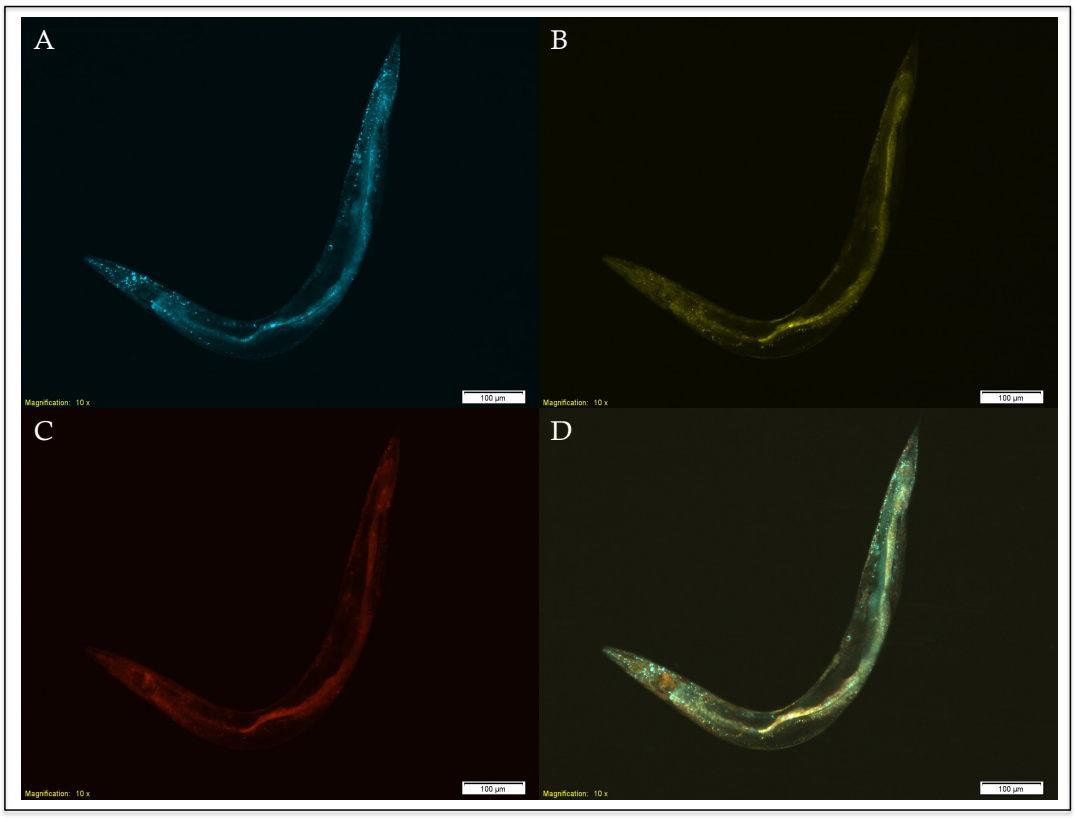

Figure 58. Expression patterns observed when the NCS is reversed and complemented. Adult hermaphrodite worms, transgenic for the fosmid construct where the conserved sequence was reversed and complemented, were imaged. Expression for *T08B2.7b::F-CFP* (Panel A) was seen as dots of expression in body wall cells and faintly in the intestine. *T08B2.8::F-YFP* (Panel B) was observed faintly in the muscle and intestine and *T08B2.9:Mc-mCherry* (Panel C) expression was visualized faintly in the intestine and muscle. Panel D is all the fluorescent images superimposed and the scale bar indicates 100 μm.

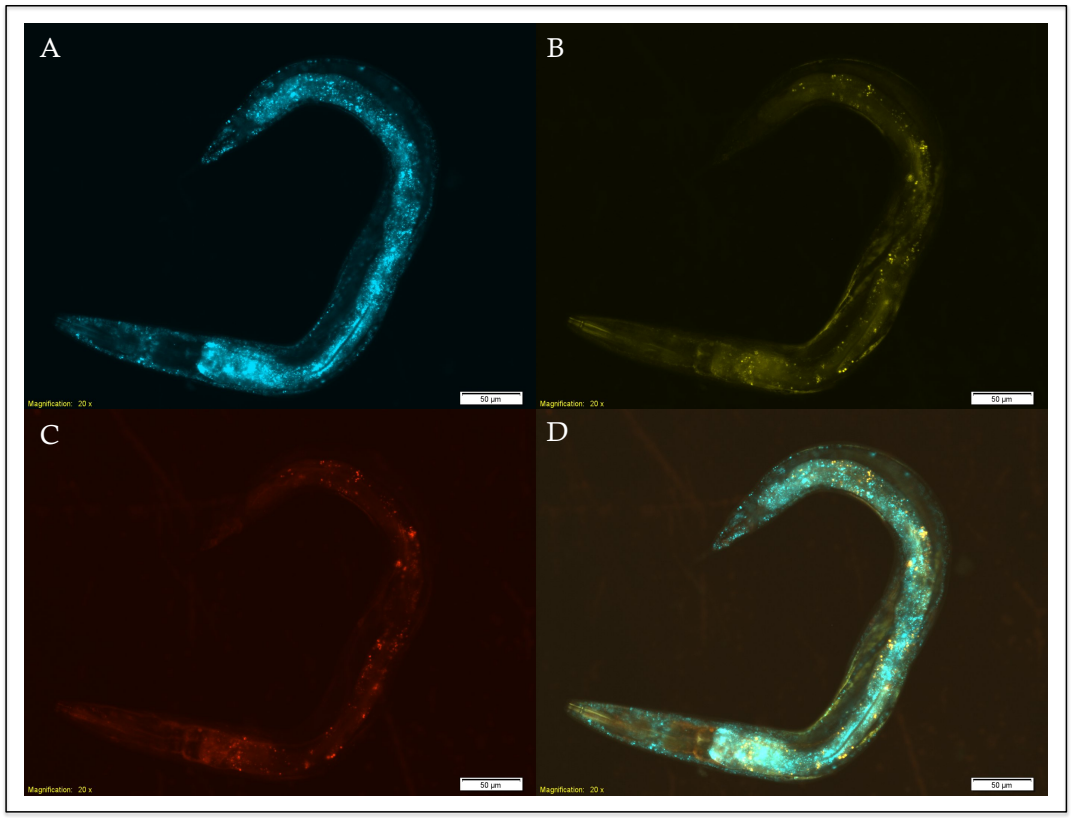

Figure 59. Expression pattern observed when the conserved region is replaced with a scrambled sequence. The resulting independently-generated transgenic lines were imaged and expression of *T08B2.7::F-CFP* (Panel A) was visualized as foci in body wall cells and brightly in the intestine, expression of *T08B2.8::F-YFP* (Panel B) was seen faintly in the muscle and intestine and *T08B2.9::Mc-mCherry* (Panel C) was faintly in muscle. All the images were superimposed (Panel D) and scale bar indicates 100 μm.

6.4 Discussion

Classical operons are clusters of genes driven by a single major upstream promoter. Operons are commonly found in prokaryotes and genes that are found in them have related functions or the proteins, translated from the genes, are required in the same pathway. Operon structures, albeit more complex, have also been identified in various eukaryotic organisms including the worm. The complexity of operons in the worm arose from the discovery of internal promoter elements in the intergenic sequences of 25 % of the operons in the worm and operons containing such elements were termed "hybrid operons". These intergenic sequences were found to drive expression of GFP and thus demonstrated as containing an internal promoter sequence (Huang et al., 2007). Data generated by Huang et al was via the use of transcriptional-style reporters that examined each excised section of the operon in

isolation. In contrast, our recombineering methods allows us to uniquely investigate the entire operon structure within the context of a genomic clone, approximately. 40 kb in size. Additionally, our method also allows us to dissect out the internal structures whilst retaining the overall genomic context of the operon structure. We chose to do so by employing counter-selection recombineering to subtly modify such sequences in the context of a fosmid reporter construct. The method of counter-selection recombineering was employed as it is the only method which would allow the subtle and precise modifications to be made without leaving a scar sequence.

From the list of potential operon targets identified (Chapter 2.6), two operons were selected for interrogation. These two operons were selected as they are complex, have large intergenic sequences (larger than the average 100 – 400bp intergenic regions found in operons) and unusual large introns which could be potential locations for extra internal promoter elements. Interrogation of the two operons CEOP1312 and CEOP1358, via bioinformatic analysis, revealed CEOP1358 to be one of the 4% of operons which are not conserved between *C. elegans* and *C. briggsae* as alignment of the two gene operon showed that the structure is not syntenic in *C. briggsae*, *C. brenneri* or *C. remanei* (Figure 50). Operon CEOP1312 on the other hand is conserved in the *Caenorhabditis* genus with the orientation and sizes of the genes along with the intergenic regions syntenic in *C. elegans*, *C. briggsae*, *C. brenneri* and *C. remanei* (Figure 51). Careful, detailed analysis of Pustell dot blots of CEOP1312 in *C. elegans*, *C. briggsae*, *C. brenneri* and *C. remanei* also identified a short stretch of conserved sequence in the non-coding region, intron/intergenic region, of T08B2.7 and further investigation highlighted a 48 bp GC-rich sequence which is conserved (Figure 54) and also appeared to overlap with some regions that were identified as TF-binding sites via ModEncode, more specifically binding sites for ALR-1, BLMP-1, GEI-11, AMA-1, HLH-8, DAF-16, EGL-5, MAB-5 and ELT-3 (Figure 53). As a conserved sequence usually infers functional relevance this short sequence was investigated for any potential regulatory role it may possess. The conserved sequence, in the triple-tagged fosmid-based translational reporter, was deleted, reversed and complemented and finally replaced with a scrambled sequence of an equivalent length. Strains transgenic for the modified conserved sequence showed no visible difference to the expression observed of the three genes within CEOP1312 when gravid hermaphrodite worms were imaged. However, worms at different life stages, e.g. L3/L4s and also starved worms were not imaged. The resolution was also limited by the imaging system and imaging with a confocal system may have revealed subtle differences.

Expression of *T08B2.7b::F-CFP* was visualized throughout the intestine and as foci in body wall cells, *T08B2.8::F-YFP* was observed in the intestine and also faintly in muscle cells and *T08B2.9::Mc-mCherry* was viewed as faint expression in the hypodermis.

The *trans*-splice site of *T08B2.7* receives both SL1 and SL2 (50 /50) (T. Blumenthal, pers. Comm) and therefore if the conserved short sequence identified does indeed play a functional role and upon modification of the sequence its role has been modified/diminished *T08B2.7* will still receive SL2 when transcribed from the operon promoter and therefore still be transcribed. Due to the method used here to visualize and determine the expression patterns from the fosmid-based translational reporters, microinjection and formation of extra-chromosomal arrays, there is no simple way to determine if expression of T08B2.7::F-CFP was changed in our transgenic strains as each extra-chromosomal array is unique and contains a different number of fusion gene copies. The transcriptional reporter expression pattern of *T08B2.7b* (transgenic strain BC15507) utilized the intergenic sequence as the promoter sequence and therefore allowed visualization of the expression pattern from the internal promoter.

The NCS is conserved within the *Caenorhabditis* genus indicating it is functional and may represent a potential internal promoter controlling the expression of one or more of the operon genes perhaps independently of the global operon promoter. The NCS identified may also, instead of an internal promoter, represent a transcriptional element that regulates the time of expression i.e. at a certain life stage, or it may influence the strength of the expression of T08B2.7, i.e. increase its expression at a specific stage in the worms' life cycle. As such, the role of this conserved sequence was investigated via recombineered translational reporters.

The expression pattern obtained from the fosmid-based translational reporter containing *T08B2.7b::F-CFP* would be a mix of the operon and internal promoter and therefore if the intense bright foci, all over the body wall cells, of fluorescence observed in the strains transgenic for *T08B2.7b::F-CFP* is due to the operon promoter it may have obscured any potential additional fluorescence resulting from the internal promoter and also any potential changes resulting from the modifications to the conserved non-coding sequence. Additional constructs, which could have been generated to test this, would have been to delete the upstream operon promoter and

observe if the internal promoter identified did drive expression of any of the operon genes and if so where F-CFP was observed. Also the NCS could have been deleted along with the upstream operon promoter to observe if the NCS does act as an internal promoter (if deleted no F-CFP should be expressed) or if it plays a role in regulating transcription (F-CFP expression would be expected).

7

Overall conclusion

This thesis has aimed to elucidate and dissect out the function of any conserved sequences, which may act as internal regulatory elements, within selected operons by examining expression patterns of genes using fusion reporter constructs generated in fosmid genomic clones. These constructs were built via counter-selection, seamless recombineering. To facilitate this two resource sets were designed and constructed to simplify and facilitate the generation of fosmid-based FP reporter constructs *via* counter-selection recombineering. The first set of resources were designed to deliver new FPs for use as reporters encoded by codon-optimized CDSs with, or without, internal artificial introns. Unfortunately, the spectral characteristics of these FPs did not come up to our expectations. mCerulean exhibited rapid photobleaching upon imaging resulting in no expression pattern data being obtained with mCerulean. The introduction of the two artificial introns into the CDS of mCerulean did not increase the fluorescent signal of mCerulean. The reasons behind this are not understood as the amino acid sequence of our codon-optimized mCerulean is the same as the published sequence. There may have been potential problems with the folding of the protein or due to the incorporation of rare codons in the FP sequence, the other cellular components may be limited during translation affecting the expression levels of the protein. mCitrine and mCherry FPs did produce a sufficient fluorescent signal from which expression pattern could be obtained and analyzed. However, when they were compared to their spectral counterparts, F-YFP and Mc-mCherry, respectively, the signal was not as bright as judged by the exposure times required to obtain images with similar fluorescent signals. Results obtained from these experiments illustrate that codon-optimization is not the consideration to maximize *in vivo* expression of the FPs. The number and placement of introns may also play an equally or more important role in obtaining maximum expression of a sequence *in vivo*. The *C. elegans* codon-optimized FP CDS variants particularly those containing two artificial introns, apart from mCerulean, can potentially be utilized in transcriptional reporters.

The second collection of constructs has been created to simplify and streamline the counter-selection recombineering process when utilizing the RT cassette. The use of the two resource sets created here allows the generation of recombineered products with the final variants free of unwanted sequence artifacts with minimal need to verify the final constructs prior to their use in the creation of transgenic animals. Although some of the FPs encoded displayed undesirable characteristics, including rapid photobleaching and poor brightness, the complete collections represent a valuable resource that simplify the overall procedure and significantly increase the

recombineering efficiency when using the RT cassette for counter-selection recombineering. This is important, especially in the negative selection step, as there is always a small amount of background. By increasing the efficiency of the negative selection step, this will increase the number of colonies containing successfully created recombinants and reduce the background.

Within *C. elegans* >17% of the genes in the 100Mb genome are found in clustered in operon like structures. The significance of the operons are not clearly understood as there are instances where the genes within the operons have related function, for example containing genes *deg-3* and *des-2* whose protein products are subunits for the acetylcholine receptor (Treinin et al., 1998), or have no relation between them. A study in 2007 by Huang and colleagues identified internal promoters in 25% of the operons they investigated. Transcriptional reporters of the operon and internal (intergenic sequence) promoters revealed two different expression patterns differing in tissue and the importance of the internal promoters were demonstrated when they were used to rescue lethal knockout phenotypes in mutant strains with deletions of the internal promoter regions.

With the resources created here, a subset were utilized and two operon targets, one containing 3 genes and the other containing two, were tagged with F-CFP, F-YFP and Mc-mCherry FP CDS 6 codons upstream of the native genes stop codon. Expression patterns of genes within operons CEOP1312 and CEOP1358, via fosmid-based translational reporters, in some cases revealed very different and distinct expression patterns when compared to the expression patterns visualized via transcriptional reporters. Fosmid-based reporters of CEOP1312, the 3 gene operon, revealed wide-spread expression through out the transgenic animal consistent with the genes within the operon which encode for respectively, a gene orthologous to human hydroxyacyl-coenzyme A dehydrogenase (*T08B2.7*), a large mitochondrial ribosome protein (*T08B2.8*) and a predicted phenylalanyl-t-RNA synthetase (*T08B2.9*) one of which is involved in metabolism (*T08B2.7*) and the remaining two are required in protein synthesis. This is consistent with the classes of gene found in *C. elegans* operons as approximately. 30% of dehydrogenases, approximately. 50% of mitochondrial and over 50% of t-RNA synthetases are found in operons (Blumenthal and Gleason, 2003). In contrast the fosmid-based translational reporters of CEOP1358, the 2 gene operon, revealed two distinct expression patterns. *T10B11.3* (zinc finger TF) expression, via F-CFP, was observed in head and tail neurons, intestine and also in the somatic gonad. Fosmid-based expression of *T10B11.2*

(predicted to encode a ceramide kinase) was visualized as bright Mc-mCherry expression in the spermathecae with additional expression in the nervous system and also faintly in muscle cells. The two genes contained within operon CEOP1358 encode for a kinase and transcription factor, and 25% and 50%, respectively, of the total number of genes within the same classes are located in operons (Blumenthal and Gleason, 2003).

