Fire Lab 1999 Vector Supplement- May 1999

***gfp* color variants** | **Bacterial RNAi vectors**
Some new localization signals | **Hairpin constructs**
RNA-interference controls | **T7polymerase-based RNA expression**

Jamie Fleenor, Lisa Timmons, SiQun Xu, Kelly Liu, Bill Kelly, Andrew Fire
Carnegie Institution of Washington
Department of Embryology
115 West University Parkway
Baltimore, Md. 21210 USA

Also thanks to the following
For making & characterizing the Plasma Membrane Localization signal:
Harald Hutter (Max-Planck-Institut für medizinische Forschung, Heidelberg, Germany)
For identifying GFP isoforms suitable for differential labeling:
D.M. Miller, N.S. Desai, D.C. Hardin, D.W. Piston, G.H. Patterson (Vanderbilt University, Nashville, USA)

The vector kit has three components
1. This handout, with preliminary description of the new vectors
2. The microtiter vector array, containing vectors described in this handout.
Store -20°C. Instructions for handling the samples are described on the microtiter wills and at the end of this handout
4. The address of the ciw vector archive (www.ciwemb.edu)**.**

I. Summary

This kit contains several sets of vectors that should be useful in studies of gene expression and function in *C. elegans*. The vectors in this supplementary kit are in many cases derivatives of our earlier vectors. Documentation on the earlier vectors has been provided, and is still available from our web site (www.ciwemb.edu). In many cases, the vectors in this current kit have components or segments that can be simply excised and inserted in previous constructions to obtain a plasmid with desired properties. In working with this kit, we are assuming that the user has experience and knowledge of sequence-based design and construction of plasmids. We are providing both a general description and the full predicted DNA sequence for each of the constructs. Users should then be able to use their own software and design facilities to generate the restriction maps and other information needed for detailed construct design. The present kit contains the following
A. Fusion constructs with the coding region for Cyan-shifted and Yellow-shifted GFP
B. Fusion constructs with the coding regions for several newer GFP protein variants
C. Vectors for localization of expressed proteins to nucleoli and plasma membrane
D. Control plasmids for RNA-mediated interference
E. Vectors for Bacterial-mediated RNAi
F. Control constructs for Bacterial-mediated RNAi
G. Hairpin sequences to test promoters for ability to trigger RNAi in vivo
H. Components of an experimental system using T7RNA polymerase to express RNAs in *C. elegans*
I. Two miscellaneous plasmids for diverse applications

Suggestions for the "vector novice" and the "vector pro"

This supplementary kit is not intended to replace earlier expression vector kits sent out by this laboratory [see references 1,2,3]. If you are new to *C. elegans*, to studies with *gfp*, or to DNA-mediated transformation, you may wish to review literature from this lab and elsewhere concerning the nature of the system [see references]. If you have already produced *gfp* fusions and are hoping to obtain improved expression, we make the following suggestions (in order):

1. If you are using a single GFP construct to track expression, we still find that the best variant to use is S65C. This is the form of GFP that has been predominantly used in our previous kits.

2. For GFP double labeling, Miller and colleagues have found a combination of Cyan and Yellow to be most effective. Numerous vectors carrying these two GFPs are provided in this kit.

3. For carrying out initial RNA-mediated interference, we recommend synthesis of the RNA in vitro followed by injection into animals. Bacterial-mediated RNAi is generally much less effective. For in vitro synthesis of RNA, you can use a variety of readily available vectors (you don't need anything from us). We have included in this kit a number of control templates for producing characterized dsRNAs in vitro. We have also included a number of vectors and controls for Bacterial-mediated RNAi.

II. Description of individual vectors

A. Cyan-FP and Yellow-FP constructs

David Miller and colleagues have demonstrated that two distinct isoforms of GFP can be used with appropriate filter sets for double-labeling of transgenic worm strains. Details of appropriate filter sets for distinction of these GFP isoforms can be found in reference 4. Most of the constructs below have the myo-3 promoter [5] driving a localized version of one of the two GFP isoforms. These constructs are all built using our standard system of restriction enzyme cassettes, so that the GFP and/or promoter sequences can easily be swapped with any other vector or pre-existing construct that has been produced using our vectors. Note that all of these vectors have multiple introns interrupting the GFP coding region. This greatly stimulates expression level in *C. elegans* (apparently by improving RNA processing or transport steps (see the 1995 vector kit documentation for more description of intron use to stimulate gene expression). A downside of the multi-intron character of these vectors is that they will likely fail to function in any non-nematode species (although this has not been tested, the intron sequences are likely to be poorly removed in other species).

The amino acid changes present in the shifted GFPs are

CFP [referred to below as gf15]: Y66W, N146I, M153T, V163A
The gf15 isoform was actually included in our 1997 kit. It was based on the work of Heim and Tsien [6]

YFP [referred to below as gf42]: S65G, V68A, S72A, T203Y
The gf42 isoform was based on the work of Ormo et al. [7]

Table 1: constructs with CFP [gf15]: Y66W, N146I, M153T, V163A

plasmid	lig	promoter	Intracellular localization	reporter	introns	3' end	decoy
pPD136.61	L4816	*myo-3*	Nucleus≈cytoplasm	cfp [gf15]	åßð∑L	*unc-54*	no
pPD133.48	L4663	*myo-3*	Nucleus>cytoplasm	cfp [gf15]	åßð∑L	*unc-54*	no
pPD133.91	L4686	*myo-3*	Nucleolus>Nucleus	cfp [gf15]	Aåßð∑	*let-858*	yes
pPD133.54	L4665	*myo-3*	Mitochondria	cfp [gf15]	Aåßð∑L	*unc-54*	no
pPD133.51	L4664	*myo-3*	Plasma membrane	cfp [gf15]	Aåßð∑L	*unc-54*	yes
pPD133.82	L4682	*myo-3*	Cytoplasm only (GFP-ßgal)	cfp [gf15]::lacZ	åßð∑B-L	*unc-54*	no
pPD133.45	L4660	*myo-3*	Nucleus only (GFP-ßgal)	cfp [gf15]::lacZ	åßð∑B-L	*unc-54*	no
pPD135.41	L4780	*F22B7.9*	Mitochondria	cfp [gf15]	åßð∑L	*unc-54*	no
pPD134.96	L4752	*pes-10*	Nucleus≈cytoplasm	cfp [gf15]	åßð∑L	*unc-54*	yes