Investigation of the two operons through out the sequenced *Caenorhabditis* species (*C. elegans*, *C. briggsae*, *C. brenneri* and *C. remanei*) revealed CEOP1358 to be one of the operons in the 4% that are not conserved within the *Caenorhabditis* genus (Stein et al., 2003). However CEOP1312 was identified as being conserved and alongside the three genes being conserved a short non-coding sequence was also identified to be conserved in the intergenic region of *T08B2.7b*, the gene in the operon that receives both *trans*-splice leaders (SL1 And SL2), and this region was selected for further investigation via manipulation of the sequence within a triple-tagged fosmid-based translational reporter. ModEncode data, specifically the transcription factor binding sites GFP-ChiP data, revealed that the NCS sits within the binding site of several transcription factors. Modification of the conserved non-coding sequence (deleted, inverted and replaced with a scrambled sequence) did not reveal a difference in fosmid-based expression patterns that were visualized by fluorescence microscopy. The bright punctate foci in the body wall cells were still present and if any subtle changes to any internal faint expression had occurred as a result of modifying the non-coding conserved sequence it would have be obscured by the bright foci present all over the transgenic worms. An alternative explanation to the results observed is that the non-coding conserved sequence may not play a functional role to *T08B2.7*, its activity could be a gene further downstream, and the result of altering the sequence was not visualized as only the genes in the operon were tagged with an FP CDS. An additional reason may be that the conserved non-coding sequence does not have a functional role in gene regulation and is present in the genomes by chance.

The use of fosmid-based reporter constructs, generated by counter-selection seamless recombineering, is an ideal way to investigate individual elements within operon structures for potential regulatory activity in the context of the entire operon structure contained on a single fosmid clone.

Overall the expression of the genes within the two operons selected illustrated expression of each gene in a different tissue type. As both operons had evidence to

support the hypothesis that they are both 'hybrid' operons and the expression pattern data illustrates fosmid-based expression of the genes in gravid adults are in different tissues hence the need for an internal promoter to control the expression of the gene in the different tissue. The modification of the NCS identified in operon CEOP1312 did not yield a change in the expression pattern (one which could be detected by our fluorescence set-up). In order to further investigate this, animals at different stages and also males would need to be imaged using a higher resolution fluorescence microscope i.e. confocal to be able to infer with more confidence what actually occurs upon modification of the NCS. If upon further detailed analysis of the expression patterns at the different stages, and also in males, reveals no change in the expression pattern then this raises the question of if the NCS identified is significant and if so what role it plays.

8

References

C. ELEGANS SEQUENCING CONSORTIUM. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science,* 282**,** 2012-8.

AI, H. W., HENDERSON, J. N., REMINGTON, S. J. & CAMPBELL, R. E. 2006. Directed evolution of a monomeric, bright and photostable version of Clavularia cyan fluorescent protein: structural characterization and applications in fluorescence imaging. *Biochem J,* 400**,** 531-40.

ALLEN, M. A., HILLIER, L. W., WATERSTON, R. H. & BLUMENTHAL, T. 2011. A global analysis of *C. elegans trans*-splicing. *Genome Res,* 21**,** 255-64.

ALTUN, Z. F. A. H., D. H 2005. *Handbook of C. elegans Anatomy*.

ANDREWS, J., SMITH, M., MERAKOVSKY, J., COULSON, M., HANNAN, F. & KELLY, L. E. 1996. The stoned locus of *Drosophila melanogaster* produces a dicistronic transcript and encodes two distinct polypeptides. *Genetics,* 143**,** 1699-711.

ANTOSHECHKIN, I. & STERNBERG, P. W. 2007. The versatile worm: genetic and genomic resources for *Caenorhabditis elegans* research. *Nat Rev Genet,* 8**,** 518-32.

BAEK, D., VILLEN, J., SHIN, C., CAMARGO, F., GYGI, S. & BARTEL, D. 2008. The impact of microRNAs on protein output. *Nature,* 455**,** 64 - 71.

BAMPS, S. & HOPE, I. A. 2008. Large-scale gene expression pattern analysis, in situ, in *Caenorhabditis elegans*. *Brief Funct Genomic Proteomic,* 7**,** 175-83.

BASU, U., SI, K., WARNER, J. & MAITRA, U. 2001. The *Saccharomyces cerevisiae* TIF6 gene encoding translation initiation factor 6 is required for 60S ribosomal subunit biogenesis. *Mol Cell Biol,* 21**,** 1453 - 1462.

BECKWITH, J. R. 1967. Regulation of the lac operon. Recent studies on the regulation of lactose metabolism in *Escherichia coli* support the operon model. *Science,* 156**,** 597-604.

BEREZIKOV, E., BARGMANN, C. I. & PLASTERK, R. H. 2004. Homologous gene targeting in *Caenorhabditis elegans* by biolistic transformation. *Nucleic Acids Res,* 32**,** e40.

BHATTACHARYYA, S., HABERMACHER, R., MARTINE, U., CLOSS, E. & FILIPOWICZ, W. 2006. Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell,* 125**,** 1111 - 1124.

BLUMENTHAL, T. 2005. *Trans*-splicing and operons. *WormBook* (http://www.wormbook.org/)**,** 1-9.

BLUMENTHAL, T., EVANS, D., LINK, C. D., GUFFANTI, A., LAWSON, D., THIERRY-MIEG, J., THIERRY-MIEG, D., CHIU, W. L., DUKE, K.,

KIRALY, M. & KIM, S. K. 2002. A global analysis of *Caenorhabditis elegans* operons. *Nature,* 417**,** 851-4.

BLUMENTHAL, T. & GLEASON, K. S. 2003. *Caenorhabditis elegans* operons: form and function. *Nat Rev Genet,* 4**,** 112-20.

BLUMENTHAL, T. & STEWARD, K. 1997. RNA Processing and Gene Structure.

BOULIN, T., ETCHBERGER, J. F. & HOBERT, O. 2006. Reporter gene fusions. *WormBook* (http://www.wormbook.org/)**,** 1-23.

BOULIN, T. & HOBERT, O. 2012. From genes to function: the *C. elegans* genetic toolbox. *Wiley Interdisciplinary Reviews: Developmental Biology,* 1**,** 114-137.

BRENNER, S. 1974. The genetics of *Caenorhabditis elegans*. *Genetics,* 77**,** 71-94.

BRODERSEN, P., SAKVARELIDZE-ACHARD, L., BRUUN-RASMUSSEN, M., DUNOYER, P., YAMAMOTO, Y., SIEBURTH, L. & VOINNET, O. 2008. Widespread translational inhibition by plant miRNAs and siRNAs. *Science,* 320**,** 1185 - 1190.

BUCHMAN, A. R. & BERG, P. 1988. Comparison of intron-dependent and intron-independent gene expression. *Mol Cell Biol,* 8**,** 4395-405.

BÜRGLIN, T. R., LOBOS, E. , BLAXTER, M. L. 1998. *Caenorhabditis elegans* as a model for parasitic nematodes. *Int J Parasitol,* 28**,** 395 - 411.

CAMPBELL, R. E., TOUR, O., PALMER, A. E., STEINBACH, P. A., BAIRD, G. S., ZACHARIAS, D. A. & TSIEN, R. Y. 2002. A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A,* 99**,** 7877-82.

CARTHEW, R. & SONTHEIMER, E. 2009. Origins and mechanisms of miRNAs and siRNAs. *Cell,* 136**,** 642 - 655.

CELNIKER, S. E., DILLON, L. A., GERSTEIN, M. B., GUNSALUS, K. C., HENIKOFF, S., KARPEN, G. H., KELLIS, M., LAI, E. C., LIEB, J. D., MACALPINE, D. M., MICKLEM, G., PIANO, F., SNYDER, M., STEIN, L., WHITE, K. P. & WATERSTON, R. H. 2009. Unlocking the secrets of the genome. *Nature,* 459**,** 927-30.

CHALFIE, M., JORGENSEN, E.M. 1998. *C. elegans* neuroscience: genetics to genome. *Trends Genet.,* 14**,** 506 - 512.

CHALFIE, M., TU, Y., EUSKIRCHEN, G., WARD, W. W. & PRASHER, D. C. 1994. Green fluorescent protein as a marker for gene expression. *Science,* 263**,** 802-5.

CHEKULAEVA, M. & FILIPOWICZ, W. 2009. Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. *Curr Opin Cell Biol,* 21**,** 452 - 460.

CHEN, N., HARRIS, T. W., ANTOSHECHKIN, I., BASTIANI, C., BIERI, T., BLASIAR, D., BRADNAM, K., CANARAN, P., CHAN, J., CHEN, C. K., CHEN, W. J., CUNNINGHAM, F., DAVIS, P., KENNY, E., KISHORE, R., LAWSON, D., LEE, R., MULLER, H. M., NAKAMURA, C., PAI, S., OZERSKY, P., PETCHERSKI, A., ROGERS, A., SABO, A., SCHWARZ, E. M., VAN AUKEN, K., WANG, Q., DURBIN, R., SPIETH, J., STERNBERG, P. W. & STEIN, L. D. 2005. WormBase: a comprehensive data resource for *Caenorhabditis* biology and genomics. *Nucleic Acids Res,* 33**,** D383-9.

CHEN, N., LAWSON, D., BRADNAM, K., HARRIS, T. W. & STEIN, L. D. 2004. WormBase as an integrated platform for the *C. elegans* ORFeome. *Genome Res,* 14**,** 2155-61.

CHENDRIMADA, T., FINN, K., JI, X., BAILLAT, D., GREGORY, R., LIEBHABER, S., PASQUINELLI, A. & SHIEKHATTAR, R. 2007. MicroRNA silencing through RISC recruitment of eIF6. *Nature,* 447**,** 823 - 8.

CHUDAKOV, D. M., MATZ, M. V., LUKYANOV, S. & LUKYANOV, K. A. 2010. Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues. *Physiological Reviews,* 90**,** 1103-1163.

CLARKE, N. D. & BERG, J. M. 1998. Zinc fingers in *Caenorhabditis elegans*: finding families and probing pathways. *Science,* 282**,** 2018-22.

COGHLAN, A. 2005. Nematode genome evolution. *WormBook* (http://www.wormbook.org/)**,** 1-15.

CONRAD, R., LEA, K. & BLUMENTHAL, T. 1995. SL1 *trans*-splicing specified by AU-rich synthetic RNA inserted at the 5′ end of *Caenorhabditis elegans* pre-mRNA. *RNA,* 1**,** 164-70.

CONRAD, R., LIOU, R. F. & BLUMENTHAL, T. 1993. Conversion of a *trans*-spliced *C. elegans* gene into a conventional gene by introduction of a splice donor site. *EMBO J,* 12**,** 1249-55.

CONRAD, R., THOMAS, J., SPIETH, J. & BLUMENTHAL, T. 1991. Insertion of part of an intron into the 5′ untranslated region of a *Caenorhabditis elegans* gene converts it into a *trans*-spliced gene. *Mol Cell Biol,* 11**,** 1921-6.

COPELAND, N. G., JENKINS, N. A. & COURT, D. L. 2001. Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet,* 2**,** 769-79.

CORCORAN, M. M., HAMMARSUND, M., ZHU, C., LERNER, M., KAPANADZE, B., WILSON, B., LARSSON, C., FORSBERG, L., IBBOTSON, R. E., EINHORN, S., OSCIER, D. G., GRANDER, D. &

SANGFELT, O. 2004. DLEU2 encodes an antisense RNA for the putative bicistronic RFP2/LEU5 gene in humans and mouse. *Genes Chromosomes Cancer,* 40**,** 285-97.

CUBITT, A. B., HEIM, R., ADAMS, S. R., BOYD, A. E., GROSS, L. A. & TSIEN, R. Y. 1995. Understanding, improving and using green fluorescent proteins. *Trends Biochem Sci,* 20**,** 448-55.

DENG, X., YIN, X., ALLAN, R., LU, D. D., MAURER, C. W., HAIMOVITZ-FRIEDMAN, A., FUKS, Z., SHAHAM, S. & KOLESNICK, R. 2008. Ceramide biogenesis is required for radiation-induced apoptosis in the germ line of *C. elegans*. *Science,* 322**,** 110-5.

DING, L. & HAN, M. 2007. GW182 family proteins are crucial for microRNA-mediated gene silencing. *Trends Cell Biol,* 17**,** 411 - 416.

DING, X. & GROSSHANS, H. 2009. Repression of *C. elegans microRNA targets at the initiation level of translation requires GW182 proteins*. *Embo J,* 28**,** 213 - 222.

DJURANOVIC, S., ZINCHENKO, M., HUR, J., NAHVI, A., BRUNELLE, J., ROGERS, E. & GREEN, R. 2010. Allosteric regulation of Argonaute proteins by miRNAs. *Nat Struct Mol Biol,* 17**,** 144 - 50.

DOLPHIN, C. T. & HOPE, I. A. 2006. *Caenorhabditis elegans* reporter fusion genes generated by seamless modification of large genomic DNA clones. *Nucleic Acids Res,* 34**,** e72.

DUERR, J. S. 2006. Immunohistochemistry. *WormBook* (http://www.wormbook.org/)**,** 1-61.

EECKMAN, F. H. & DURBIN, R. 1995. ACeDB and macace. *Methods Cell Biol,* 48**,** 583-605.

EULALIO, A., BEHM-ANSMANT, I., SCHWEIZER, D. & IZAURRALDE, E. 2007. P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Mol Cell Biol,* 27**,** 3970 - 3981.

EVANS, D., ZORIO, D., MACMORRIS, M., WINTER, C. E., LEA, K. & BLUMENTHAL, T. 1997. Operons and SL2 *trans*-splicing exist in nematodes outside the genus *Caenorhabditis*. *Proc Natl Acad Sci U S A,* 94**,** 9751-6.

FILIPOWICZ, W., BHATTACHARYYA, S. & SONENBERG, N. 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet,* 9**,** 102 - 114.

FIRE, A. 1986. Integrative transformation of *Caenorhabditis elegans*. *EMBO J,* 5**,** 2673-80.

FITCH, D. H. 2005. Introduction to nematode evolution and ecology. *WormBook* (http://www.wormbook.org/), 1-8.

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, 325-30.

GANDIN, V., MILUZIO, A., BARBIERI, A., BEUGNET, A., KIYOKAWA, H., MARCHISIO, P. & BIFFO, S. 2008. Eukaryotic initiation factor 6 is rate-limiting in translation, growth and transformation. *Nature,* 455, 684 - 688.

GARCIA-RIOS, M., FUJITA, T., LAROSA, P. C., LOCY, R. D., CLITHERO, J. M., BRESSAN, R. A. & CSONKA, L. N. 1997. Cloning of a polycistronic cDNA from tomato encoding gamma-glutamyl kinase and gamma-glutamyl phosphate reductase. *Proc Natl Acad Sci U S A,* 94, 8249-54.

GERSTEIN, M. B., LU, Z. J., VAN NOSTRAND, E. L., CHENG, C., ARSHINOFF, B. I., LIU, T., YIP, K. Y., ROBILOTTO, R., RECHTSTEINER, A., IKEGAMI, K., ALVES, P., CHATEIGNER, A., PERRY, M., MORRIS, M., AUERBACH, R. K., FENG, X., LENG, J., VIELLE, A., NIU, W., RHRISSORRAKRAI, K., AGARWAL, A., ALEXANDER, R. P., BARBER, G., BRDLIK, C. M., BRENNAN, J., BROUILLET, J. J., CARR, A., CHEUNG, M. S., CLAWSON, H., CONTRINO, S., DANNENBERG, L. O., DERNBURG, A. F., DESAI, A., DICK, L., DOSE, A. C., DU, J., EGELHOFER, T., ERCAN, S., EUSKIRCHEN, G., EWING, B., FEINGOLD, E. A., GASSMANN, R., GOOD, P. J., GREEN, P., GULLIER, F., GUTWEIN, M., GUYER, M. S., HABEGGER, L., HAN, T., HENIKOFF, J. G., HENZ, S. R., HINRICHS, A., HOLSTER, H., HYMAN, T., INIGUEZ, A. L., JANETTE, J., JENSEN, M., KATO, M., KENT, W. J., KEPHART, E., KHIVANSARA, V., KHURANA, E., KIM, J. K., KOLASINSKA-ZWIERZ, P., LAI, E. C., LATORRE, I., LEAHEY, A., LEWIS, S., LLOYD, P., LOCHOVSKY, L., LOWDON, R. F., LUBLING, Y., LYNE, R., MACCOSS, M., MACKOWIAK, S. D., MANGONE, M., MCKAY, S., MECENAS, D., MERRIHEW, G., MILLER, D. M., 3RD, MUROYAMA, A., MURRAY, J. I., OOI, S. L., PHAM, H., PHIPPEN, T., PRESTON, E. A., RAJEWSKY, N., RATSCH, G., ROSENBAUM, H., ROZOWSKY, J., RUTHERFORD, K., RUZANOV, P., SAROV, M., SASIDHARAN, R., SBONER, A., SCHEID, P., SEGAL, E., SHIN, H., SHOU, C., SLACK, F.

J., et al. 2010. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science,* 330**,** 1775-87.

GOEDHART, J., VON STETTEN, D., NOIRCLERC-SAVOYE, M., LELIMOUSIN, M., JOOSEN, L., HINK, M. A., VAN WEEREN, L., GADELLA, T. W., JR. & ROYANT, A. 2012. Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nat Commun,* 3**,** 751.

GU, S., JIN, L., ZHANG, F., SARNOW, P. & KAY, M. 2009. Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nat Struct Mol Biol,* 16**,** 144 - 150.

GU, S. & KAY, M. 2010. How do miRNAs mediate translational repression? *Silence,* 1**,** 11.

GUSTAFSSON, C., GOVINDARAJAN, S., & MINSHULL, M. 2004. Codon bias and heterologous protein expression. *Trends in Biotech*. 22 (7) 346 – 353.

GUSTAFSSON, C., MINSHULL, J., GOVINDARAJAN, S., NESS, J., VILLALOBOS, A., & WELCH, M. 2012. Engineering genes for predictable protein expression. *Protein Expression and Purification* 83, 37–46.

GUPTA, B. P., JOHNSEN, R. & CHEN, N. 2007. Genomics and biology of the nematode *Caenorhabditis briggsae*. *WormBook* (http://www.wormbook.org/)**,** 1-16.

HARRIS, T. W., ANTOSHECHKIN, I., BIERI, T., BLASIAR, D., CHAN, J., CHEN, W. J., DE LA CRUZ, N., DAVIS, P., DUESBURY, M., FANG, R., FERNANDES, J., HAN, M., KISHORE, R., LEE, R., MULLER, H. M., NAKAMURA, C., OZERSKY, P., PETCHERSKI, A., RANGARAJAN, A., ROGERS, A., SCHINDELMAN, G., SCHWARZ, E. M., TULI, M. A., VAN AUKEN, K., WANG, D., WANG, X., WILLIAMS, G., YOOK, K., DURBIN, R., STEIN, L. D., SPIETH, J. & STERNBERG, P. W. 2010. WormBase: a comprehensive resource for nematode research. *Nucleic Acids Res,* 38**,** D463-7.

HARRIS, T. W., CHEN, N., CUNNINGHAM, F., TELLO-RUIZ, M., ANTOSHECHKIN, I., BASTIANI, C., BIERI, T., BLASIAR, D., BRADNAM, K., CHAN, J., CHEN, C. K., CHEN, W. J., DAVIS, P., KENNY, E., KISHORE, R., LAWSON, D., LEE, R., MULLER, H. M., NAKAMURA, C., OZERSKY, P., PETCHERSKI, A., ROGERS, A., SABO, A., SCHWARZ, E. M., VAN AUKEN, K., WANG, Q., DURBIN, R., SPIETH, J., STERNBERG, P. W. & STEIN, L. D. 2004. WormBase: a

multi-species resource for nematode biology and genomics. *Nucleic Acids Res,* 32**,** D411-7.

HARRIS, T. W., LEE, R., SCHWARZ, E., BRADNAM, K., LAWSON, D., CHEN, W., BLASIER, D., KENNY, E., CUNNINGHAM, F., KISHORE, R., CHAN, J., MULLER, H. M., PETCHERSKI, A., THORISSON, G., DAY, A., BIERI, T., ROGERS, A., CHEN, C. K., SPIETH, J., STERNBERG, P., DURBIN, R. & STEIN, L. D. 2003. WormBase: a cross-species database for comparative genomics. *Nucleic Acids Res,* 31**,** 133-7.

HEIM, R., PRASHER, D. C. & TSIEN, R. Y. 1994. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc Natl Acad Sci U S A,* 91**,** 12501-4.

HEIM, R. & TSIEN, R. Y. 1996. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr Biol,* 6**,** 178-82.

HOBERT, O. 2002. PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans*. *Biotechniques,* 32**,** 728-30.

HOPE, I. A. 1999. *C. elegans: A Practical Approach*, Oxford University Press.

HORTON, R. M., HUNT, H. D., HO, S. N., PULLEN, J. K. & PEASE, L. R. 1989. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene,* 77**,** 61-8.

HUANG, P., PLEASANCE, E. D., MAYDAN, J. S., HUNT-NEWBURY, R., O'NEIL, N. J., MAH, A., BAILLIE, D. L., MARRA, M. A., MOERMAN, D. G. & JONES, S. J. 2007. Identification and analysis of internal promoters in *Caenorhabditis elegans* operons. *Genome Res,* 17**,** 1478-85.

HUMPHREYS, D., WESTMAN, B., MARTIN, D. & PREISS, T. 2005. MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proc Natl Acad Sci USA,* 102**,** 16961 - 16966.

HUNT-NEWBURY, R., VIVEIROS, R., JOHNSEN, R., MAH, A., ANASTAS, D., FANG, L., HALFNIGHT, E., LEE, D., LIN, J., LORCH, A., MCKAY, S., OKADA, H. M., PAN, J., SCHULZ, A. K., TU, D., WONG, K., ZHAO, Z., ALEXEYENKO, A., BURGLIN, T., SONNHAMMER, E., SCHNABEL, R., JONES, S. J., MARRA, M. A., BAILLIE, D. L. & MOERMAN, D. G. 2007. High-throughput in vivo analysis of gene expression in *Caenorhabditis elegans*. *PLoS Biol,* 5**,** e237.

IBDAH, J. A., DASOUKI, M. J. & STRAUSS, A. W. 1999. Long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency: variable expressivity of maternal illness during pregnancy and unusual presentation with infantile cholestasis and hypocalcaemia. *J Inherit Metab Dis,* 22**,** 811-4.

IBDAH, J. A., PAUL, H., ZHAO, Y., BINFORD, S., SALLENG, K., CLINE, M., MATERN, D., BENNETT, M. J., RINALDO, P. & STRAUSS, A. W. 2001. Lack of mitochondrial trifunctional protein in mice causes neonatal hypoglycemia and sudden death. *J Clin Invest,* 107**,** 1403-9.

IMBODEN, M. A., LAIRD, P. W., AFFOLTER, M. & SEEBECK, T. 1987. Transcription of the intergenic regions of the tubulin gene cluster of *Trypanosoma brucei*: evidence for a polycistronic transcription unit in a eukaryote. *Nucleic Acids Res,* 15**,** 7357-68.

INOUYE, S. & TSUJI, F. I. 1994. Aequorea green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein. *FEBS Lett,* 341**,** 277-80.

IWASAKI, S., KAWAMATA, T. & TOMARI, Y. 2009. Drosophila argonaute1 and argonaute2 employ distinct mechanisms for translational repression. *Mol Cell,* 34**,** 58 - 67.

KAMATH, R. S., FRASER, A. G., DONG, Y., POULIN, G., DURBIN, R., GOTTA, M., KANAPIN, A., LE BOT, N., MORENO, S., SOHRMANN, M., WELCHMAN, D. P., ZIPPERLEN, P. & AHRINGER, J. 2003. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature,* 421**,** 231-7.

KAMATH, R. S., MARTINEZ-CAMPOS, M., ZIPPERLEN, P., FRASER, A. G. & AHRINGER, J. 2001. Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans*. *Genome Biol,* 2**,** RESEARCH0002.

KESHAVA, N. & MUSTARD, J. F. 2002. Spectral unmixing. *Signal Processing Magazine, IEEE,* 19**,** 44-57.

KIM, S. K. 2001. Functional genomics: the worm scores a knockout. *Curr Biol.,* 11**,** R85 - 87.

KIMURA, Y., KURABE, N., IKEGAMI, K., TSUTSUMI, K., KONISHI, Y., KAPLAN, O. I., KUNITOMO, H., IINO, Y., BLACQUE, O. E. & SETOU, M. 2010. Identification of tubulin deglutamylase among *Caenorhabditis elegans* and mammalian cytosolic carboxypeptidases (CCPs). *J Biol Chem,* 285**,** 22936-41.

KINCH, L. & GRISHIN, N. 2009. The human Ago2 MC region does not contain an eIF4E-like mRNA cap binding motif. *Biol Direct,* 4**,** 2.

KIRIAKIDOU, M., TAN, G., LAMPRINAKI, S., DE PLANELL-SAGUER, M., NELSON, P. & MOURELATOS, Z. 2007. An mRNA m7G cap binding-like motif within human Ago2 represses translation. *Cell,* 129**,** 1141 - 1151.

KONG, Y., CANNELL, I., DE MOOR, C., HILL, K., GARSIDE, P., HAMILTON, T., MEIJER, H., DOBBYN, H., STONELEY, M., SPRIGGS, K., WILLIS, A. & BUSHELL, M. 2008. The mechanism of micro-RNA-mediated translation repression is determined by the promoter of the target gene. *Proc Natl Acad Sci USA,* 105**,** 8866 - 8871.

KRAUSE, M. & HIRSH, D. 1987. A *trans*-spliced leader sequence on actin mRNA in *C. elegans*. *Cell,* 49**,** 753-61.

KUBAGAWA, H. M., WATTS, J. L., CORRIGAN, C., EDMONDS, J. W., SZTUL, E., BROWSE, J. & MILLER, M. A. 2006. Oocyte signals derived from polyunsaturated fatty acids control sperm recruitment in vivo. *Nat Cell Biol,* 8**,** 1143-8.

LAMESCH, P., MILSTEIN, S., HAO, T., ROSENBERG, J., LI, N., SEQUERRA, R., BOSAK, S., DOUCETTE-STAMM, L., VANDENHAUTE, J., HILL, D. E. & VIDAL, M. 2004. *C. elegans* ORFeome version 3.1: increasing the coverage of ORFeome resources with improved gene predictions. *Genome Res,* 14**,** 2064-9.

LANET, E., DELANNOY, E., SORMANI, R., FLORIS, M., BRODERSEN, P., CRETE, P., VOINNET, O. & ROBAGLIA, C. 2009. Biochemical evidence for translational repression by Arabidopsis microRNAs. *Plant Cell,* 21**,** 1762 - 1768.

LI, Y. 2010. Commonly used tag combinations for tandem affinity purification. *Biotechnol Appl Biochem,* 55**,** 73-83.

LI, Y. 2011. The tandem affinity purification technology: an overview. *Biotechnol Lett,* 33**,** 1487-99.

LINTS, R. A. H., D. H. 2005. *Handbook of C. elegans Male Anatomy*.

LIU, H., JANG, J. K., GRAHAM, J., NYCZ, K. & MCKIM, K. S. 2000. Two genes required for meiotic recombination in Drosophila are expressed from a dicistronic message. *Genetics,* 154**,** 1735-46.

LIU, Y., KUERSTEN, S., HUANG, T., LARSEN, A., MACMORRIS, M. & BLUMENTHAL, T. 2003. An uncapped RNA suggests a model for *Caenorhabditis elegans* polycistronic pre-mRNA processing. *RNA,* 9**,** 677-87.

LONGMAN, D., ARRISI, P., JOHNSTONE, I. L. & CACERES, J. F. 2008. Chapter 7. Nonsense-mediated mRNA decay in *Caenorhabditis elegans*. *Methods Enzymol,* 449**,** 149-64.

LYTLE, J., YARIO, T. & STEITZ, J. 2007. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5′ UTR as in the 3' UTR. *Proc Natl Acad Sci USA,* 104**,** 9667 - 9672.

MACMORRIS, M., KUMAR, M., LASDA, E., LARSEN, A., KRAEMER, B. & BLUMENTHAL, T. 2007. A novel family of *C. elegans* snRNPs contains proteins associated with *trans*-splicing. *RNA,* 13**,** 511-20.

MANGO, S. E. 2001. Stop making nonSense: the *C. elegans smg* genes. *Trends Genet,* 17**,** 646-53.

MARKWARDT, M. L., KREMERS, G. J., KRAFT, C. A., RAY, K., CRANFILL, P. J., WILSON, K. A., DAY, R. N., WACHTER, R. M., DAVIDSON, M. W. & RIZZO, M. A. 2011. An improved cerulean fluorescent protein with enhanced brightness and reduced reversible photoswitching. *PLoS One,* 6**,** e17896.

MARONEY, P., YU, Y., FISHER, J. & NILSEN, T. 2006. Evidence that microRNAs are associated with translating messenger RNAs in human cells. *Nat Struct Mol Biol,* 13**,** 1102 - 1107.

MATHONNET, G., FABIAN, M., SVITKIN, Y., PARSYAN, A., HUCK, L., MURATA, T., BIFFO, S., MERRICK, W., DARZYNKIEWICZ, E., PILLAI, R., FILIPOWICZ, W., DUCHAINE, T. & SONENBERG, N. 2007. MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F. *Science,* 317**,** 1764 - 1767.

MATZ, M. V., FRADKOV, A. F., LABAS, Y. A., SAVITSKY, A. P., ZARAISKY, A. G., MARKELOV, M. L. & LUKYANOV, S. A. 1999. Fluorescent proteins from nonbioluminescent *Anthozoa* species. *Nat Biotechnol,* 17**,** 969-73.

MCKAY, S. J., JOHNSEN, R., KHATTRA, J., ASANO, J., BAILLIE, D. L., CHAN, S., DUBE, N., FANG, L., GOSZCZYNSKI, B., HA, E., HALFNIGHT, E., HOLLEBAKKEN, R., HUANG, P., HUNG, K., JENSEN, V., JONES, S. J., KAI, H., LI, D., MAH, A., MARRA, M., MCGHEE, J., NEWBURY, R., POUZYREV, A., RIDDLE, D. L., SONNHAMMER, E., TIAN, H., TU, D., TYSON, J. R., VATCHER, G., WARNER, A., WONG, K., ZHAO, Z. & MOERMAN, D. G. 2003. Gene expression profiling of cells, tissues, and developmental stages of the nematode *C. elegans*. *Cold Spring Harb Symp Quant Biol,* 68**,** 159-69.

MCNALLY, K., AUDHYA, A., OEGEMA, K. & MCNALLY, F. J. 2006. Katanin controls mitotic and meiotic spindle length. *J Cell Biol,* 175**,** 881-91.

MELLO, C. & FIRE, A. 1995. DNA transformation. *Methods Cell Biol,* 48**,** 451-82.

MELLO, C. C., KRAMER, J. M., STINCHCOMB, D. & AMBROS, V. 1991. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J,* 10**,** 3959-70.

MERZLYAK, E. M., GOEDHART, J., SHCHERBO, D., BULINA, M. E., SHCHEGLOV, A. S., FRADKOV, A. F., GAINTZEVA, A., LUKYANOV, K. A., LUKYANOV, S., GADELLA, T. W. & CHUDAKOV, D. M. 2007. Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat Methods,* 4**,** 555-7.

MOERSCHELL, R. P., TSUNASAWA, S. & SHERMAN, F. 1988. Transformation of yeast with synthetic oligonucleotides. *Proc Natl Acad Sci U S A,* 85**,** 524-8.

NAGAI, T., IBATA, K., PARK, E. S., KUBOTA, M., MIKOSHIBA, K. & MIYAWAKI, A. 2002. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat Biotechnol,* 20**,** 87-90.

NISSAN, T. & PARKER, R. 2008. Computational analysis of miRNA-mediated repression of translation: implications for models of translation initiation inhibition. *Rna,* 14**,** 1480 - 1491.

NOTTROTT, S., SIMARD, M. & RICHTER, J. 2006. Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nat Struct Mol Biol,* 13**,** 1108 - 1114.

OLAHOVA, M., TAYLOR, S. R., KHAZAIPOUL, S., WANG, J., MORGAN, B. A., MATSUMOTO, K., BLACKWELL, T. K. & VEAL, E. A. 2008. A redox-sensitive peroxiredoxin that is important for longevity has tissue- and stress-specific roles in stress resistance. *Proc Natl Acad Sci U S A,* 105**,** 19839-44.

OLSEN, P. & AMBROS, V. 1999. The lin-4 regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol,* 216**,** 671 - 680.

ORMO, M., CUBITT, A. B., KALLIO, K., GROSS, L. A., TSIEN, R. Y. & REMINGTON, S. J. 1996. Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science,* 273**,** 1392-5.

ORR-WEAVER, T. L., SZOSTAK, J. W. & ROTHSTEIN, R. J. 1983. Genetic applications of yeast transformation with linear and gapped plasmids. *Methods Enzymol,* 101**,** 228-45.

PETERSEN, C., BORDELEAU, M., PELLETIER, J. & SHARP, P. 2006. Short RNAs repress translation after initiation in mammalian cells. *Mol Cell,* 21**,** 533 - 542.

PILLAI, R., ARTUS, C. & FILIPOWICZ, W. 2004. Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. *Rna,* 10**,** 1518 - 1525.

PILLAI, R., BHATTACHARYYA, S., ARTUS, C., ZOLLER, T., COUGOT, N., BASYUK, E., BERTRAND, E. & FILIPOWICZ, W. 2005. Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science,* 309**,** 1573 - 1576.

PORTMAN, D. S. 2006. Profiling *C. elegans* gene expression with DNA microarrays. *WormBook* (http://www.wormbook.org/)**,** 1-11.

PUIG, O., CASPARY, F., RIGAUT, G., RUTZ, B., BOUVERET, E., BRAGADO-NILSSON, E., WILM, M. & SERAPHIN, B. 2001. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods,* 24**,** 218-29.