Table 2: constructs with YFP [gf42]: S65G, V68A, S72A, T203Y

plasmid	lig	promoter	Intracellular localization	reporter	introns	3' end	decoy
pPD136.64	L4817	*myo-3*	Nucleus≈cytoplasm	yfp [gf42]	åßð∑L	*unc-54*	no
pPD132.112	L4643	*myo-3*	Nucleus>cytoplasm	yfp [gf42]	åßð∑L	*unc-54*	no
pPD133.97	L4687	*myo-3*	Nucleolus>Nucleus	yfp [gf42]	Aåßð∑	*let-858*	yes
pPD133.60	L4667	*myo-3*	Mitochondria	yfp [gf42]	Aåßð∑L	*unc-54*	no
pPD133.58	L4666	*myo-3*	Plasma membrane	yfp [gf42]	Aåßð∑L	*unc-54*	yes
pPD133.86	L4683	*myo-3*	Cytoplasm only (GFP-ßgal)	yfp [gf42]::lacZ	åßð∑B-L	*unc-54*	no
pPD133.63	L4671	*myo-3*	Nucleus only (GFP-ßgal)	yfp [gf42]::lacZ	åßð∑B-L	*unc-54*	no
pPD135.49	L4781	*F22B7.9*	Mitochondria	yfp [gf42]	åßð∑L	*unc-54*	no
pPD134.99	L4753	*pes-10*	Nucleus≈cytoplasm	yfp [gf42]	åßð∑L	*unc-54*	yes
pPD132.102	L4640	*myo-2*	Nucleus≈cytoplasm	yfp [gf42]	Aåßð∑L	*unc-54*	yes

Promoters

myo-3: This promoter is generally active in body muscles (striated bodywall muscles as well as pharyngeal, vulval and intestine associated muscles; [5,8]). None of the coding region is contained in these constructs. Hence, the resulting fusions are transcriptional.

F22B7.9: This promoter is active from early stages in intestinal cells. Constructs with this promoter are translational fusions (derived from a construct originally received from I. Hope).

pes-10: This promoter is active in the embryo in each of the early somatic lineages as the lineage diverges from the germline. The above constructs are transcriptional fusions and carry the upstream activator sequences from the pes-10 promoter. Note: this promoter segment should not be confused with a shorter minimal promoter segment in some other vectors from this lab that has no activity in isolation but can be activated by numerous enhancers: the pes-10 promoter in L4752 and L4752 is complete in that with no additional enhancer, expression is driven in the early embryonic lineages. Despite the presence of two decoy sequences in these vectors, a small degree of posterior gut activity of the promoters is still present.

myo-2: This promoter is generally active in pharyngeal muscles [5]). None of the coding region is contained in these constructs. Hence, the resulting fusions are transcriptional. Constructs with this promoter can be toxic when present at high copy number, thus we generally dilute myo-2::GFP or myo-2::lacZ constructs substantially with the appropriate carrier or marker DNA before transformation.

Intracellular Localizations:

Nucleus≈cytoplasm: The reporter in these construct consists of the GFP coding region with no specific localization signal. If these reporters are used without any additional coding sequence, the result is a cytosol-filling fluorescent signal. Most cells expressing the reporter will show fairly uniform nucleus+cytoplasmic fluorescence, although some show slightly stronger fluorescence in nuclei.

Nucleus>cytoplasm: The reporter in this construct consists of GFP with the Nuclear Localization Signal (NLS) from SV40 T-Antigen at the N-terminus. Native GFP diffuses readily out of the nucleus. Hence if no additional protein coding sequence is attached in this construct, only partial nuclear localization of the fluorescence is observed. Hence the native construct gives only a slight preference of nuclear>cytoplasm. Addition of any additional mass to the protein can produce a more complete nuclear localization.

Nucleolus>Nucleus: The reporter in this construct consists of the GFP coding region with four tandem copies of the Nuclear Localization Signal (NLS) from SV40 T-Antigen at the N-terminus. The resulting protein (for reasons not known) is highly concentrated in the nucleolus, with a lower concentration in the nucleus. We have found that high level expression of this protein can be deleterious to some cells.

Mitochondrion: The reporter in this construct consists of GFP with the Mitochondrial import signal (MtLS) from Chicken Aspartate aminotransferase [9] at the N-terminus. This reporter is efficiently taken into mitochondria, which label strongly. This reporter protein may be more stable than cytosolic GFP. Protein fusions with additional sequence at the N-terminus of the protein would probably not be imported into mitochondria.

Plasma Membrane: The reporter in this construct consists of GFP with a hybrid signal-sequence/transmembrane domain from Pat-3 at the N-terminus. This Mb localization signal was developed by Harald Hutter and provides effective localization of the fluorescent reporter to the plasma membrane in many cells.

Cytoplasm only (GFP-ßGal): The reporter in this construct consists of the GFP coding region fused to an *E. coli lacZ* coding region. The resulting fusion protein has both GFP (fluorescence) and ß-galactosidase activities. The product is sufficiently large (due to the size and ability of ß-galactosidase to tetramerize) that it remains outside of the nucleus, yielding a completely cytoplasmic localization. Under certain circumstances, the distribution of this reporter protein is relatively uniform, with fluorescence spread throughout the cell. Under conditions that lead to brief high-level expression, the distribution may be less uniform, with formation of localized concentrations (like inclusion bodies) within the cytoplasm.