PULAK, R. & ANDERSON, P. 1993. mRNA surveillance by the *Caenorhabditis elegans smg* genes. *Genes Dev,* 7**,** 1885-97.

PUSTELL, J. & KAFATOS, F. C. 1984. A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis and homology determination. *Nucleic Acids Res,* 12**,** 643-55.

QIAN, W. & ZHANG, J. 2008. Evolutionary dynamics of nematode operons: easy come, slow go. *Genome Res,* 18**,** 412-21.

REECE-HOYES, J. S., SHINGLES, J., DUPUY, D., GROVE, C. A., WALHOUT, A. J., VIDAL, M. & HOPE, I. A. 2007. Insight into transcription factor gene duplication from *Caenorhabditis elegans* Promoterome-driven expression patterns. *BMC Genomics,* 8**,** 27.

REINKE, V. & CUTTER, A. D. 2009. Germline expression influences operon organization in the *Caenorhabditis elegans* genome. *Genetics,* 181**,** 1219-28.

RIDDLE, D. L., BLUMENTHAL, T., MEYER, B.J. , AND PRIESS, J.R. 1997. *C. elegans II*, Cold Spring Harbor Laboratory Press.

RIGAUT, G., SHEVCHENKO, A., RUTZ, B., WILM, M., MANN, M. & SERAPHIN, B. 1999. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol,* 17**,** 1030-2.

RIZZO, M. A., DAVIDSON, M. W. & PISTON, D. W. 2009. Fluorescent protein tracking and detection: fluorescent protein structure and color variants. *Cold Spring Harb Protoc,* 2009**,** pdb top63.

RIZZO, M. A., SPRINGER, G. H., GRANADA, B. & PISTON, D. W. 2004. An improved cyan fluorescent protein variant useful for FRET. *Nat Biotechnol,* 22**,** 445-9.

ROGERS, A., ANTOSHECHKIN, I., BIERI, T., BLASIAR, D., BASTIANI, C., CANARAN, P., CHAN, J., CHEN, W. J., DAVIS, P., FERNANDES, J., FIEDLER, T. J., HAN, M., HARRIS, T. W., KISHORE, R., LEE, R., MCKAY, S., MULLER, H. M., NAKAMURA, C., OZERSKY, P., PETCHERSKI, A., SCHINDELMAN, G., SCHWARZ, E. M., SPOONER, W., TULI, M. A., VAN AUKEN, K., WANG, D., WANG, X., WILLIAMS, G., YOOK, K., DURBIN, R., STEIN, L. D., SPIETH, J. & STERNBERG, P. W. 2008. WormBase 2007. *Nucleic Acids Res,* 36**,** D612-7.

SAROV, M., SCHNEIDER, S., POZNIAKOVSKI, A., ROGUEV, A., ERNST, S., ZHANG, Y., HYMAN, A. A. & STEWART, A. F. 2006. A recombineering pipeline for functional genomics applied to *Caenorhabditis elegans*. *Nat Methods,* 3**,** 839-44.

SAWITZKE, J. A., THOMASON, L. C., COSTANTINO, N., BUBUNENKO, M., DATTA, S. & COURT, D. L. 2007. Recombineering: in vivo genetic engineering in *E. coli, S. enterica*, and beyond. *Methods Enzymol,* 421**,** 171-99.

SCHNEIDER, T. D. & STEPHENS, R. M. 1990. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res,* 18**,** 6097-100.

SEGGERSON, K., TANG, L. & MOSS, E. 2002. Two genetic circuits repress the *Caenorhabditis elegans* heterochronic gene lin-28 after translation initiation. *Dev Biol,* 243**,** 215 - 225.

SELBACH, M., SCHWANHAUSSER, B., THIERFELDER, N., FANG, Z., KHANIN, R. & RAJEWSKY, N. 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature,* 455**,** 58 - 63.

SHANER, N. C., CAMPBELL, R. E., STEINBACH, P. A., GIEPMANS, B. N., PALMER, A. E. & TSIEN, R. Y. 2004. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma sp*. red fluorescent protein. *Nat Biotechnol,* 22**,** 1567-72.

SHANER, N. C., LIN, M. Z., MCKEOWN, M. R., STEINBACH, P. A., HAZELWOOD, K. L., DAVIDSON, M. W. & TSIEN, R. Y. 2008.

Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nat Methods,* 5**,** 545-51.

SHANER, N. C., PATTERSON, G. H. & DAVIDSON, M. W. 2007. Advances in fluorescent protein technology. *J Cell Sci,* 120**,** 4247-60.

SHANER, N. C., STEINBACH, P. A. & TSIEN, R. Y. 2005. A guide to choosing fluorescent proteins. *Nat Methods,* 2**,** 905-9.

SHARAN, S. K., THOMASON, L. C., KUZNETSOV, S. G. & COURT, D. L. 2009. Recombineering: a homologous recombination-based method of genetic engineering. *Nat Protoc,* 4**,** 206-23.

SHAYE, D. D. & GREENWALD, I. 2011. OrthoList: a compendium of *C. elegans* genes with human orthologs. *PLoS One,* 6**,** e20085.

SHCHERBO, D., MERZLYAK, E. M., CHEPURNYKH, T. V., FRADKOV, A. F., ERMAKOVA, G. V., SOLOVIEVA, E. A., LUKYANOV, K. A., BOGDANOVA, E. A., ZARAISKY, A. G., LUKYANOV, S. & CHUDAKOV, D. M. 2007. Bright far-red fluorescent protein for whole-body imaging. *Nat Methods,* 4**,** 741-6.

SHEN, P., AND HUANG, H. V. 1986. Homologous recombination in *Escherichia coli*: Dependence on substrate length and homology. *Genetics* 112, 441–457.

SILHAVY, T. J. & BECKWITH, J. R. 1985. Uses of lac fusions for the study of biological problems. *Microbiol Rev,* 49**,** 398-418.

SLOAN, J., KINGHORN, J. R. & UNKLES, S. E. 1999. The two subunits of human molybdopterin synthase: evidence for a bicistronic messenger RNA with overlapping reading frames. *Nucleic Acids Res,* 27**,** 854-8.

SOMMER, R. J. 2005. Evolution of development in nematodes related to *C. elegans*. *WormBook* (http://www.wormbook.org/)**,** 1-17.

SONENBERG, N. & HINNEBUSCH, A. 2009. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell,* 136**,** 731 - 745.

SPIETH, J., BROOKE, G., KUERSTEN, S., LEA, K. & BLUMENTHAL, T. 1993. Operons in *C. elegans*: polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. *Cell,* 73**,** 521-32.

SPIETH, J. & LAWSON, D. 2006. Overview of gene structure. *WormBook* (http://www.wormbook.org/)**,** 1-10.

STAVROPOULOS, T. A. & STRATHDEE, C. A. 2001. Synergy between tetA and rpsL provides high-stringency positive and negative selection in bacterial artificial chromosome vectors. *Genomics,* 72**,** 99-104.

STEIN, L., STERNBERG, P., DURBIN, R., THIERRY-MIEG, J. & SPIETH, J. 2001. WormBase: network access to the genome and biology of *Caenorhabditis elegans*. *Nucleic Acids Res,* 29**,** 82-6.

STEIN, L. D., BAO, Z., BLASIAR, D., BLUMENTHAL, T., BRENT, M. R., CHEN, N., CHINWALLA, A., CLARKE, L., CLEE, C., COGHLAN, A., COULSON, A., D'EUSTACHIO, P., FITCH, D. H., FULTON, L. A., FULTON, R. E., GRIFFITHS-JONES, S., HARRIS, T. W., HILLIER, L. W., KAMATH, R., KUWABARA, P. E., MARDIS, E. R., MARRA, M. A., MINER, T. L., MINX, P., MULLIKIN, J. C., PLUMB, R. W., ROGERS, J., SCHEIN, J. E., SOHRMANN, M., SPIETH, J., STAJICH, J. E., WEI, C., WILLEY, D., WILSON, R. K., DURBIN, R. & WATERSTON, R. H. 2003. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol,* 1**,** E45.

SU, H., TROMBLY, M., CHEN, J. & WANG, X. 2009. Essential and overlapping functions for mammalian Argonautes in microRNA silencing. *Genes Dev,* 23**,** 304 - 317.

SUGIURA, M., KONO, K., LIU, H., SHIMIZUGAWA, T., MINEKURA, H., SPIEGEL, S. & KOHAMA, T. 2002. Ceramide kinase, a novel lipid kinase. Molecular cloning and functional characterization. *J Biol Chem,* 277**,** 23294-300.

SULSTON, J. E., SCHIERENBERG, E., WHITE, J. G. & THOMSON, J. N. 1983. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Developmental Biology,* 100**,** 64-119.

THERMANN, R. & HENTZE, M. 2007. Drosophila miR2 induces pseudo-polysomes and inhibits translation initiation. *Nature,* 47**,** 875 - 8.

THIMMAPURAM, J., DUAN, H., LIU, L. & SCHULER, M. A. 2005. Bicistronic and fused monocistronic transcripts are derived from adjacent loci in the *Arabidopsis* genome. *RNA,* 11**,** 128-38.

THOMASON, L., COURT, D. L., BUBUNENKO, M., COSTANTINO, N., WILSON, H., DATTA, S. & OPPENHEIM, A. 2007. Recombineering: genetic engineering in bacteria using homologous recombination. *Curr Protoc Mol Biol,* Chapter 1**,** Unit 1 16.

TREININ, M., GILLO, B., LIEBMAN, L. & CHALFIE, M. 1998. Two functionally dependent acetylcholine subunits are encoded in a single *Caenorhabditis elegans* operon. *Proc Natl Acad Sci U S A,* 95**,** 15492-5.

TSANG, P., GILLES, F., YUAN, L., KUO, Y. H., LUPU, F., SAMARA, G., MOOSIKASUWAN, J., GOYE, A., ZELENETZ, A. D., SELLERI, L. & ET AL. 1995. A novel L23-related gene 40 kb downstream of the imprinted H19 gene is biallelically expressed in mid-fetal and adult human tissues. *Hum Mol Genet,* 4**,** 1499-507.

TSIEN, R. Y. 1998. The green fluorescent protein. *Annu Rev Biochem,* 67**,** 509-44.

TURSUN, B., COCHELLA, L., CARRERA, I. & HOBERT, O. 2009. A toolkit and robust pipeline for the generation of fosmid-based reporter genes in *C. elegans*. *PLoS One,* 4**,** e4625.

VELCULESCU, V. E., ZHANG, L., VOGELSTEIN, B. & KINZLER, K. W. 1995. Serial analysis of gene expression. *Science,* 270**,** 484-7.

VILLALOBOS, A., NESS, J. E., GUSTAFSSON, C., MINSHULL, J. & GOVINDARAJAN, S. 2006. Gene Designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics,* 7**,** 285.

WACHTER, R. M., ELSLIGER, M. A., KALLIO, K., HANSON, G. T. & REMINGTON, S. J. 1998. Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure,* 6**,** 1267-77.

WAKIYAMA, M., TAKIMOTO, K., OHARA, O. & YOKOYAMA, S. 2007. Let-7 microRNA-mediated mRNA deadenylation and translational repression in a mammalian cell-free system. *Genes Dev,* 21**,** 1857 - 1862.

WATT, V. M., INGLES, C. J., URDEA, M. S., AND RUTTER, W. J. 1985. Homology requirements for recombination in *Escherichia coli. Proc. Natl. Acad. Sci. USA* 82, 4768–4772.

WANG, B., LOVE, T., CALL, M., DOENCH, J. & NOVINA, C. 2006. Recapitulation of short RNA-directed translational gene silencing in vitro. *Mol Cell,* 22**,** 553 - 560.

WANG, B., YANEZ, A. & NOVINA, C. 2008. MicroRNA-repressed mRNAs contain 40S but not 60S components. *Proc Natl Acad Sci USA,* 105**,** 5343 - 5348.

WANG, L., JACKSON, W. C., STEINBACH, P. A. & TSIEN, R. Y. 2004. Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc Natl Acad Sci U S A,* 101**,** 16745-9.

WELCH, M., VILLALOBOS, A., GUSTAFSSON, C. & MINSHULL, J. 2011. Designing genes for successful protein expression. *Methods Enzymol,* 498**,** 43-66.

WESTENBERG, M., BAMPS, S., SOEDLING, H., HOPE, I. A. & DOLPHIN, C. T. 2010. *Escherichia coli* MW005: lambda Red-mediated recombineering and copy-number induction of oriV-equipped constructs in a single host. *BMC Biotechnol,* 10**,** 27.

WILLIAMS, C., XU, L. & BLUMENTHAL, T. 1999. SL1 trans splicing and 3'-end formation in a novel class of *Caenorhabditis elegans* operon. *Mol Cell Biol,* 19**,** 376-83.

WU, L. & BELASCO, J. 2008. Let me count the ways: mechanisms of gene regulation by miRNAs and siRNAs. *Mol Cell,* 29**,** 1 - 7.

WU, L., FAN, J. & BELASCO, J. 2006. MicroRNAs direct rapid deadenylation of mRNA. *Proc Natl Acad Sci USA,* 103**,** 4034 - 4039.

XU, X., SONG, Y., LI, Y., CHANG, J., ZHANG, H. & AN, L. 2010. The tandem affinity purification method: an efficient system for protein complex purification and protein interaction identification. *Protein Expr Purif,* 72**,** 149-56.

YOOK, K., HARRIS, T. W., BIERI, T., CABUNOC, A., CHAN, J., CHEN, W. J., DAVIS, P., DE LA CRUZ, N., DUONG, A., FANG, R., GANESAN, U., GROVE, C., HOWE, K., KADAM, S., KISHORE, R., LEE, R., LI, Y., MULLER, H. M., NAKAMURA, C., NASH, B., OZERSKY, P., PAULINI, M., RACITI, D., RANGARAJAN, A., SCHINDELMAN, G., SHI, X., SCHWARZ, E. M., ANN TULI, M., VAN AUKEN, K., WANG, D., WANG, X., WILLIAMS, G., HODGKIN, J., BERRIMAN, M., DURBIN, R., KERSEY, P., SPIETH, J., STEIN, L. & STERNBERG, P. W. 2012. WormBase 2012: more genomes, more data, new website. *Nucleic Acids Res,* 40**,** D735-41.

ZACHARIAS, D. A., VIOLIN, J. D., NEWTON, A. C. & TSIEN, R. Y. 2002. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science,* 296**,** 913-6.

ZAHLER, A. M. 2005. Alternative splicing in *C. elegans*. *WormBook* (http://www.wormbook.org/)**,** 1-13.

ZEKRI, L., HUNTZINGER, E., HEIMSTADT, S. & IZAURRALDE, E. 2009. The silencing domain of GW182 interacts with PABPC1 to promote translational repression and degradation of microRNA targets and is required for target release. *Mol Cell Biol,* 29**,** 6220 - 6231.

ZHANG, Y., KASHYAP, L., FERGUSON, A. A. & FISHER, A. L. 2011. The production of *C. elegans* transgenes via recombineering with the galK selectable marker. *J Vis Exp*.

ZHANG, Y., NASH, L. & FISHER, A. L. 2008. A simplified, robust, and streamlined procedure for the production of *C. elegans* transgenes via recombineering. *BMC Dev Biol,* 8**,** 119.

ZIMMERMANN, T. 2005. Spectral Imaging and Linear Unmixing in Light Microscopy

Microscopy Techniques. *In:* RIETDORF, J. (ed.). Springer Berlin / Heidelberg.

ZIMMERMANN, T., RIETDORF, J. & PEPPERKOK, R. 2003. Spectral imaging and its applications in live cell microscopy. *FEBS Letters,* 546**,** 87-92.

ZORIO, D. A., CHENG, N. N., BLUMENTHAL, T. & SPIETH, J. 1994. Operons as a common form of chromosomal organization in *C. elegans*. *Nature,* 372**,** 270-2.

Appendix 1 - Publications

Hirani et al. BMC Biotechnology 2013, 13:1
http://www.biomedcentral.com/1472-6750/13/1

BMC Biotechnology

METHODOLOGY ARTICLE Open Access

A simplified counter-selection recombineering protocol for creating fluorescent protein reporter constructs directly from *C. elegans* fosmid genomic clones

Nisha Hirani[1], Marcel Westenberg[1,4], Minaxi S Gami[1], Paul Davis[2], Ian A Hope[3] and Colin T Dolphin[1*]

Abstract

Background: Recombineering is a genetic engineering tool that enables facile modification of large episomal clones, e.g. BACs, fosmids. We have previously adapted this technology to generate, directly from fosmid-based genomic clones, fusion gene reporter constructs designed to investigate gene expression patterns in *C. elegans*. In our adaptation a *rpsL-tet(A)* positive/negative-selection cassette (RT-cassette) is first inserted and then, under negative selection, seamlessly replaced with the desired sequence. We report here on the generation and application of a resource comprising two sets of constructs designed to facilitate this particular recombineering approach.