Nucleus only (GFP-ßGal): The reporter in this construct is an in frame fusion of SV40 T-antigen nuclear localization signal, GFP coding region, and E. coli lacZ coding region. The resulting protein has both GFP (fluorescence) and ß-galactosidase activities. The product is sufficiently large to be retained in the nucleus after transport, yielding a completely nuclear localization (in most cells). No activity is seen in the nucleolus). Under certain circumstances, the distribution of this reporter protein is relatively uniform, with fluorescence spread throughout the nucleus. Under other conditions localized concentrations (like inclusion bodies) may form.

Introns: The use of multiple introns to expression for C. elegans transgenes is described in the 1995 vector kit documentation. For convenience, the identities of the individual introns are described below

A A 42base intron upstream of the reporter coding regions. Although useful in some constructs, this intron can produce a situation in some fusions where only a short exon is present at the 5' end of the mRNA. Since it is not clear how short an exon is tolerated for C. elegans, we have omitted intron A in a subset of constructs

å,ß,∂ Three 51 base introns interrupting the *gfp* coding region. (Detailed description in the 1995 vector docs).

Π A fourth intron within GFP in some constructs, this is slightly longer and has a number of unique restriction sites to allow specific segments to be placed in an intron. (Detailed description in the 1997 vector docs)

Σ A 51b intron just downstream of GFP in some constructs. (Detailed description in the 1995 vector docs).

B-K Ten 51b introns spaced through the coding region for lacZ. (Detailed description in the 1995 vector docs).

L A 51b intron in the *unc-54* 3' UTR. (Detailed description in the 1995 vector docs)

3' end: Two different 3' ends are used in our general expression vectors, from unc-54 [5] and let-858 [10]. In somatic cells, these 3' ends seem interchangeable. The *let-858* 3' end constructs may be advantageous in producing C-terminal fusions to GFP as these offer several unique restriction sites in the relevant region. For germline expression, the *let-858* 3' end shows some advantages [M. Montgomery, personal communication]. Recent data from Dunn, Reese and Seydoux [11], however, suggests that an alternative 3' end (from the pie-1 gene) may be most useful for germline expression (pie-1 3' vectors can be obtained from the Seydoux lab [11]).

Decoy: This is a sequence (described in the 1995 vector kit) that is placed upstream of the polylinker that is used for promoter insertion. This sequence serves to prevent some readthrough transcripts from the plasmid vector from producing a functional GFP mRNA. Without the sequences, many constructs (particularly those with short promoter regions) can produce a background activity, often seen in the posterior gut and pharynx. The decoy doesn't complete eliminate this background, but can help to reduce it.

B. Other GFP variants

We have produced a number of other amino-acid sequence variants in the context of our intron-rich *gfp* coding region. These were produced for various reasons and have not been extensively tested. From the literature, some of the newer variants might be expected to have different spectral properties and thus be advantageous in specific applications. We have no simple quantitative way of comparing GFP spectra; we would certainly encourage any enthusiastic laboratory to generate such data. These mutations have been produced in test constructs carrying the *gfp* variant driven in body wall muscle by the *unc-54* or *myo-3* promoters. Any variants can be exchanged into existing vectors or fusion constructs by a simple restriction fragment swap. GFP variants 1-41 were originally described and distributed in our 1997 vector kit and are listed below for completeness. Variants 42-54 are new to this kit.

Variant GFP forms (including those in 1997 kit)

"Wild type"	[gf1]
S65T	[gf2]
S65C	[gf3]
F64L S65T	[gf4]
S65A V68L S72A	[gf5]
S65G S72A	[gf6]
Y66H	[gf7]
Y66W	[gf8]
F64L S65T Y145F	[gf9]
F64L S65T N146I	[gf10]
S65C M153A	[gf11]
F64L S65T N146I M153A	[gf12]
F64L S65T N146I M153T V163A	[gf13]
Y66H N146I M153T V163A	[gf14]
Y66W N146I M153T V163A	[gf15]
Y66H N146I	[gf16]
Y66W N146I	[gf17]
F64L S65T Y145F M153A	[gf18]
F64L S65T Y145F M153T V163A	[gf19]
Y66H Y145F M153T V163A	[gf20]
Y66W Y145F M153T V163A	[gf21]
Y66H Y145F "Blue *gfp* above"	[gf22]
Y66W Y145F	[gf23]
S65T Y145F	[gf24]
F64L S65T M153A	[gf25]
S65T M153A	[gf26]
F64L S65T M153T V163A	[gf27]
Y66H M153T V163A	[gf28]
Y66W M153T V163A	[gf29]
Y66H N146I M153A	[gf30]
Y66W N146I M153A	[gf31]
S65T N146I M153A	[gf32]
S65T N146I M153T V163A	[gf33]
S65T N146I	[gf34]
Y66H Y145F M153A	[gf35]
Y66W Y145F M153A	[gf36]
S65T Y145F M153A	[gf37]
S65T Y145F M153T V163A	[gf38]
Y66H M153A	[gf39]
Y66W M153A	[gf40]
S65T M153T V163A	[gf41]
S65G V68A S72A T203Y	[gf42]
S65C Y66H N146I M153T V163A	[gf43]
S65C Y66H Y145F	[gf44]
S65C Y66W N146I M153T V163A	[gf45]
F64L S65T Y66W Y145F	[gf46]
F64L S65T Y66H N146I M153T V163A	[gf47]
F64L S65T Y66H Y145F	[gf48]
F64L S65T Y66W N146I M153T V163A	[gf49]
F64L S65T Y66W Y145F	[gf50]
S65C T203Y	[gf51]
F64L S65T T203Y	[gf52]
S65G S72A T203Y	[gf53]
S65G V68A S72A	[gf54]

Except for the YFP and CFP isoforms described in section **A.**, we have little experience with the new gfp isoforms [gf43-gf54]. These were designed with the hope of producing more stable CFP and BFP fluorescence, but we have not tested this.

Table 3: constructs with variant *gfp* forms gf43-gf54

plasmid	lig	promoter[1]	localization[2]	reporter	designation	introns[3]	3' end	decoy
pPD122.61	L4061	*myo-3*	None	S65C Y66H N146I M153T V163A	[gf43]	åßð∑L	*unc-54*	yes
pPD122.64	L4062	*myo-3*	None	S65C Y66H Y145F	[gf44]	åßðL	*unc-54*	yes
pPD122.66	L4063	*myo-3*	None	S65C Y66W N146I M153T V163A	[gf45]	åßð∑L	*unc-54*	yes
pPD122.70	L4064	*myo-3*	None	F64L S65T Y66W Y145F	[gf46]	åßðL	*unc-54*	yes
pPD122.72	L4065	*myo-3*	None	F64L S65T Y66H N146I M153T V163A	[gf47]	åßð∑L	*unc-54*	yes
pPD122.75	L4066	*myo-3*	None	F64L S65T Y66H Y145F	[gf48]	åßðL	*unc-54*	yes
pPD123.19	L4095	*myo-3*	None	F64L S65T Y66W N146I M153T V163A	[gf49]	åßð∑L	*unc-54*	yes
pPD123.24	L4096	*myo-3*	None	F64L S65T Y66W Y145F	[gf50]	åßðL	*unc-54*	yes
pPD129.71	L4458	*unc-54*	SV40-NLS	S65C T203Y	[gf51]	Aåßð∑L	*unc-54*	no
pPD130.04	L4506	*unc-54*	SV40-NLS	F64L S65T T203Y	[gf52]	Aåßð∑L	*unc-54*	no
pPD130.07	L4507	*unc-54*	SV40-NLS	S65G S72A T203Y	[gf53]	Aåßð∑L	*unc-54*	no
pPD132.59	L4632	*unc-54*	SV40-NLS	S65G V68A S72A	[gf54]	Aåßð∑L	*unc-54*	no

[1]The *unc-54* and *myo-3* promoters both function in body muscle; the *unc-54* promoter shows distinct mosaicism in repetitive array contexts.

[2]NLS-GFP is relatively small and appears not to be efficiently retained in the nucleus.

[3]The Σ intron present in some of these constructs is just 3' to the *gfp* coding region. This intron significantly boosts expression of certain fusions. An S65C version of gfp carrying intron Σ is available in the 1995 kit (pPD94.81=L2406)

C. Vectors for localization of expressed proteins to nucleoli

We fortuitously found a concatamer of the SV40 T-Antigen Nuclear Localization Signal (SV40NLS) to effectively localize GFP to the nucleolus (some material is also present in the nucleus, but the majority is nucleolar). Several advantages and potential (and significant) problems will be noted for the use of these constructs as reporter vectors. In addition to their use as reporter vectors, these vectors should all be useful for directed localization of products to the nucleolus (although it should be stressed that the degree of localization with GFP is not 100%).

Advantages:

1. Fluorescence signal is concentrated in a small area of the cell
2. Signal is easily distinguished from other (non-nuclear) GFP patterns

Disadvantages:

1. High expression of 4xNLS-GFP apparently can hurt the organism (in that few transgenic lines are obtained). We don't know how strong the promoter needs to be to cause trouble or why these constructs are deleterious.
2. The 4xNLS tandem repeat structure, although short (42nt), could be a potential trigger of silencing.
3. A potential cryptic splice site in the 4xNLS might lead to abberantly spiced products (this has not been tested).
4. Hypodermal nucleoli are themselves autofluorescent under certain conditions in C. elegans [12]. This fluorescence has a somewhat different spectrum from GFP but in some filter sets would be easily confused.