Results: Two complementary sets of constructs were generated. The first contains different fluorescent protein reporter coding sequences and derivatives while the second set of constructs, based in the copy-number inducible vector pCC1Fos, provide a resource designed to simplify RT-cassette-based recombineering. These latter constructs are used in pairs the first member of which provides a template for PCR-amplification of an RT-cassette while the second provides, as an excised restriction fragment, the desired fluorescent protein reporter sequence. As the RT-cassette is flanked by approximately 200 bp from the ends of the reporter sequence the subsequent negative selection replacement step is highly efficient. Furthermore, use of a restriction fragment minimizes artefacts negating the need for final clone sequencing. Utilizing this resource we generated single-, double- and triple-tagged fosmid-based reporters to investigate expression patterns of three *C. elegans* genes located on a single genomic clone.

Conclusions: We describe the generation and application of a resource designed to facilitate counter-selection recombineering of fosmid-based *C. elegans* genomic clones. By choosing the appropriate pair of 'insertion' and 'replacement' constructs recombineered products, devoid of artefacts, are generated at high efficiency. Gene expression patterns for three genes located on the same genomic clone were investigated via a set of fosmid-based reporter constructs generated with the modified protocol.

Keywords: *C. elegans*, Recombineering, Fosmid, Fluorescent protein, Deoxyribose-phosphate aldolase, Peroxiredoxin, Metallocarboxypeptidase

* Correspondence: colin.dolphin@kcl.ac.uk
[1]Institute of Pharmaceutical Science, King's College London, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NH, UK
Full list of author information is available at the end of the article

Background

Since its introduction as a model metazoan animal [1] the experimental convenience, simple anatomy, which includes most differentiated tissues including a nervous system, and, in particular, genetic tractability has made the nematode *Caenorhabditis elegans* a popular research platform for many biologists and geneticists [2]. Sequence analysis of the *C. elegans* genome, completed in 1998 [3], has revealed the presence of approximately twenty thousand protein-coding genes. Although many of these have human homologues many are yet to have a function assigned. A number of complementary methodological approaches are available to functionally analyze *C. elegans* genes amongst which the determination of the spatial-temporal pattern of gene expression is particularly informative. While such patterns can be investigated at the level of transcription or translation, *via* approaches such as, respectively, *in situ* hybridization or immunohistochemistry, it is more common to employ reporter technology and analyze, in transgenic animals, the expression pattern of the surrogate reporter.

In its simplest guise, reporter technology involves cloning either the known promoter region or, if this sequence has not been precisely defined, as is often the case, then all or part of the 5′ intergenic sequence upstream of a reporter gene, such as *gfp*, in an appropriate base plasmid (e.g. reference [4]) to generate a so-called transcriptional reporter. Because the cloned sequence may not contain all necessary regulatory elements resulting expression patterns should be interpreted cautiously as they may not fully reflect that of the gene under study. If, on the other hand, a more complex translational reporter, in which the reporter is fused in-frame to the protein-coding sequence of the gene of interest, is constructed then, in addition to deriving information about promoter activity, potentially other related aspects of gene expression, such as the sub-cellular localization of the corresponding gene product, can also be derived. In order to include distantly located regulatory elements, and thus generate a construct more likely to recapitulate the expression pattern of the endogenous gene, such a reporter would, ideally, contain not only the genomic locus under study but also significant stretches of the 5′ and 3′ flanking sequence. Because *C. elegans* genes are relatively compact (5 kb on average [3]) and the fosmid-based genomic clone insert is typically 35–40 kb, a translational reporter based upon such a fosmid would, when the locus under study is located within the central region of the insert, provide that broad genomic DNA environment. Fosmid-based *C. elegans* genomic clones are thus ideal foundations for translational reporter construction [5].

To facilitate the generation of such reporter constructs from *C. elegans* fosmid clones we [5,6], and others [7-9], have developed tools and techniques designed to leverage the power of recombineering. Recombineering is a homologous recombination (HR)-based genetic engineering system mediated by transient expression of λ-encoded recombinases within an *E. coli* host (reviewed in [10]). We reported previously [5] on the use of a *rpsL-tet(A)* positive/negative-selection cassette (RT-cassette) [11] in a two-step counter-selection recombineering approach that enables the seamless creation of fosmid-based reporters devoid of undesired, extraneous sequence changes. Although some groups [8,9] have reported poor recombineering efficiencies using the RT-cassette as a selectable marker others [12,13] have, like us, employed it successfully to generate highly informative reporter fusions.

Although the various strategies and associated resources provide *C. elegans* researchers with a range of powerful recombineering tools the methodology is perhaps not as commonly practiced as it might be. To both further encourage its uptake and to facilitate an ongoing project designed to interrogate putative internal promoter elements within operons we generated two sets of constructs designed to provide tools to simplify the RT-cassette-based counter-selection recombineering approach employed in our laboratories. The first set contained coding sequences (CDSs) encoding potentially useful fluorescent protein (FP) reporters intended to expand the FP palette currently in common use by *C. elegans* researchers. In order to maximise the utility of the final resource, *N*- and *C*-terminal TAP- (tandem affinity purification) tag [14,15] sequences were also included. The FP-encoding sequences were both extracted from the previously described vectors pPD136.61, pPD136.64 [4] and pAA64 [16] that encode, respectively, CFP, YFP and the red-shifted mCherry [17], and synthesized *de novo* to encode the less frequently used mCerulean [18], mTFP1 (Teal FP) [19] and mCitrine [20]. A set of these inserts was subsequently introduced into the single-copy fosmid vector pCC1Fos generating the second set of constructs that provide the tools to faciliate RT-cassette-based recombineering. The resulting method requires only a single PCR, to generate the initial RT-cassette, and generates final products of high fidelity. We report here on the generation of these constructs and illustrate their utility by building reporter fusion constructs designed to examine the expression patterns of three *C. elegans* genes. These genes were all located on a single fosmid and a series of reporter constructs were built, *via* iterative rounds of counter-selection recombineering, in which one, two or all three genes were tagged with a different FP reporter. Fluorescence microscopy of lines transgenic for a subset of these constructs revealed expression patterns that either confirmed those previously reported or, in the case of one gene, indicated an unexpected aggregation of the fusion gene product.

Unexpectedly, and somewhat disappointingly, mCerulean, mCitrine and mCherry, each encoded by a codon-optimized CDS, exhibited either rather fast bleaching or were not as bright as had been envisaged particularly when expressed as fusion proteins. As these characteristics make their corresponding constructs less useful than hoped only those constructs that provided useful recombineering tools are being made available *via* the plasmid depository Addgene. By selecting constructs from this resource appropriate to their aims *C. elegans* researchers will be able to recombineer fosmid-based reporter constructs in a simplified, streamlined manner.

Results and discussion

Approximately 15% of all *C. elegans* genes are clustered together into polycistronic transcriptional units termed operons [21]. We have begun to investigate the extent that downstream operon genes may be transcriptionally regulated *via* internal *cis*-acting sequence elements (Hirani et al., unpublished). Operons containing such internal promoter elements, so-called hybrid operons, have been identified using transcriptional reporters designed to interrogate, in isolation, the intergenic regions between neighbouring operon genes [22]. In contrast, we are employing fosmid-based reporters to enable us to both manipulate precisely such putative internal regulatory elements and, at the same time, tag multiple operon genes with different FPs, without, in either case, otherwise disturbing the gross operon structure. To facilitate the recombineering of these reporter constructs we designed and built a resource consisting of two complementary sets of plasmids. The first of these comprises a modular set of FP CDSs including versions augmented by addition of *N*- and *C*-terminal TAP-tags or a nuclear-localization signal. The rationale for this set of plasmids was our initial aim to tag, with different FP reporters, up to four genes within any one operon. Spectral discrimination between each FP would be achieved by a combination of careful choice of FP with associated filter set and combining this, if necessary, with post-acquisition spectral unmixing. However, as discussed below, ultimately only three different FPs were combined within a single fosmid-based reporter enabling us to discriminate readily between their respective spectral signals simply using appropriate filter sets. The second set of plasmids, built using inserts from the first, was designed to both simplify and speed up the recombineering procedure and generate high fidelity FP fusion gene reporter products.

Sequences encoding fluorescent protein reporters

To identify FPs with sufficient spectral separation to allow multiple tagging and, at the same time, minimize the requirement for computational separation of the resulting fluorescent signals we considered the relative biophysical and spectral properties of the FPs available [23,24]. Monomeric FPs with reported high photostability and brightness were selected from the cyan, green, yellow and red regions of the visible light spectrum and led to the choice of the cyan mCerulean [18], the cyan/green mTFP1 [19], the yellow mCitrine [20] and the red mCherry [17]. Although DNA sequences encoding each of the FPs were available we decided to design, and have commercially synthesized, novel CDSs as this would not only permit codon optimization but also, by careful choice and placement of restriction enzyme (RE) sites, enable us to simplify all envisaged downstream sub-cloning events. In addition, because introns have been demonstrated to generally increase levels of heterologous gene expression [25], and are routinely inserted into sequences encoding reporter proteins in *C. elegans*, we also designed and included silent, blunt-cutting internal RE sites to enable facile insertion of up to two artificial introns per CDS (Additional file 1: Figure S1 and Additional file 1: Table S1).

We expanded the overall utility of this first set of constructs by including additional sequences encoding *N*- and *C*-terminal TAP-tags (pHN001, pNH002) to enable combined gene expression analysis and TAP-based protein purification [14,15]. We chose, as the terminal epitope, the S-Tag sequence due to its strong retention on S-protein resins [26] followed by, as the protease site, the recognition sequence for human rhinovirus 3C protease. For the internal tag we decided upon StrepTag II [27] because of its highly specific interaction with Strep-Tactin resins and the ability to subsequently elute bound protein(s) under gentle conditions. Although not yet experimentally tested these TAP-tags can be excised from constructs available from Addgene (Table 1). The modular design of the synthesized sequences and their provision in a vector, pGOv5, stripped of common restriction sites enabled us to build quickly a range of sub-clones encompassing the more useful potential module combinations (Additional file 1: Table S1).

Table 1 pGOv5-based constructs[a]

Construct	Insert	Intron 1 [e]	Intron 2
pNH001	$[(G_4S)_3]$::C-TAP-tag::2xNLS [b]	-	-
pNH002	N-TAP-tag::$[(G_4S)_3]$::mTFP1 [c]	-	-
pNH009	N-TAP-tag::mTFP1	-	-
pNH013	mTFP1 [d]	-	-
pNH026	mTFP1::$[(G_4S)_3]$::C-TAP-tag	-	-
pNH030	mTFP1::$[(G_4S)_3]$::2xNLS	-	-
pNH078	mTFP1[1I] [d]	B	-
pNH082	mTFP1[2I] [d]	B	C

[a] All constructs are in *E. coli* DH5α and are available from Addgene [b] lacks intact open-reading frame; [c] mTFP1 sequence lacks N-terminal Met-Val and C-terminal Lys residues present in native sequence; [d] N-terminal Met-Val residues and C-terminal Lys residue, present in native coding sequence, replaced, respectively, with Met-Ala-Ala and Val-Ser-Ala; [e] see Additional file 1: Table S2.

Prior to their incorporation into fosmid-based reporters the spectral and biophysical properties of the FPs were briefly investigated *via in vitro* and *in vivo* approaches. First, each FP was expressed *in vitro* from a DNA template and the resulting protein immunocaptured to agarose beads *via* an appropriate anti-FP antibody. Fluorescence microscopy revealed discrete, concentrated fluorescence of the intended colour (Additional file 1: Figure S5) confirming that each FP had been successfully generated from its respective CDS and each exhibited the expected spectral profile. Second, we observed fluorescence expression patterns in transgenic *C. elegans* strains transformed with constructs built from the *myo*-3^{PROM}-containing vector pPD95.86 [4] that encoded either an unembellished FP, with a contiguous or two intron-containing (2I) CDS, or with a C-terminal nuclear localization signal (NLS). Control strains were also generated by transformation with DNA of equivalent pPD95.86-based sub-clones that contained CDSs encoding CFP, YFP and mCherry excised, respectively, from pPD136.61, pPD136.64 [4] and pAA64 [16]. To distinguish these FPs from those encoded by CDSs designed as part of this work the prior versions of CFP and YFP are prefixed with F (F-CFP, F-YFP) and mCherry with Mc (Mc-mCherry). All strains revealed the expected body muscle expression pattern (Additional file 1: Figure S5) with clear nuclear localization for the FPs equipped with a NLS (data not shown). Although these results indicated that the FPs possessed the anticipated spectral properties, both *in vitro* and *in vivo*, two additional observations were made. First, rapid photobleaching was exhibited by mCerulean, particularly when expressed *in vivo* within transgenic worms, using the same illumination intensity and exposure time parameters employed to visualize the other FPs including F-CFP. As the FP had effectively completely bleached in the time it took to focus on the fine detail of a specimen it was concluded that mCerulean would be impractical as a FP partner in multiple *in vivo* gene tagging strategies and that the three FP combination mTFP1/mCitrine/mCherry would be used in future constructs. The reasons for the rapid *in vivo* fluorescence decay of mCerulean, especially in comparison to that observed with the ancestral enhanced CFP (ECFP) (visualized here as F-CFP), is not clear as the amino acid sequence is the same as that of the original mCerulean [18]. Interestingly, recent crystallographic data [28] has suggested that the additional Y145A change present in mCerulean might exacerbate the likelihood of fluorescence quenching in comparison to CFP. Whatever the underlying cause(s), further improved versions of cyan-shifted FPs have been developed, e.g. mCerulean3 [29] and mTurquoise2 [28], which would be less likely to suffer from such fluorescence instability.

The second observation was the apparent lack of improvement, in terms of relative brightness, exhibited by mCitrine and mCherry when compared, respectively, to F-YFP and Mc-mCherry despite the codon optimization undertaken specifically to maximize expression *in vivo*. Both latter FPs are encoded by CDSs containing three short, equally spaced artificial introns. As expected, the introduction of two similar sized artificial introns, albeit not spaced equally, into the CDSs encoding mCitrine and mCherry improved, as judged by eye, protein expression levels in transgenic worms over the equivalent non-intron-containing CDSs (data not shown). However, longer exposure times were still required to achieve fluorescence emissions equivalent to worms transformed with corresponding F-YFP- and Mc-mCherry-encoding constructs. In contrast, mTFP1 proved to be consistently bright and photostable requiring only rather short (2–5 msec) exposure times when encoded by the two intron-containing CDS. Many factors are involved in determining the overall brightness of the FP reporter including the efficiency of transcription/translation of the encoding sequence as well as the intrinsic spectral properties of the resulting fluorophore. With respect to the latter, we expected mCitrine, especially when encoded by the two-intron-containing CDS, to be at least as bright as F-YFP, when illuminated under equivalent exposure conditions, as this FP is considered superior to standard YFP in both brightness and stability [23,24]. Although strains transformed with *myo*-3^{PROM}-driven Mc-mCherry reproducibly exhibited brighter images than equivalent strains transformed with mCherry the two FPs had identical amino acid sequences and differed only by being encoded by different, "codon-optimized" CDSs that were globally 77% identical (data not shown). The differences in brightness observed between the respective yellow and red FP pairs are more likely due to differences in the efficiency with which each CDS is transcribed and/or the resulting mRNA translated into protein resulting in reduced FP abundance than any significant differences between the spectral properties of the fluors. Although we attempted to ensure, by designing sequences from scratch, that the CDSs encoding mCitrine and mCherry would promote efficient transcription and translation it is clear that simple assumptions concerning, for example, adjusting codon choice to reflect usage in highly expressed endogenous genes, does not necessarily guarantee optimized protein expression [30].

pCC1Fos-based constructs

The second set of plasmids, built with inserts derived from the first, was designed to enable the recombineering procedure to be simplified, take less time and generate final FP fusion gene reporter constructs of high fidelity. As a subset of the plasmids were to contain the

RT-cassette which, we had found, confers instability when present in a standard high-copy number vector (unpublished observations), it was decided to base them all in pCC1Fos. As well as maintaining clones as single copies for stable, routine propagation pCC1Fos-based constructs can, when hosted in a suitable strain such as EPI300 or MW005 [6], also be transiently induced to approximately 50 copies per bacterial cell thus improving DNA isolation yield and purity. These plasmids were designed as pairs the first member of which provides the template for PCR-amplification of the RT-cassette, used in the first, positive selection recombineering step, while the second provides the desired replacement sequence, excised as a restriction fragment, for use in the subsequent counter selection step. Generation of this pCC1Fos-based construct set greatly simplifies the RT-cassette-mediated counter-selection recombineering protocol as only one PCR needs to be performed and a restriction fragment is used to replace the inserted RT-cassette in the second recombineering step. The first step involves PCR-amplification of an RT-cassette from a template sequence excised by restriction enzyme digestion from the appropriate construct. The product is flanked by approximately 200 bp, derived from the template, matching the 5′ and 3′ ends of the chosen FP CDS and 50 bp terminal homology arms, derived from the primers, designed to direct insertion into the chosen site within the target gene. Replacement of the inserted RT-cassette is achieved by recombineeering transformation with a *Not*I restriction fragment, excised from a second plasmid, containing the desired FP CDS. The inclusion of the extended regions of flanking FP CDS homology with the RT-cassette ensures this negative selection replacement step is highly efficient with far fewer false positives, e.g. from when the RT-cassette is deleted rather than being replaced. In our hands, for clones selected under negative selection with chloramphenicol and streptomycin, successful replacement of the RT-cassette with the desired sequence is essentially 100% (data not shown). This improvement in HR brings a number of additional methodological benefits including smaller amounts of linear replacement DNA (50 ng) being required and, as far fewer bacterial cells need be to electroporated, reduced cell culture volumes and shorter incubation times. Furthermore, use of a restriction fragment ensures final recombineered products are essentially free of sequence artefacts introduced by PCR requiring only RE analysis to confirm construction fidelity. Lists of paired constructs designed to provide template for PCR-amplification of the RT-cassette donor and the subsequent replacement fragment are provided in Table 2 and Additional file 1: Table S6. Additional methodological details are provided in the supplemental data.