Table 4: constructs with strong nuclear/nucleolar expression signal

plasmid	lig	promoter/frame	localization	reporter	introns	3' end	decoy	notes
pPD122.56	L4054	none/Frame #0	4xSV40-NLS	GFP[S65C]	Aåß∂∑L	*unc-54*	yes	Promoter insertion construct frame #0
pPD122.34	L4052	none/Frame #0	4xSV40-NLS	GFP[S65C]	Aåß∂	*let-858*	yes	Promoter insertion construct frame #0
pPD121.86	L4019	none/Frame #1	4xSV40-NLS	GFP[S65C]	Aåß∂∑L	*unc-54*	yes	Promoter insertion construct frame #1
pPD122.24	L4047	none/Frame #1	4xSV40-NLS	GFP[S65C]	Aåß∂	*let-858*	yes	Promoter insertion construct frame #1
pPD121.83	L4018	none/Frame #2	4xSV40-NLS	GFP[S65C]	Aåß∂∑L	*unc-54*	yes	Promoter insertion construct frame #2
pPD122.22	L4046	none/Frame #2	4xSV40-NLS	GFP[S65C]	Aåß∂	*let-858*	yes	Promoter insertion construct frame #2
pPD121.89	L4020	none/Frame S	4xSV40-NLS	GFP[S65C]	Aåß∂∑L	*unc-54*	yes	Promoter insertion construct frame S
pPD122.27	L4048	none/Frame S	4xSV40-NLS	GFP[S65C]	Aåß∂	*let-858*	yes	Promoter insertion construct frame S
pPD122.13	L4041	none	4xSV40-NLS	GFP[S65C]	Aåß∂	*let-858*	yes	Has rare restriction sites to insert PCR fragments
pPD121.95	L4025	Δ*myo-2*	4xSV40-NLS	GFP[S65C]	Aåß∂∑L	*unc-54*	yes	Enhancer assays (inactive deleted *myo-2* promoter)
pPD122.28	L4049	Δ*myo-2*	4xSV40-NLS	GFP[S65C]	Aåß∂	*let-858*	yes	Enhancer assays (inactive deleted *myo-2* promoter)
pPD122.53	L4053	Δ*pes-10*	4xSV40-NLS	GFP[S65C]	Aåß∂∑L	*unc-54*	yes	Enhancer assays (inactive deleted *pes-10* promoter)
pPD122.32	L4051	Δ*pes-10*	4xSV40-NLS	GFP[S65C]	Aåß∂	*let-858*	yes	Enhancer assays (inactive deleted *pes-10* promoter)
pPD121.92	L4026	Δ*glp-1*	4xSV40-NLS	GFP[S65C]	Aåß∂∑L	*unc-54*	yes	Enhancer assays (inactive deleted *glp-1*promoter)
pPD122.11	L4040	*myo-2*	4xSV40-NLS	GFP[S65C]	Aåß∂	*let-858*	yes	Pharyngeal muscle expression
pPD122.45	L4037	*myo-3*	4xSV40-NLS	GFP[S65C]	Aåß∂	*let-858*	yes	Body muscle expression
pPD122.15	L4042	*hsp16-41*	4xSV40-NLS	GFP[S65C]	Aåß∂	*let-858*	yes	Heat shock inducible expression
pPD122.18	L4043	*hsp16-2*	4xSV40-NLS	GFP[S65C]	Aåß∂	*let-858*	yes	Heat shock inducible expression
pPD122.08	L4038	*let-858*	4xSV40-NLS	GFP[S65C]	Aåß∂	*let-858*	yes	General expression (*let-858* promoter/enhancer)
pPD129.51	L4453	*rps-5*	4xSV40-NLS	GFP[S65C]	åß∂	*let-858*	yes	General promoter (ribosome protein S5 gene)
pPD129.57	L4455	*rpl-28*	4xSV40-NLS	GFP[S65C]	åß∂	*let-858*	yes	General promoter (ribosome protein L28 gene)

Promoterless constructs

These have an upstream polylinker for insertion of your favorite promoter region. Note that these constructs are suitable for production of translational fusions, but that the effect of the 4xNLS segment internal to a fusion protein has not been studied. For making translational fusions with promoterless vectors, it is critical that the fusion between your inserted DNA and the reporter be in frame. The reading frame that continues into *lacZ* is shown. The entire upstream multiple-cloning sites reproduced below is present in each of the promoterless vectors (see also Fire et al. [1]; The frame need not be considered in enhancer assay vectors.) The TAG sequence in the *XbaI* site of frame I precludes the use of upstream sites in this set of vectors. This has been remedied by producing a special set of vectors (labeled frame "S"), which provide the missing frame for HindIII, SphI, PstI, and SalI fusion junctions.

```
HindIII      SphI    PstI  SalI   XbaI   BamHI   SmaI       BalI
  v           v       v  v       v       v       v          v      Frame 0
A AGC TTG CAT GCC TGC AGG TCG ACT CTA GAG GAT CCC CGG GAT TGG CCA ...

HindIII     SphI    PstI SalI    XbaI   BamHI    SmaI      BalI
 v           v       v v       v       v       v          v       Frame 1
aa gct tgc atg cct gca ggt cga ctc tag AGG ATC CCC GGG ATT GGC CA ...

HindIII    SphI   PstI   SalI   XbaI   BamHI     SmaI      BalI
 v          v       v  v       v       v       v          v       Frame 2
AAG CTT GCA TGC CTG CAG GTC GAC TCT AGA GGA TCC CCG GGA TTG GCC A ...

HindIII     SphI   PstI  SalI          BamHI  SmaI       BalI
 v           v       v v             v       v          v       Frame S
AA GCT TGC ATG CCT GCA GGT CGA CTA GAG GAT CCC CGG GAT TGG CCA ...
```

Enhancer Assay Vectors

Δ*myo-2*: This is a deleted version of the myo-2 promoter that is not active without additional enhancement [5]. This promoter can be activated in the proximity of enhancers active in different muscle types, providing an assay within the musculature for the activity pattern of a given enhancer. Some enhancement in non-muscle pharyngeal cells has been seen, but the activity of this promoter in the body has so far been limited to muscle.

Δ*pes-10* This is a deleted version of the pes-10 promoter that is essentially inactive without additional enhancement. This promoter can be activated in many (if not all) somatic tissues by an appropriate enhancer.

Specific promoter constructs:

None of the coding region is contained in these constructs. Hence, the resulting fusions are transcriptional.

myo-2:This promoter is generally active in pharyngeal muscles [5]). Constructs with this promoter can be toxic when present at high copy number, thus we generally dilute myo-2::GFP or myo-2::lacZ constructs substantially with the appropriate carrier or marker DNA before transformation.

myo-3: This promoter is generally active in body muscles (striated bodywall muscles as well as pharyngeal, vulval and intestine associated muscles; [5,8]).

let-858: This promoter/enhancer combination is active in the whole animal [10]. Expression is moderate-to-strong

rps-5, rpl-28: These promoters are also active at moderate-to-high levels in multiple tissues.

hsp16-41, hsp16-2:These are heat/stress inducible promoters. Activity patterns for these promoters have been described previously [1,13].

Second generation NLS segment

The following constructs contain a modified SV40-NLS region that eliminates a potential cryptic splice site and also includes a sequence degeneracy in third base positions that avoids a duplication in the coding sequence. We don't have any data yet which would indicate whether these factors improve expression. A limited set of restriction enzymes is available for transfer of the NLSn-GFP coding region to other constructs (KpnI/Asp718I upstream of GFP is the only straightforward site. A variety of sites within or downstream of GFP can be used for the swap (NcoI, XhoI, MfeI, EcoRI, NgoMIV, SpeI, ApaI, Bsp120I).

Table 5: constructs with second generation nuclear/nucleolar expression signals

plasmid	lig	promoter/frame	localization	reporter	introns	3' end	decoy	localization of simple GFP fusion
pPD135.02	L4759	*let-858*	1xSV40-NLS	GFP[S65C]	åßð	*let-858*	yes	limited nuclear localization
pPD135.83	L4796	*let-858*	2xSV40-NLS	GFP[S65C]	åßð	*let-858*	yes	nucleolus & nucleus
pPD136.15	L4809	*let-858*	3xSV40-NLS	GFP[S65C]	åßð	*let-858*	yes	nucleolus > nucleus

D. Constructs with Harald Hutter's plasma-membrane localization signal.

These constructs contain the gfp coding sequence preceded by a plasma membrane localization signal devised by Harald Hutter. The signal consists of a secretion signal sequence followed by a transmembrane domain; both are from coding region for pat-3 integrin [14]. These vectors can serve as a source of a Membrane-targeted GFP reporter, or as a means to target a specific protein product to the membrane. Note that the secretion signal at the beginning is designed to function on the N-terminus of a protein, and would likely not function if placed more than a few amino acids past the N-terminus of a protein. These vectors are easily used for transcriptional fusions. Alternatively, protein coding sequences C-terminal to GFP (this will be easiest in L4057), so that the appended sequence will remain in the cytosol. Finally, a BspE1 site between the signal sequence and transmembrane domain has been put in by Hutter to allow insertion of segments that should then end up outside the cell.

Table 6: constructs with plasma membrane localization signal

plasmid	lig	promoter/frame	localization	reporter	introns	3' end	decoy	notes
pPD122.36	L4057	*myo-3*	pat3-MbLS	GFP[S65C]	Aåßð	*let-858*	yes	Bodywall muscle expression
pPD122.39	L4058	none (polylinker)	pat3-MbLS	GFP[S65C]	Aåßð	*unc-54*	yes	Cloning sites for GFP fusion
pPD122.07	L4035	*pes-10*	pat3-MbLS	GFP[S65C]	åßð	*let-858*	yes	Promoter active in early somatic lineages

E. Control plasmids for RNA-mediated interference

RNA-mediated interference is an effective tool for specifically disrupting the function of individual genes or groups of genes in C. elegans [15,16]. A full description of the protocols used in this lab for RNA-mediated interference can be found in [17]. The following constructs are provided for in vitro synthesis of RNA to be injected into the worm. Note that one very effective control construct (containing the GFP coding region with flanking T3 and T7 promoters) can be found in our earlier vector kits (pPD79.44 in the 1995 kit), or in the original Chalfie lab GFP vector set (pTU#65). The following clones provide a series of additional controls.

Table 7. Control plasmids for RNA-mediated interference

plasmid	lig	gene	insert size	Sense RNA	Antisense RNA	notes
pPD34.09	L453	unc-22	743 (all from exon)	T3 > XbaI	T7 > XhoI	Affected worms twitch strongly
pPD123.70	L4129	unc-54	652 (all from exon)	T3 > Bsp120I	T7 > ClaI	Affected worms are slow or paralyzed
pPD123.58	L4119	fem-1	662 (532 from exon)	T3 > XhoI	T7 > SpeI	Affected worms are female
pPD123.102	L4138	*E. coli lacZ*	830 (all exon)	T7 > ClaI	T3 > Ecl136II	Interferes with lacZ reporter constructs
pPD136.68	L4819	CAT	547 (all exon)	T7 > NcoI	T3 > SalI	Bacterial chloramphenicol acetyl transferase

F. Vectors and controls for Bacterial-mediated RNAi

Plasmid **pPD129.36 (L4440)** carries two promoters for T7 RNA polymerase facing each other [18]. It is designed for facile insertion of coding regions for genes to be targeted for interference. We generally produce the relevant clones using a standard *E. coli* cloning host used. After characterization, the plasmid is transferred to a bacterial strain expressing T7-RNA polymerase from an inducible promoter. We have sent such a bacterial strain to the C. elegans genetic stock center. Bacterial-mediated RNA interference is less effective than injection in almost every case that we have tested. We are actively working on improvements to this technique... for the moment, we recommend the injection as a first approach if your primary goal is to obtain a preliminary indication of loss-of-function phenotype. We have also included a number of plasmids which are derived from 129.36 by insertion of coding regions for which interference can be assayed

Table 8. Plasmids for Bacterial-mediated RNA interference [18]

plasmid	lig	insert	notes
pPD129.36	L4440	*none*	Insert coding region of interest between T7 polymerase sites
pLT61.1	LT61	*unc-22*	Affected worms twitch strongly
pLT63.1	LT63	*fem-1*	Affected worms are female
pPD128.110	L4417	*gfp*	Interferes with gfp reporter constructs

G. Hairpin sequences

Inverted repeat or hairpin sequences provide a ready means to produce a single RNA that will fold back on itself to form a duplex. We have used both in-vitro and in vivo synthesized hairpin RNAs as triggers for interference. In order to propagate the hairpin in an E. coli plasmid, it is necessary to interrupt the hairpin with a region of unique sequence. This is the loop in each of these cases. Each of these plasmids carries sequences from a target gene that are present in two copies in inverted repeat orientation. The two copies are separated by a spacer DNA. The constructs are in vectors with a promoter for T7 RNA polymerase. In some cases also a promoter for T3 RNA polymerase and/or for E. coli RNA polymerase [the lac promoter] is also present.