Gene expression analysis

Having generated this pCC1Fos-based resource set we utilized some of its components to build, by iterative rounds of counter-selection recombineering, a series of translational reporter constructs from a single fosmid clone. Although we are using this resource to investigate the transcriptional complexity of operon gene regulation we chose, for the sake of simplicity, to confirm its utility here by tagging genes present on the same fosmid clone but not located within a single operon. A list of fosmid clones each containing three or more such genes with

Table 2 Resources for simplified counter-selection recombineering

Desired insertion sequence[a]	RT	kb[d]	RT-cassette-containing construct[b]		Replacement construct[c]		
			Fwd (5′-3′)[e]	Rev (5′-3′)[e]	pCC1Fos-based	pGOv5-based	kb[f]
RT-cassette	pNH034	2.0	GCTGTCGAGATATGACGGTGTTCA	TCTTGGAGTGGTGAATCCGTTAGC	-	-	
F-CFP	pNH050	2.4	ATGAGTAAAGGAGAAGAACTTTTC	[*]TTTGTATAGTTCATCCATGCCATG	pNH039	n/a	0.9
F-GFP	pNH051	2.4	ATGAGTAAAGGAGAAGAACTTTTC	[*]TTTGTATAGTTCATCCATGCCATG	pNH040	n/a	0.9
F-YFP	pNH052	2.4	ATGAGTAAAGGAGAAGAACTTTTC	[*]TTTGTATAGTTCATCCATGCCATG	pNH041	n/a	0.9
Mc-mCherry	pNH053	2.4	ATGGTCTCAAAGGGTGAAGAAGAT	[*]GGATCCACTAGTCTTATACAATTC	pNH042	n/a	0.9
mTFP1	pNH054	2.4	ATGGCCGCCTCAAAAGGAGAAGAA	[*]AGCGCTTACGTAGAGCTCGTCCAT	pNH043	pNH013	0.7
N-TAP-tag::[$(G_4S)_3$]::mTFP1	pNH066	2.6	ATGGTTAAAGAAACAGCAGCAGCG	[*]AGCGCTTACGTAGAGCTCGTCCAT	pNH058	pNH002	0.9
mTFP1::[$(G_4S)_3$]::C-TAP-tag	pNH094	2.3	ATGGCCGCCTCAAAAGGAGAAGAA	[*]AGCCCATGAGTCCATATGCTGTCT	pNH090	pNH026	0.9
mTFP1::[$(G_4S)_3$]::2xNLS	pNH070	2.6	ATGGCCGCCTCAAAAGGAGAAGAA	[*]AGCGCTAACTTTTCGCTTCTTCTT	n/a	pNH030	0.8
mTFP1[B]	pNH054	2.4	ATGGCCGCCTCAAAAGGAGAAGAA	[*]AGCGCTTACGTAGAGCTCGTCCAT	n/a	pNH078	0.8
mTFP1[BC]	pNH054	2.4	ATGGCCGCCTCAAAAGGAGAAGAA	[*]AGCGCTTACGTAGAGCTCGTCCAT	n/a	pNH082	0.8

[a] sequence to be seamlessly inserted into target; [b] the counter-selection RT-cassette should be PCR-amplified from this construct using, as template, the RT-cassette-containing *Not*I-*Not*I fragment; [c] the desired replacement cassette can be excised, as a *Not*I-*Not*I fragment, from either the pCC1Fos- (if available) or parental pGOv5-based (Table 1) constructs; [d] size of the isolated *Not*I-fragment (or *Nco*I-*Nco*I fragment for pNH034) to be used as PCR template; [e] suggested 3′ sequences of 'recombineering' ODNs designed to PCR-amplify FP-flanked, RT-cassette-containing sequence. Reverse ODN sequence lacks a 5′ stop triplet [*]; [f] size of the isolated *Not*I-fragment to be used to replace the inserted FP-flanked, RT-cassette-containing sequence. All constructs are available from Addgene and are in *E. coli* EPI300 apart from those built in pGOv5 which are in *E. coli* DH5α.

published expression patterns was scanned by eye and clone WRM069dD11, containing three such genes located approximately centrally within the insert, was identified. WRM069dD11 contains the genes *F09E5.3*, *F09E5.15* (*prdx-2*) and *EEED8.6* (*ccpp-6*) that encode, respectively, a putative deoxyribose-phosphate aldolase (DERA), a peroxiredoxin and a metallocarboxypeptidase (Figure 1). The iterative recombineering workflow generated single, double and triple-tagged constructs for both the mTFP1/mCitrine/mCherry and F-CFP/F-YFP/Mc-mCherry FP combinations (Additional file 1: Figure S4 and Figure 2). Transgenic lines were generated for the single- and triple-tagged variants and resulting expression patterns investigated (Figure 3). Microscopic examination of the lines revealed that, as noted previously for the *myo-3*^{PROM}-driven transcriptional constructs, the mCitrine and mCherry fusion proteins, even when encoded by CDSs with 2 introns, were, as determined by eye, visibly less bright that the respective F-YFP and Mc-mCherry equivalents. Whatever the cause(s) the low signal levels made discerning clear expression patterns for the two fusion genes, *F09E5.15::mCitrine(2I)* and *EEED8.6::mCherry (2I)*, difficult and further examination of the corresponding lines was not pursued. Interestingly, mTFP1-expressing worms were, consistently, as bright as the otherwise equivalent F-CFP-expressing strains confirming that mTFP1, as encoded by the CDS designed here, is a useful addition to the FP palette for reporter analysis in *C. elegans* (Figure 3 panels E, F). All lines transformed with a reporter construct built with one or more of the F-CFP/F-YFP/Mc-mCherry FP set exhibited relatively bright and easily interpretable expression patterns (Figure 3).

Subsequent investigation of expression patterns proceeded by examination of lines transformed with the F-CFP/F-YFP/Mc-mCherry-tagged constructs. *F09E5.3* encodes a DERA homolog that shares 72% similarity and 54% identity with human DERA (data not shown). DERA, as part of the pentose phosphate shunt, catalyzes the reversible reaction of 2-deoxy-D-ribose 5-phosphate to D-glyceraldehyde 3-phosphate and acetaldehyde [31] and, as such, might be expected to exhibit widespread expression consistent with an enzyme involved in intermediate metabolism. Indeed, examination of strains CTD1059 and CTD1055 (Additional file 1: Table S7), that were transformed, respectively, with the *F09E5.3::F-CFP* (fNH058) and the triple gene-tagged construct (fNH086), revealed diffuse F-CFP expression throughout the hypodermis, intestine and pharynx (Figure 3, panels A-D, R). Similarly, strain CTD1050 (Additional file 1: Table S7), transformed with the corresponding *F09E5.3::mTFP1* construct fNH068, exhibited an essentially equivalent expression pattern (Figure 3, panels E, F). This generalized pattern of expression was in agreement with that previously published for strain BC14910 that had been generated by transformation with a corresponding *F09E5.3*^{PROM} transcriptional style reporter construct (Figure 1) [32,33].

F09E5.15 (*prdx-2*) encodes one of the two 2-Cys peroxiredoxins expressed in *C. elegans*. These conserved thioredoxin-coupled peroxidases, *via* their capacity for H_2O_2 reduction, are important components of the overall oxidative-stress response employed by multicellular organisms. However, there is also increasing evidence that they have additional, complex functions including important roles as regulators of H_2O_2-mediated redox

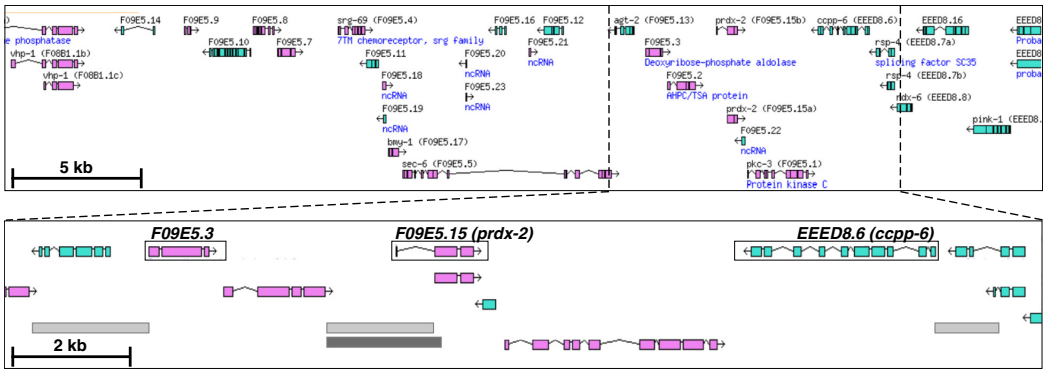

Figure 1 Genomic location of fosmid WRM069dD11 insert. Upper panel: genes on the forward (purple exons) and reverse (green exons) strands of the insert from fosmid genomic clone WRM069dD11. Lower panel: An expanded region of the insert demonstrating the relative locations of the genes (boxed) *F09E5.3*, *F09E5.15* (*prdx-2*) and *EEED8.6* (*ccpp-6*) for which published, independent expression data was available. Locations of the genomic regions used to drive *gfp* expression in the transcriptional reporters described by McKay *et al.* (2003) ref [32] and Hunt-Newbury *et al.* (2007) ref [33] (light grey boxes) for *F09E5.3*, *F09E5.15* and *EEED8.6*, strains BC14910, BC13145 and BC11803, respectively, and Isermann *et al.* (2004) ref [35] (dark grey box) for *F09E5.15* are indicated (lower panel). Figure derived from WormBase.

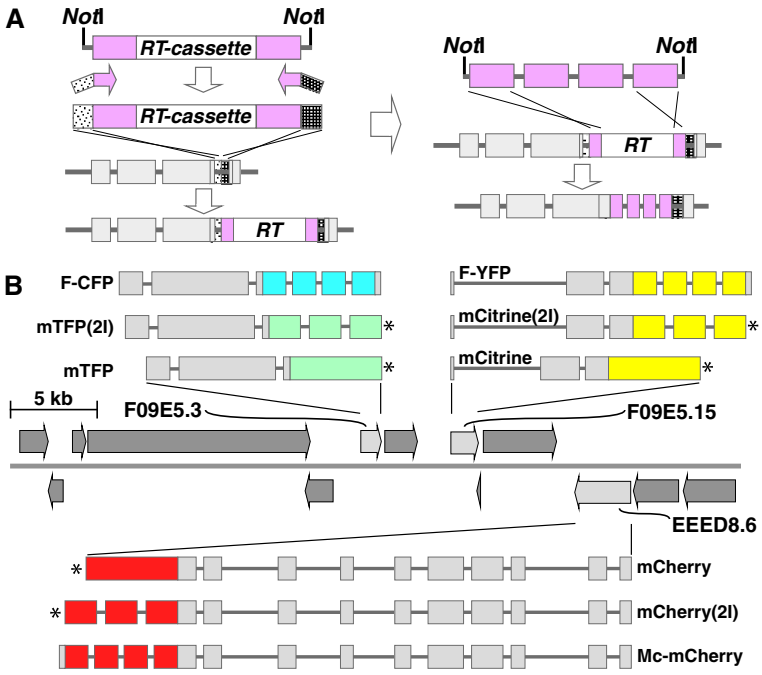

Figure 2 Counter-selection recombineering-mediated generation of fosmid-based fluorescent protein-fusion reporter constructs. Panel **A**: schematic representation of the modified counter-selection protocol. A *rpsL-tet*A(C) (RT) counter-selection cassette, PCR-amplified from a *Not*I fragment excised from the appropriate pCC1Fos-based construct, flanked by approx. 200 bp from the 5′ and 3′ ends of the fluorescent protein reporter to be inserted (purple boxes) and terminal 50 bp homology arms (stippled boxes), is recombineered into the insertion site within the target gene *via* tetracycline selection of the positive marker (*tet*A(C)). In the subsequent replacement step a *Not*I fragment, containing the fluorescent protein coding sequence, replaces the inserted RT-cassette, *via* streptomycin selection conferred due to loss of the negative marker (*rps*L), generating an in-frame *gene::fp* fusion. Panel **B**: a schematic representation, within the centre of the panel and drawn approximately to scale, illustrating sizes and orientations of the genes located on the insert of the fosmid clone WRM069dD11. The genes *F09E5.3*, *F09E5.13* and *EEED8.6* (light grey solid arrows) were tagged, respectively, with one of either F-CFP, mTFP1(2I) or mTFP1, F-YFP, mCitrine(2I) or mCitrine, or Mc-mCherry, mCherry(2I) or mCherry. Reporter gene fusions marked with an asterisk lack the 5 terminal codons of the target gene.

signaling [34]. While the generalized anti-oxidant role might suggest a relatively widespread distribution for PRDX-2 in *C. elegans* the previously reported expression patterns indicated a more tissue-, even cell-specific localization suggesting that PRDX-2 may play one of these more discrete, specialized functional role(s). For example, using a *F09E5.15*^{PROM} transcriptional reporter (Figure 1), Isermann et al. (2004) ref. [35] reported that PRDX-2 expression was restricted to the pharyngeal interneurons I2 and I4. A more distributed PRDX-2 expression pattern, including head and tail neurons and intestine, was reported in strain BC13145 transformed with a similar transcriptional reporter (Figure 1) [32,33]. Olahova et al. (2008) ref. [36] demonstrated subsequently, *via* an immunohistochemical approach, that PRDX-2 was also expressed in gonad and intestine. However, examination of strain CTD1051 (Additional file 1: Table S7), transformed with the single *F09E5.15::F-YFP* construct (fNH058), demonstrated a highly unusual expression pattern that did not conform with either of these previous reports (Figure 3, panels G-N). F-YFP expression was observed as intense punctate foci distributed, apparently randomly, throughout the body and that did not co-localize with any specific tissue or cell type. Equivalent patterns were observed in other lines transformed with fNH058 as well as those generated from the triple-tagged construct (Figure 3, panel S). This pattern of concentrated fluorescent foci is difficult to interpret in the context of PRDX-2 function. It might reflect some uncharacterized role but, and possibly most likely, may also result from unexpected, artifactual aggregation of the fusion protein perhaps initiated and promoted by the presence of the *C*-terminal F-YFP sequence. Whilst this is entirely possible, one of us (IAH), with considerable experience in interpreting gene expression patterns in *C. elegans*, has not previously observed such aggregations suggesting that if they do occur