Table 9. Plasmids carrying inverted repeat hairpins

plasmid	lig	stem of inverted repeat	loop of inverted repeat	promoters	terminator/3'
pPD128.117	L4400	segment from *unc-22*	segment from *gfp*	T3,T7,lac	None
pPD126.25	L4271	*gfp*	segment from *unc-22*	T3,T7,lac	None
pPD128.86	L4404	segment from *unc-54*	another *unc-54* segment	T3,T7,lac	None
pPD129.23	L4433	segment from *unc-22*	segment from *gfp*	T7,lac	T7-ter
pPD129.17	L4431	*gfp*	segment from *unc-22*	T7,lac	T7-ter
pPD137.21	L4842	chloramphenicol acetyl transferase	segment from *cat*	T7,T3,lac	None

H. Components for expression of T3 and T7 RNA polymerase in *C. elegans*

We have had limited success producing RNA interference by expressing T3 or T7 polymerase in C. elegans, then using the polymerase to express an inverted repeat and thereby produce gene- and tissue-specific silencing. The T3 and T7 polymerases cassettes are described below. The following constructs have an intron in the T-polymerase coding region (to maximize c. elegans expression while minimizing bacterial toxicity). The pes-10 promoter can be easily removed and a promoter of choice substituted. We have carried out only limited testing of these constructs.

Table 10. Expression cassettes for T3 and T7 RNA polymerase

plasmid	lig	promoter	coding region	3' UTR	notes
pPD129.27	L4437	*pes-10*	T7 RNA polymerase with flanking introns (5' & 3') and internal intron	*unc-54*	
pPD129.29	L4438	*pes-10*	T3 RNA polymerase with flanking introns (5' & 3') and internal intron	*unc-54*	

I. Two miscellaneous plasmids for diverse applications

A modular *gfp* reporter construct with no intron and S65C mutation

pBK125: This construct has been useful for splicing and early embryo metabolism studies. After heat shock, the construct expresses at high levels in embryos independent of a functional splicing system (B. Kelly, unpublished)

A modular construct with the CeTwist promoter.

pKL464.1: After some detailed analysis of the promoter for the C. elegans homolog of Twist (Harfe et al., 1998; K. Liu, unpublished), we have come up with a promoter fragment that strongly drives a reporter coding region with no additional sequences in the M Cell lineage. This promter (with a GFP insert that can be easily removed) is incorporated into plasmid pKL464.1. Note that the ATG from CeTwist, but no additional amino acids are present. This provides a facile means for expressing a gene of interest either by precise engineering of the coding region following the twist ATG or using polylinker sites just after the ATG.

Appendix B. Explanation of a few of the more obscure terms in describing these vectors.

What follows below are overall descriptions of the vectors being distributed with this kit. The convenience of the microtiter well distribution format allows us to distribute a large number of vectors that might have applications in several different types of studies. It is anticipated that only a fraction of the vectors will be used by any given lab. These descriptions should allow the experimenter to see which vectors are most appropriate for the task at hand. In order to actually carry out the constructions, it is anticipated that the experimenter will retrieve the complete sequence of the vector from the **web** archive and use this to choose specific restriction sites and cloning strategies. All of the predicted structures are based on knowledge of the individual elements and a reconstruction of the cloning steps used to produce the vectors. At each step, we've done considerable checking with restriction enzymes to confirm structures. We can't rule out unexpected sequence changes, although most of the vectors below are at most one construct removed from a vector that has been tested in vivo for efficacy.

Plasmid #: This gives the unique number of the DNA preparation that is being distributed.

Ligation #: Ligation numbers uniquely indicate the structure of the construct described. These allow you to reference the proper sequence in the electronic archive.

Promoter: The new reporter segments have been tested with a variety of previously characterized promoter segments. The precise details of fragments used for promoter activity can be derived from the sequences.

IVS: Number and distribution of synthetic intervening sequences

𝔸 is the original synthetic IVS inserted upstream of the reporter coding region in the older *lacZ* vectors [1]

𝔹, ℂ , 𝔻 , 𝔼 , 𝔽 , 𝔾 , ℍ , 𝕀 , 𝕁 , 𝕂 are unique synthetic introns inserted into the coding region of *lacZ*

å , ß , ∂ are unique synthetic introns inserted into the *gfp* coding region

𝕄 is a synthetic intron with an internal multiple cloning region inserted into *gfp*

𝕊 is a synthetic intron inserted just upstream of the *unc-54* 3' UTR sequence

𝕃 is a synthetic intron inserted into the 3' UTR of *unc-54*

3' end: Vector 3' end sequences from *unc-54* and *let-858* are used.

decoy: decoy **+** vectors have a (synthetic intron--->short coding region) minigene upstream of the MCS, to decrease background from readthrough transcription. "double" decoy vectors have two different decoys upstream.

References

1. Fire, A, Harrison, S., and Dixon, D. (1990). A modular set of *lacZ* fusion vectors for studying gene expression in *Caenorhabditis elegans*. Gene 93, 189-198.
2. Fire, A., Ahnn, J., Seydoux, G., and S. Xu (1995) FireLab 1995 Vector kit documentation. www.ciwemb.edu
3. Xu, S., Kelly, W., Harfe, B., Montgomery, M., Ahnn, J., Getz, S., and Fire, A. (1997) FireLab 1997 Vector supplement documentation. www.ciwemb.edu
4. Miller, D.M., Desai, N., Hardin, D., Piston, D.W., Patterson, G.H., Fleenor, J., Xu, S., and Fire, A. (1999) A two-color GFP expression system for *C. elegans*. Biotechniques 26, 914
5. Okkema, P., White-Harrison, S., Plunger, V., Aryana, A., and Fire, A. (1993). Sequence requirements for myosin gene expression and regulation in *C. elegans*. Genetics *135,* 385-404.
6. Heim, R., and Tsien, R.Y. 1996. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. Current Biology 6:178-182.
7. Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ (1996) Crystal structure of the Aequorea victoria green fluorescent protein. Science 273:1392-1395
8. Ardizzi, J. P., and H. F. Epstein 1987 Immunochemical localization of myosin heavy chain isoforms and paramyosin in developmentally and structurally diverse muscle cell types of the nematode Caenorhabditis elegans. J. Cell Biol. 105: 2763-2770.
9. Jaussi, R., Cotton, B., Juretic, N., Christen, P., and Schumperli, D. (1985) The primary structure of the precursor of chicken mitochondrial aspartate aminotransferase. Cloning and sequence analysis of cDNA. J. Biol Chem. 260,16060-160633
10. Kelly, W.G., Xu, S., Montgomery, M., and Fire, A. (1997) Distinct Requirements for Somatic and Germline Expression of a Generally Expressed *C. elegans* Gene. **Genetics**, 146: 227-238.
11. Dunn, M., Reese, K. and Seydoux, G. (1999) A pie-1-based vector for maternal expression of gfp fusions. Worm Breeder's Gazette 15: 18
12. Fire, A., Ahnn, J., Kelly, W., Harfe, B., Kostas, S., Hsieh, J., Hsu, M., and Xu, S. (1998) GFP applications in *C. elegans*. in GFP Strategies and Applications, M. Chalfie and S. Kain eds, John Wiley and Sons, NY. pages 153-168
13. Stringham, E.G., Dixon, D.K., Jones, D., and Candido, E.P.M. (1992) Temporal and spatial expression patterns of the small heat-shock (HSP16) genes in transgenic *Caenorhabditis elegans*. Mol. Biol. Cell 3, 221-233.

14. Gettner SN, Kenyon C, Reichardt LF (1995) Characterization of beta-pat-3 heterodimers, a family of essential integrin receptors in C. elegans. J. Cell Biol. 129: 1127-1141

15. Guo, S. & Kemphues, K. *par-1*, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* **81**, 611-620 (1995).

16. Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998) Potent and specific genetic interference by double stranded RNA in *C. elegans*, **Nature** 391, 806-811

17. "FireLab RNAi Docs" v 1.0 (2/98) For most recent version see our web site: www.ciwemb.edu

18. Timmons, L., and Fire, A. (1998) Specific Interference by ingested dsRNA. Nature, 395, 854

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

20. Hope, I.A. (1991). 'Promoter trapping' in *Caenorhabditis elegans*. Development. 113, 399-408.

21. Marck, C.: 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res. 16 (1988) 1829-1836.

22. Mello, C., and Fire, A. "DNA transformation" (Methods in Cell Biology: *C. elegans*, D. Shakes and H. Epstein eds., Academic Press [publication 1995].)

23. Mello, C.C., Kramer, J.M., Stinchcomb, D. and Ambros, V. (1991). Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. EMBO J. 10:3959-3970.

A Note about printing this document: This document should print well using the standard MacIntosh Laserwriter priniter drivers. You may experience some loss of special characters (primarily seen in intron names) if you use another printer driver (Windows or HP laserjet).

Acknowledgments:

We would like to thank S. Kostas, J. Hsieh, A. Pinder C. Mello and our other colleagues at this institution for their help and suggestions. Work in this lab is supported by the National Institutes of Health (Grant R01-GM37706) and the Carnegie Institution.

Fire Lab 1999 Vector Supplement- Physical Instructions

This is a microtiter rack with vectors from the Fire Lab. Each well has one individual vector. Stocks were dried from 5μl [10mM Tris-Cl (pH8) 1mM EDTA 5% Glycerol 1μg/ml tRNA 0.1mg/ml BSA]. DNA stocks used were very dilute (~~1μg/ml). The vectors should be rehydrated by adding 40μl H20, then transformed into a suitable *E. coli* host to obtain working stocks. The bacteria we use are DH5-alpha. All plasmids provide amp resistance.

Do not peel off the sealing film on top of the microtiter plate. To obtain a given plasmid, puncture the corresponding well with a standard pipette tip prefilled with 40μl of H2O(wipe the cover gently with a wet, then a dry kimwipe before this to avoid contamination). Pipet up and down to resuspend DNA, then take 1-2μl to transform *E. coli*. Cover opened wells with small pieces of lab tape or adhesive sealing film. As each plasmid is grown up, use restriction digestion (e.g. DdeI) compared with known sequence to help confirm that no cross- contamination has occurred. **Store -20°C this side up**

For locations of vectors, see the map with the kit documentation.

Remember to check the web archive for corrections. Always wear your seatbelt. "transgene activity patterns cannot under any circumstances be used as the sole means to determine the physiological expression pattern of an endogenous gene." We are sending this kit out (one per institution) as of June 1999. You are expected to share this kit with others at your institution. If you are leaving the institution, you can "duplicate" the kit by pipetting small aliquots from each well into duplicate wells. We make no guarantee to continue to distribute the kit.

Revision history
May 1999: Version 1.0. This is the first posting of documentation for 1999 Kit. This is a preliminary version of the documentation. Users are advised to continue to check the archive for updates over the next few weeks.

Version 1.1
Alerts July 26, 2000
1999 Vector Kit

1.
Due to a clerical error on our part, vector pPD121.92 was incorrectly annotated. This vector should be annotated L4026 (not L4025 as stated in the original documentation). The vector that was included with the kit, pPD121.92(L4026), actually contains a deleted version of the glp-1 promoter (not myo-2) driving the 4xNLS-tagged GFP. This promoter is of (at best) limited usefulness in enhancer assays.

Menwhile, the correct vector for L4025 (i.e., the deleted myo-2 promoter driving the 4xNLS-GFP) is pPD121.95. This vector was not included with the 1999 kit; anyone who needs this vector, and has filled out the paperwork for the 1999 kit should send us a note requesting it.

2.
The promoter referred to as "rpl-5" in the original documentation is actually for the small ribosomal protein-5 gene, and has thus been renamed "rps-5"

Thanks to Adrian Streit and an anonymous reviewer for pointing out these errors.