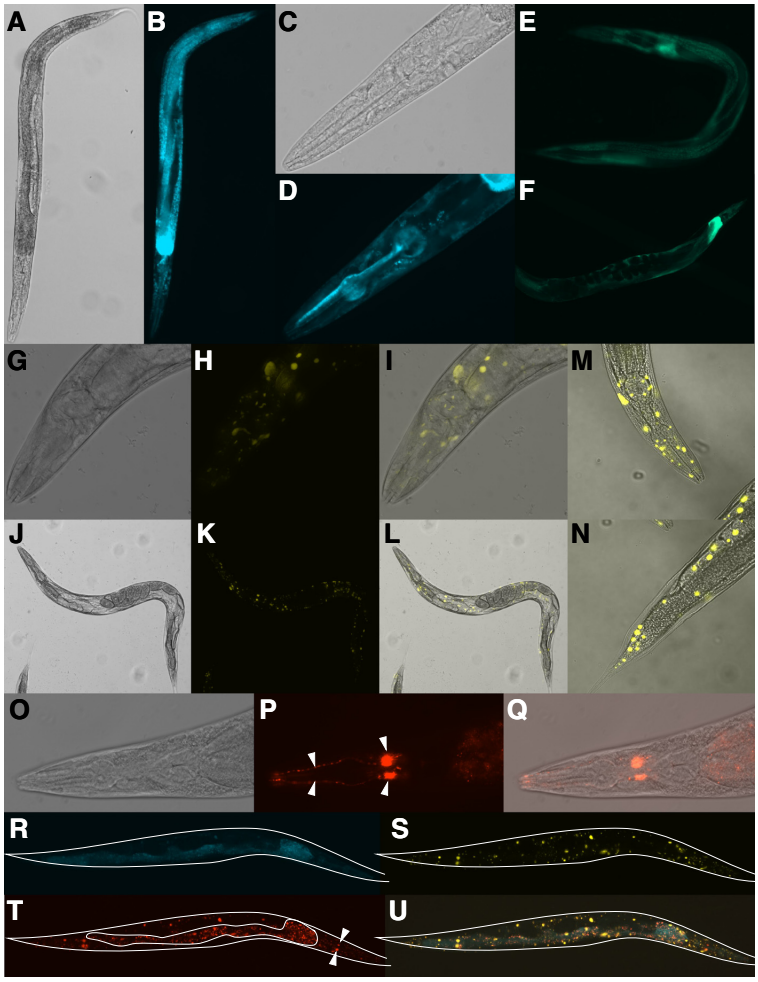

Figure 3 Expression patterns. Reporter expression patterns in adult hermaphrodites transgenic for fosmid-based constructs that tag each of the genes *F09E5.3* (**A**-**F**), *F09E5.15* (*prdx-2*) (**G**-**N**) and *EEED8.6* (*ccpp-6*) (**O**-**Q**) individually, or collectively (**R**-**U**). F-CFP (**B**, **D**, **R**), mTFP1 (**E**, **F**), F-YFP (**H**, **I**, **K**, **L**, **M**, **N**, **S**) and Mc-mCherry (**P**, **Q**, **T**) and merged F-CFP/F-YFP/Mc-mCherry (**U**) distributions are presented. DIC was used to observe anatomical details either alone (**A**, **C**, **G**, **J**, **O**) or after merging with the associated FP image(s) (**I**, **L**, **M**, **N**, **Q**). The *C. elegans* strains photographed were CTD1059 (**A**-**D**), CTD1050 (**E, F**), CTD1051 (**G**-**N**), and CTD1052 (**O**-**Q**) and CTD1055 (**R**-**U**) (Additional file 1: Table S7). Images were captured at 100x (**A**, **B**, **E**, **F**, **J**, **K**, **L**, **R**, **S**, **T**, **U**) or 400x (**C**, **D**, **G**, **H**, **I**, **O**, **P**, **Q**) magnification and fluorescence image capture times were 2 msec (**A**, **C**, **G**, **H**, **J**, **K**, **O**, **U**), 5 msec (**S**), 250 msec (**E**, **F**) or 1 sec (**B**, **D**, **P**, **R**, **T**). Confocal images (**M**, **N**) were captured at 400x magnification. Mc-mCherry expression in the anterior dendritic projections and cell bodies of one pair of the four head cephalic neurons are indicated (arrowheads) in panel P. The location of these same cell bodies is also indicated in panel T (arrowheads). In panels R-U the approximate outline of the imaged worm is provided and, in addition, in panel T a region of granular intestinal autofluorescence prominent following the 1 sec excitation of Mc-mCherry is also outlined.

with reporter gene fusions it is a rare event. We were subsequently made aware (E. Veal, pers. comm.) that very similar punctate expression was observed with a *prdx-2::gfp* translational reporter fusion but not when an N-terminal *gfp::prdx-2* construct was used. Taken together these results indicate that the punctate expression pattern, most likely artifactual aggregation, is due to the C-terminal extension and appears not to be FP-specific.

EEED8.6 (*ccpp-6*), the third gene to be tagged within fosmid WRM069dD11, and *ccpp-1* each encode one of two *C. elegans* cytosolic carboxypeptidases. CCPP-6 has been identified [37] as a deglutamylase that, in conjunction

with tubulin tyrosine ligase-like (TTLL) glutamylases, mediate the respective deglutamylation and glutamylation of tubulin thereby regulating the microtubular network present in the *C. elegans* neuronal sensory cilia. A *ccpp-1::gfp* transcriptional reporter revealed CCPP-1 expression in head neurons, specifically amphids, which were also demonstrated to exhibit TTLL-dependent tubulin glutamylation activity [37]. Interestingly, a *ccpp-6::gfp* transcriptional reporter suggested labial neuron expression for CCPP-6 whereas, in contrast, a translational version indicated expression in putative amphid cell bodies [37]. This discrepancy was attributed to the inclusion of a possible enhancer(s) in the latter construct. Examination of strain CTD1052 (Additional file 1: Table S7), transformed with the single *EEED8.6::Mc-mCherry* construct (fNH060), demonstrated clear, distinct reporter expression in head neuron cell bodies and anterior dendritic extensions. Careful examination of the relative location of the cell bodies (Figure 3, panel Q) identified these as likely to be the mechanosensitive CEP cephalic neurons. This interpretation was confirmed by additional confocal microscopy (data not shown). Localized CEP neuron expression was observed in other lines transformed with *EEED8.6::Mc-mCherry* (data not shown) and the triple-tagged construct (Figure 3, panel T). Lines transformed with the latter construct exhibited unrelated punctate fluorescence in the red channel due to bleed-through from excitation of the concentrated foci of *F09E5.15::F-YFP*. The clear CEP expression observed here for *EEED8.6* agreed with previously published expression data for strain BC11803 generated by transformation with a corresponding *EEED8.6*^{PROM} transcriptional style reporter (Figure 1) [32,33]. Thus, interestingly, it would appear that CCPP-6 is expressed predominantly in the cells of the cephalic sensilla, whereas CCPP-1, that lacks detectable deglutamylase activity but co-localizes with TTLL-mediated glutamylation [37] is, in contrast, expressed in the amphid neurons.

The use of recombineering to build, directly from a genomic clone, a translational-style *C. elegans* reporter fusion construct is gaining in popularity and a number of related methodologies have been described [5,7-9]. All these different approaches represent variations on a common methodological theme and, although all are welcomed as attempts to make the procedure easier and more robust, they will, as discussed [38], have relative strengths and weaknesses. As is often the case, the choice of which to use will depend upon application and personal preference. We have generated a construct resource that further simplifies the approach used by us enabling the rapid construction of fosmid-based reporter fusion constructs seamlessly tagged, if desired, with multiple FP reporters. From this resource we are making available, through the plasmid depository Addgene, those constructs required to generate fosmid reporters tagged with either mTFP1, F-CFP, F-YFP or Mc-mCherry as these will be of immediate use to others (Tables 1 and 2). Constructs containing sequences encoding mCerulean, mCitrine and mCherry, encoded by the codon-optimized sequences designed by us, are available from CTD.

Conclusions

We have described the generation and utility of construct collections designed to simplify and facilitate the building of fosmid-based FP reporter constructs *via* counter-selection recombineering. Use of the resource generates final recombineered products that are invariably free of unwanted sequence artefacts and, as such, can be used directly to create transgenic animals with the need for only minimal fidelity checking. Although some of the FPs encoded displayed undesirable characteristics, such as rapid bleaching and poor brightness, the remaining constructs represent a valuable resource that both simplifies the overall procedure and increases significantly recombineering efficiency.

Methods

C. elegans culture, strains, transformation and microscopy

C. elegans culture, handling and manipulations were performed according to standard methods. The wild-type Bristol N2 strain [1] was transformed by co-injection [39] of either plasmid (100 ng/μl) or fosmid (10 ng/μl) DNA together with marker plasmid pRF4 (100 ng/μl). pRF4 contains *rol-6(su1006)* which confers a rolling phenotype by which transformants can be recognized and transgenic lines maintained. Each transgenic strain was established from a different injected animal and therefore is independently generated. FP expression patterns were observed, in hermaphrodites only, by fluorescence microscopy on an Olympus BX61 equipped with DIC optics, filter sets designed to acquire GFP (Chroma, 41012), CFP, mCerulean and mTFP1 (Semrock, CFP-2432A), YFP and mCitrine (Semrock, YFP-2427A) and mCherry (Semrock, mCherry-A) and CellSens (v.1.4) software. Confocal images were captured with a Zeiss LSM510 META system and LSM510 (v.3.4) software.

General molecular and bioinformatic methods

Unless otherwise stated classical genetic engineering utilized standard protocols [40]. Restriction and modifying enzymes were from NEB (UK) and oligonucleotides (ODNs) from IDT (Belgium). ODNs for priming sequencing or standard PCR reactions were desalted whereas long (>70 nt) ODNs designed to generate linear recombineering amplicons were either PAGE-purified or ultramer grade. All PCRs were performed in volumes of 50 μl containing 15 pmol of each ODN primer and 200 mM of each dNTP and were catalyzed with a high-

fidelity DNA polymerase (Phusion, NEB) using conditions designed to minimize mis-incorporation. ODN design, cloning strategies and sequence alignments were performed using MacVector software (MacVector, Inc.). A construct derived directly from a clone originating from the *C. elegans* genomic library constructed in the copy number-inducible fosmid vector pCC1Fos (D. Moerman, pers. comm.) and which thus retains the fosmid backbone is prefixed with an "f" (fosmid). In contrast, non-genomic DNA insert-containing constructs, irrespective of the vector backbone type, are considered plasmids and prefixed "p" (plasmid). Linear RT-cassette-containing counter-selection markers for insertion were PCR-amplified essentially as described [5] using, as template, a restriction fragment excised from the appropriate construct (Additional file 1: Table S6) and an annealing temperature of 50°C.

Identification of fosmid genomic clones with associated gene expression data

Data from WormBase [41] release WS226 was interrogated to identify candidate pCC1Fos-based genomic clones containing a cluster of genes with associated published expression pattern data. Appropriate filters were applied to these data to remove any data point that did not have an associated reference. The remaining data were then converted into a list of genes by exploiting the high levels of connectivity in the ACeDB [42] schema developed by the WormBase project. The gene co-ordinate data were used to align with the fosmid clone genomic positions by performing a GFF (General Feature Format) union to generate Gene::Fosmid connections. These data were then clustered and sorted on fosmid ID to produce a candidate list of over 4000 clones that contain three or more genes each of which has a published expression pattern.

Two-step, counter-selection recombineering

All recombineering was undertaken in *E. coli* strain MW005 [6] that supports both λ Red- mediated recombineering and copy-number induction of pCC1Fos-based fosmid clones. Positive selection of recombinants *via* the RT-cassette employed tetracycline (Tc, 5 μg/ml) plus chloramphenicol (Cm, 10 μg/ml) to select for the pCC1Fos backbone. On occasion, and as indicated in the text, ampicillin (amp, 50 μg/ml) either replaced, or was used in combination with, Tc. Negative selection – used when selecting for the targeted replacement of the RT-cassette – was achieved with streptomycin (Sm, 500 μg/ml) and Cm (10 μg/ml). Recombineering, when employed at the start of the project as a tool to generate resource components, was performed essentially as described [5]. Subsequent access to these resources enabled the following simplified, more streamlined recombineering protocol to be developed requiring only a single PCR to generate the initial RT-cassette, smaller culture volumes and incubation times, and no formal requirement to sequence final recombinants. A volume (2–3 ml for each RT-cassette-containing PCR product to be introduced) of liquid medium (SOB (–Mg) plus Cm), was inoculated with an aliquot (1 in 10 v/v) from an overnight 'target fosmid clone'-containing MW005 mini-culture (150 r.p.m., 32°C), incubated on (approx. 2 h, 150 r.p.m., 32°C) to mid log phase (Ab600 approx. 0.6), aliquots (1 ml) transferred into each of an appropriate number of micro-tubes, centrifuged, supernatant removed, cells resuspended (1 ml) in pre-warmed (45°C) medium and the tube(s) transferred to a shaking incubator (Labnet VorTemp, 100 r.p.m., 45°C, 5 min). Each tube was removed, chilled on ice (10 min), harvested by brief centrifugation, cells washed (2 × 1 ml ice-cold ddH_2O), resuspended (50 μl, ddH_2O) and electroporated with an appropriate, purified (Promega, Wizard) RT-cassette-containing PCR product (500 ng). Recovery (3 h, 32°C) and subsequent positive selection of colonies harbouring RT-cassette-containing recombinant clones on LB-agar (Tc, Cm, 36-48 h, 32°C) was performed as described [5]. Subsequently, the RT-cassette was replaced, *via* negative-selection recombineering in MW005 cells prepared as described above, with the appropriate gel-purified *Not*I-*Not*I fragment (50 ng) excised from the correct 'replacement' construct (see below). Selection of desired recombinant-containing colonies on No Salt (NS)-LB agar (Sm, Cm, 36-48 h, 32°C), subsequent propagation in liquid culture, induction of copy-number ('CopyControl', Epicentre) and fosmid DNA isolation were performed as described [5]. Gross fidelity of final constructs was determined by RE analyses.

Coding sequence design and building pGOv5- and pPD95.86-based sub-clones

DNA sequences encoding the FPs mCerulean [18], mTFP1 [19], mCitrine [20] and mCherry [17], plus *N*- and *C*-TAP-tags, were codon-optimized for expression in *C. elegans* and, by careful placement of appropriate RE sites, designed to be modular in nature to facilitate subsequent sub-clone creation (detailed in supplemental data and Additional file 1: Figure S1). The sequences were commercially synthesized and provided in proprietary vectors. The lack of common RE sites in one of these, pGOv5, facilitated the subsequent generation in this plasmid vector of sets of sub-clones, one for each FP CDS, representing those module combinations likely to be most useful. These included *N*- and *C*-TAP-tagged FPs and FP CDSs containing one or two artificial introns (Additional file 1: Table S1). Following generation of these pGOv5-based sub-clones a number of the insert sequences, together with sequences encoding F-CFP, F-YFP and Mc-mCherry, were each transferred into the

backbone of the *myo*-3^{PROM}-containing vector pPD95.86 [4] (Additional file 1: Table S1).

Generating pCC1Fos-based constructs

Both classical genetic engineering and recombineering were employed to, firstly, introduce the RT-cassette into pCC1Fos and, subsequently, build a set of pCC1Fos-based constructs (Additional file 1: Figure S2) designed to simplify the counter-selection recombineering protocol. To retrofit pCC1Fos with the RT-cassette, two separate regions of the pGOv5 backbone were PCR-amplified from pNH002 (Additional file 1: Table S1) DNA (1 ng). One resulting amplicon, generated with ODNs 12015/12016 (Additional file 1: Table S3), comprised approx. 600 bp of the 3′ end of the *b*-lactamase (*bla*) coding sequence flanked, at one end, by a stretch (50 bp) identical to a region within pCC1Fos and, at the other, by 50 bp equivalent to one end of the RT-cassette (Additional file 1: Figure S2i). The second amplicon, generated using ODNs 12017/12018 (Additional file 1: Table S3), comprised approx. 200 bp equivalent to a section within pCC1Fos and a 50 bp terminus identical to the other end of the RT-cassette (Additional file 1: Figure S2i). These two amplicons were fused with the RT-cassette in a 3-way Overlap-PCR with ODNs 12015/12018 generating an RT-cassette product with 50 and 200 bp terminal sequences identical to regions in pCC1Fos (Additional file 1: Figure S2ii). The fused product was inserted, *via* positive-selection recombineering, into pCC1Fos, resulting in the replacement of the *LacZ-α* and *loxP* regions and introduction of the partial 3′ end of *bla* to generate the RT-cassette-containing pCC1Fos-based construct pNH034 (Additional file 1: Figure S2iii). Subsequently, FP cassettes, excised from the pGOv5-based constructs pNH002, pNH006-8, pNH0013-20 and pNH0026-29 (Additional file 1: Table S1) with *Fsp*I and *Lgu*I (Additional file 1: Figure S2iv), were each recombineered into pNH034 generating a series of constructs in which the RT-cassette had been replaced by a different *Not*I-flanked FP CDS or derivative (Additional file 1: Table S4). Because the full-length *bla* gene was reconstructed during RT-cassette replacement successfully engineered recombinants were selectable with ampicillin (Additional file 1: Figure S2v). Finally, an RT-cassette, PCR-amplified with appropriate ODN pairs (Additional file 1: Tables S3 and S4) (Additional file 1: Figure S2vi) was introduced centrally, *via* positive selection recombineering (Tc, amp), into the FP CDS of each of these constructs generating the corresponding RT-cassette-containing derivatives (Additional file 1: Figure S2vii and Table S4).

A similar protocol was employed to generate an equivalent set of constructs containing F-CFP, F-GFP, F-YFP and Mc-mCherry CDSs. First, an RT-cassette, PCR-amplified with ODNs 12024/12025 (Additional file 1: Table S3) (Additional file 1: Figure S3i), was introduced into the pCC1Fos-based construct pNH043 (Additional file 1: Table S4) *via* positive-selection recombineering (Tc, amp), creating the intermediate pNH038 (note that constructs were not necessarily numbered sequentially) in which the mTFP1 CDS was replaced precisely by the RT-cassette but the flanking *Not*I sites were retained (Additional file 1: Figure S3ii). Next, F-CFP, F-GFP and F-YFP CDSs were PCR-amplified from, respectively, pPD136.61, pPD95.77 and pPD136.64 [4], using ODNs 12028/12029 (Additional file 1: Table S3), and Mc-mCherry from pAA64 [16] using ODNs 12026/12027 (Additional file 1: Table S3) (Additional file 1: Figure S3iii). The resulting PCR products replaced the RT-cassette in pNH038, *via* negative-selection recombineering, generating constructs pNH039-pNH042 (Additional file 1: Table S4) (Additional file 1: Figure S3iv). Finally, an RT-cassette, PCR-amplified using either ODN pairs 12030/12031 or 12032/12033 (Additional file 1: Table S3) and designed to target, respectively, the central region of the F-CFP/F-GFP/F-YFP and Mc-mCherry CDSs, was recombineered, *via* positive-selection recombineering (Tc, amp), into each of pNH039-42 to generate, respectively, pNH050-53 (Additional file 1: Figures S3v & S3vi and Additional file 1: Table S4).

DNA from each of these pCC1Fos-based constructs (Additional file 1: Tables S4) was isolated from appropriate individual 'mini-cultures' (5 ml, LB, Cm) in which fosmid copy number had been induced. A *Not*I-*Not*I fragment, encompassing the 'selection' or 'replacement' cassette, was excised from each, gel-purified, resuspended in water to either 1 or 100 ng/μl, depending on whether the fragment was to be used as, respectively, an RT-cassette PCR template or a direct replacement sequence, and stored in aliquots at −20°C until use.

Construction of singly and multiply-tagged fosmid-based reporters *via* iterative counter-selection recombineering

Following their generation a subset of these constructs were employed as reagents during iterative rounds of counter-selection recombineering, performed as described above, designed to tag each of the genes *F09E5.3*, *F09E5.15* (*prdx*-2) and *EEED8.6* (*ccpp*-6), located on the same fosmid genomic clone WRM069dD11, with a different FP CDS followed by derivation of double- and triple-tagged reporter constructs. Parallel workflows were initiated – one to incorporate the F-CFP, F-YFP and Mc-mCherry CDSs and two others to insert either the contiguous or two intron-containing sets of CDSs encoding mTFP1/mCitrine/mCherry designed and synthesized as part of the current work (Additional file 1: Figure S4). Briefly, a series of RT-cassettes were PCR-amplified, using ODNs (Additional file 1: Table S5) designed to

amplify the cassette from different *Not*I-*Not*I fragment templates isolated from the appropriate pCC1Fos-based selection construct (Additional file 1: Table S5), generating discrete RT-cassettes flanked by approx. 200 bp from the 5′ and 3′ ends of the final FP CDS to be inserted and with terminal homology arms designed to insert the complete PCR product into the target gene precisely 15 bp (5 codons) upstream of the stop codon (Figure 2). In order that the final fusion gene sequence terminated at the translation stop codon of the native gene the intention was for each reverse recombineering ODN to be designed to anneal immediately 5′ to the FP CDS stop signal. Such ODNs were designed successfully to generate RT-cassettes corresponding to the F-CFP, F-YFP and Mc-mCherry CDSs but the presence of *two* in-frame stop codons terminating each of the contiguous and intron-split codon-optimized FP CDSs was overlooked during ODN design resulting in an in-frame TAA triplet being incorporated into each associated RT-cassette (Additional file 1: Table S2) leading to final fusion reporter sequences lacking the last 5 codons of the endogenous gene (Figure 2).

Additional file

Additional file 1: Hirani et al Additional methods and data file [4,16-20,43-46].

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

NH generated all constructs, with assistance from MW, and transgenic *C. elegans* strains with assistance from MSG. NH undertook fluorescence microscopy and expression patterns were interpreted with guidance from IAH. PD provided bioinformatics help and analysis. CTD supervised NH, MW and MSG. NH and CTD wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank J Hodgkin, L Holden-Dye and H Bayliss for support and guidance, G Howell and C Thrasivoulou for help with confocal microscopy, E. Veal for helpful discussions on *prdx-2* and the Wellcome Trust (refs. WT082603, WT078981) and the BBSRC (ref. BB/E008038/1) for financial support.

Author details

[1]Institute of Pharmaceutical Science, King's College London, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NH, UK. [2]EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge CD10 1SD, UK. [3]Institute of Integrative and Comparative Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK. [4]Present address: Plant Protection Service (NPPO), National Reference Centre, Department of Molecular Biology, P.O. Box 9102, Wageningen 6700HC, The Netherlands.

Received: 25 June 2012 Accepted: 7 December 2012
Published: 3 January 2013

References

1. Brenner S: **Genetics of *Caenorhabditis elegans*.** *Genetics* 1974, **77**(1):71–94.
2. Antoshechkin I, Sternberg PW: **The versatile worm: genetic and genomic resources for *Caenorhabditis elegans* research.** *Nat Rev Genet* 2007, **8**(7):518–532.
3. C. elegans Sequencing Consortium: **Genome sequence of the nematode *C. elegans*: a platform for investigating biology.** *Science* 1998, **282:**2012–2018.
4. Fire A, Harrison SW, Dixon D: **A modular set of lacz fusion vectors for studying gene-expression in *Caenorhabditis elegans*.** *Gene* 1990, **93**(2):189–198.
5. Dolphin CT, Hope IA: ***Caenorhabditis elegans* reporter fusion genes generated by seamless modification of large genomic DNA clones.** *Nucleic Acids Res* 2006, **34**(9):e72.
6. Westenberg M, Bamps S, Soedling H, Hope IA, Dolphin CT: ***Escherichia coli* MW005: lambda Red-mediated recombineering and copy-number induction of oriV-equipped constructs in a single host.** *BMC Biotechnol* 2010, **10:**27.
7. Sarov M, Schneider S, Pozniakovski A, Roguev A, Ernst S, Zhang Y, Hyman AA, Stewart AF: **A recombineering pipeline for functional genomics applied to *Caenorhabditis elegans*.** *Nat Methods* 2006, **3**(10):839–844.
8. Zhang Y, Nash L, Fisher AL: **A simplified, robust, and streamlined procedure for the production of *C. elegans* transgenes via recombineering.** *BMC Dev Biol* 2008, **8:**119.
9. Tursun B, Cochella L, Carrera I, Hobert O: **A toolkit and robust pipeline for the generation of fosmid-based reporter genes in *C.*** *PLoS One* 2009, **4**(3):16.
10. Sharan SK, Thomason LC, Kuznetsov SG, Court DL: **Recombineering: a homologous recombination-based method of genetic engineering.** *Nat Protoc* 2009, **4**(2):206–223.
11. Stavropoulos TA, Strathdee CA: **Synergy between tetA and rpsL provides high-stringency positive and negative selection in bacterial artificial chromosome vectors.** *Genomics* 2001, **72**(1):99–104.
12. Budde MW, Roth MB: **The response of *Caenorhabditis elegans* to hydrogen sulfide and hydrogen cyanide.** *Genetics* 2011, **189**(2):521–U584.
13. Mansisidor AR, Cecere G, Hoersch S, Jensen MB, Kawli T, Kennedy LM, Chavez V, Tan MW, Lieb JD, Grishok A: **A conserved PHD finger protein and endogenous RNAi modulate insulin signaling in *Caenorhabditis elegans*.** *PLoS Genet* 2011, **7**(9):15.
14. Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M, Seraphin B: **The tandem affinity purification (TAP) method: A general procedure of protein complex purification.** *Methods* 2001, **24**(3):218–229.
15. Li Y: **The tandem affinity purification technology: an overview.** *Biotechnol Lett* 2011, **33**(8):1487–1499.
16. McNally K, Audhya A, Oegema K, McNally FJ: **Katanin controls mitotic and meiotic spindle length.** *J Cell Biol* 2006, **175**(6):881–891.
17. Shaner NC, Campbell RE, Steinbach PA, Giepmans BNG, Palmer AE, Tsien RY: **Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma sp* red fluorescent protein.** *Nat Biotechnol* 2004, **22**(12):1567–1572.
18. Rizzo MA, Springer GH, Granada B, Piston DW: **An improved cyan fluorescent protein variant useful for FRET.** *Nat Biotechnol* 2004, **22**(4):445–449.
19. Ai H-W, Henderson JN, Remington SJ, Campbell RE: **Directed evolution of a monomeric, bright and photostable version of *Clavularia* cyan fluorescent protein: structural characterization and applications in fluorescence imaging.** *Biochem J* 2006, **400:**531–540.
20. Griesbeck O, Baird GS, Campbell RE, Zacharias DA, Tsien RY: **Reducing the environmental sensitivity of yellow fluorescent protein - mechanism and applications.** *J Biol Chem* 2001, **276**(31):29188–29194.
21. Blumenthal T, Gleason KS: ***Caenorhabditis elegans* operons: form and function.** *Nat Rev Genet* 2003, **4**(2):112–120.
22. Huang P, Pleasance ED, Maydan JS, Hunt-Newbury R, O'Neil NJ, Mah A, Baillie DL, Marra MA, Moerman DG, Jones SJM: **Identification and analysis of internal promoters in *Caenorhabditis elegans* operons.** *Genome Res* 2007, **17**(10):1478–1485.
23. Shaner NC, Steinbach PA, Tsien RY: **A guide to choosing fluorescent proteins.** *Nat Methods* 2005, **2**(12):905–909.
24. Shaner NC, Patterson GH, Davidson MW: **Advances in fluorescent protein technology.** *J Cell Sci* 2007, **120**(24):4247–4260.
25. Buchman AR, Berg P: **Comparison of intron-dependent and intron-independent gene expression.** *Mol Cell Biol* 1988, **8**(10):4395–4405.
26. Kim JS, Raines RT: **Ribonuclease S-peptide as a carrier in fusion proteins.** *Protein Sci* 1993, **2**(3):348–356.
27. Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S: **Comparison of affinity tags for protein purification.** *Protein Expr Purif* 2005, **41**(1):98–105.

28. Goedhart J, von Stetten D, Noirclerc-Savoye M, Lelimousin M, Joosen L, Hink MA, van Weeren L, Gadella TWJ Jr, Royant A: **Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%.** *Nature Communication* 2012, **3:**751.
29. Markwardt ML, Kremers G-J, Kraft CA, Ray K, Cranfill PJC, Wilson KA, Day RN, Wachter RM, Davidson MW, Rizzo MA: **An improved cerulean fluorescent protein with enhanced brightness and reduced reversible photoswitching.** *PLoS One* 2011, **6**(3):e17896.
30. Gustafsson C, Minshull J, Govindarajan S, Ness J, Villalobos A, Welch M: **Engineering genes for predictable protein expression.** *Protein Expr Purif* 2012, **83**(1):37–46.
31. Racker E: **Enzymatic synthesis and breakdown of desoxyribose phosphate.** *J Biol Chem* 1952, **196:**347–365.
32. McKay SJ, Johnsen R, Khattra J, Asano J, Baillie DL, Chan S, Dube N, Fang L, Goszczynski B, Ha K, *et al*: **Gene expression profiling of cells, tissues, and developmental stages of the nematode *C. elegans*.** *Cold Spring Harb Symp Quant Biol* 2003, **68:**159–169.
33. Hunt-Newbury R, Viveiros R, Johnsen R, Mah A, Anastas D, Fang L, Halfnight E, Lee D, Lin J, Lorch A, *et al*: **High-throughput *in vivo* analysis of gene expression in *Caenorhabditis elegans*.** *PLoS Biol* 2007, **5**(9):1981–1997.
34. Rhee SG, Woo HA: **Multiple functions of peroxiredoxins: peroxidases, sensors and regulators of the intracellular messenger H_2O_2, and protein chaperones.** *Antioxid Redox Signal* 2011, **15**(3):781–794.
35. Isermann K, Liebau E, Roeder T, Bruchhaus I: **A peroxiredoxin specifically expressed in two types of pharyngeal neurons is required for normal growth and egg production in *Caenorhabditis elegans*.** *J Mol Biol* 2004, **338**(4):745–755.
36. Olahova M, Taylor SR, Khazaipoul S, Wang J, Morgan BA, Matsumoto K, Blackwell TK, Veal EA: **A redox-sensitive peroxiredoxin that is important for longevity has tissue- and stress-specific roles in stress resistance.** *Proc Natl Acad Sci U S A* 2008, **105**(50):19839–19844.
37. Kimura Y, Kurabe N, Ikegami K, Tsutsumi K, Konishi Y, Kaplan OI, Kunitomo H, Iino Y, Blacque OE, Setou M: **Identification of tubulin deglutamylase among *Caenorhabditis elegans* and mammalian cytosolic carboxypeptidases (CCPs).** *J Biol Chem* 2010, **285**(30):22934–22939.
38. Bamps S, Hope IA: **Large-scale gene expression pattern analysis, *in situ*, in *Caenorhabditis elegans*.** *Brief Funct Genomic Proteomic* 2008, **7**(3):175–183.
39. Mello CC, Kramer JM, Stinchcomb D, Ambros V: **Efficient gene-transfer in *C. elegans* - extrachromosomal maintenance and integration of transforming sequences.** *EMBO J* 1991, **10**(12):3959–3970.
40. Sambrook J, Fritsch EF, Maniatis T: *Molecular Cloning; A Laboratory Manual.* New York, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1989.
41. Yook K, Harris TW, Bieri T, Cabunoc A, Chan J, Chen WJ, Davis P, de la Cruz N, Duong A, Fang R, *et al*: **WormBase 2012: more genomes, more data, new website.** *Nucleic Acids Res* 2012, **40**(D1):D735–D741.
42. Eeckman FH, Durbin R: **ACeDB and Macace.** *Method Cell Biol* 1995, **48:**583–605.
43. Zacharias DA, Violin JD, Newton AC, Tsien RY: **Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells.** *Science* 2002, **296**(5569):913–916.
44. Schmidt TGM, Skerra A: **The strep-tag system for one-step purification and high-affinity detection or capturing of proteins.** *Nat Protoc* 2007, **2**(6):1528–1535.
45. Walker PA, Leong LEC, Ng PWP, Tan SH, Waller S, Murphy D, Porter AG: **Efficient and rapid affinity purification of proteins using recombinant fusion proteases.** *Bio/Technology* 1994, **12**(6):601–605.
46. Raines RT, McCormick M, Van Oosbree TR, Mierendorf RC: **The S•tag fusion system for protein purification.** *Method Enzymol* 2000, **326:**362–376.

doi:10.1186/1472-6750-13-1
Cite this article as: Hirani *et al.*: A simplified counter-selection recombineering protocol for creating fluorescent protein reporter constructs directly from *C. elegans* fosmid genomic clones. *BMC Biotechnology* 2013 **13**:1.

Appendix 2 – Solutions and media

Solutions

M9 Buffer - per litre; 3 g potassium phosphate (KH_2PO_4, Sigma), 6 g sodium dihydrogen phosphate (Na_2HPO_4, Sigma) and 5 g sodium chloride (NaCl, Sigma) dissolved in 1 l of dH_2O and autoclaved at 120 ° C for 20 minutes. After the solution had cooled down 1 ml of 1 M magnesium sulphate ($MgSO_4$, Sigma) was added prior to use.

10 x Annealing buffer – per 100 ml; 10 ml 1 M tris (Apollo scientific) at pH 6.8, 50 ml 1 M sodium chloride (NaCl, Sigma) and 2 ml 0.5 M EDTA (Sigma) at pH 8.0 diluted in dH_2O.

Medium

LB (Luria-Broth) agar – per litre; 9.5 g Tryptone (peptone, Sigma), 4.7 g yeast extract (Sigma), 9.5 g sodium chloride (NaCl, Sigma) and 14.3 g Agar (Apollo Scientific) dissolved in 1 l dH_2O and autoclaved at 120 °C for 20 minutes.

NGM (Nematode Growth Media) agar – per litre; 3 g Sodium chloride (NaCl, Sigma), 2.5 g Tryptone (peptone, Sigma) and 17.4 g Agar (Apollo Scientific) dissolved in 972 mls dH_2O. The media was autoclaved for 20 mins at 120 °C to sterilize and once cooled down to approximatelyimately 60°C the following was added: 1 ml 1 M calcium chloride ($CaCl_2$ Sigma), 1 ml 1 M magnesium sulphate ($MgSO_4$, Sigma), 1 ml 5 mg/ml cholesterol (dissolved in 95 % ethanol, Sigma) and 25 ml 1 M potassium phosphate (KH_2PO_4, both from Sigma).

LB (Luria-Burtani) medium – per litre; 10 g Tryptone (peptone, Sigma), 5 g Yeast (Sigma) and 10 g Sodium Chloride (NaCl, Sigma) dissolved in 1 l of dH_2O. The solution was autoclaved at 120 °C for 20 mins to sterilize.

SOB (-mg) medium – per litre; 19.5 g Tryptone (peptone, Sigma), 4.9 g Yeast extract (Sigma), 1 g sodium chloride (NaCl, Sigma) and 0.2 g potassium chloride (KCl, Sigma) dissolved in 1 l of dH_2O and autoclaved for 20 mins at 120 °C to sterilize.

Gel electrophoresis

50 x TAE stock solution – per litre; 242 g Tris base (Apollo scientific), 57.1 mls Glacial acetic acid (company) and 100 mls 0.5 M EDTA @ pH 8.0. Add 500 mls dH_2O until the tris base has dissolved and then adjust the total volume to 1l with dH_2O.

Antibiotics

Ampicillin (Amp) – made at a stock concentration of 50 mg/ml dissolved in dH_2O and filter sterilized (22 μM filter) before being aliquoted.

Chloramphenicol (Cm) – made at a stock concentration of 10 mg/ml, dissolved in 100 % ethanol and filter sterilized (22μM filter).

Tetracycline (A) (Tet (A)) – Stock made at a concentration of 5 mg/ml in 50 % ethanol and filter (22 μM) sterilized.

Streptomycin (Sm) – Stock concentration of 500 mg/ml made in dH_2O and filter (22 μM) sterilized.

Kanamycin (Km) – Stock of 10 mg/ml was dissolved in dH_2O and filter sterilized (22 μM filter).