FireLab 2005 Mini-vector Kit
July 2005
Documentation "alpha" version
Direct contributors to making and testing vectors in this kit
In our lab:
Jamie Fleenor, SiQun Xu, Chaya Krishna, Luiz Pantalena Filho, Daniel Blanchard, Javier Lopez, Lisa Timmons.

Friends Elsewhere: Pam Hoppe, David Miller, Rosaling Lee, Vector Ambros, Randall French, Jim Kramer, Craig Mello, Meera Sundaram, John Yochem, Min Han, Stuart Kim, Roger Tsien, Bob Waterston, Morris Maduro, David Pilgrim, Yishi Jin, Phil Anderson, Geraldine Seydoux, David Hsu, Barbara Meyer, Peg MacMorris, Tom Blumenthal, Bob Horvitz, Sydney Brenner, Kazunori Kondo, Jonathan Hodgkin... more

In honor of having passed the mantle of distributing the physical vector kits, we have released a number of vectors which we think may be of use in the field.

These are enumerated below in rough categories of which components are "new". As always, these vectors are (for the most part) not our creations but are cobbled together from parts that we have leanred about and in some cases physcially obtained from others in the scientific community (and from the worm itself). Also as always, please note that while we have done our best to ensure that these will be useful tools and reagents, nothing in life is guaranteed. Plese check the structures of the vectors as you use them and be aware that there may be errors in the documentation, the descriptions, or the sequence information.

In addition to releasing this set of vectors, I have gone through the sequence files of previous vectors in an attempt to provide some "fixes" for known errors in sequence. These errors generally have little effect on vector function but can have specific effects on restriction digests used to check the eventual clones that are produced from the vectors. I am sure there are still some errors in the virtual vector sequences... they should just be a little fewer and further between. The vector structures and some documentation has been assembled in two formats for this release. One is a standard tab-delimited excel spreadsheet format which I hope will be self explanatory. The other is as a "DNA Janitor" database

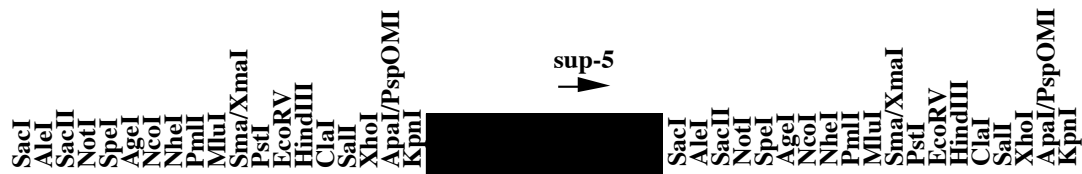which should run on any Macintosh that can still manage to run system 9.

Enjoy.

Selectable Markers

L3099 (sup5 segmnt btween MCS')

This is a conveniently placed clone of the amber-suppressing version of the sup-5 gene. It can be used as a selectable marker in a number of strains that carry readily selected amber mutations. Some nice examples of rescuable amber mutations are hlh(cc450am) and tra-3(e1107am). Also consider that nuc-1(e1392am) confers methotrexate sensitivity due to a need for purines derived from digestion of bacterial DNA in the gut. Rescue of the amber apparently rescues this. One thing to remember: high copy numbers of amber suppressor genes are lethal, even low copy numbers can cause distress to the worm population. Toxic effects are less dramatic at higher temperatures (e.g. 23C), but so is the amber suppression.

Thanks for information, etc: Sydney Brenner, Bob Waterston, Jonathan Hodgkin, Craig Mello, Dan Stinchcomb, Lihsia Chen, Jamie Fleenor, Siqun Xu, Kazunori Kondoh.

L3099 Map of sup-5 segment and flanking restriction sites

sup-5

SacI AleI SacII NotI SpeI AgeI NcoI NheI PmlI MluI Sma/XmaI PstI EcoRV HindIII ClaI SalI XhoI ApaI/PspOMI KpnI   SacI AleI SacII NotI SpeI AgeI NcoI NheI PmlI MluI Sma/XmaI PstI EcoRV HindIII ClaI SalI XhoI ApaI/PspOMI KpnI

L4624+ (lin4-687b in BS)

This contains a portable/functional (wild type) lin-4 segment inserted into a Bluescript-like polylinker. It serves as a fairly effective selectable marker in a lin-4 mutant background.

Thanks for information, etc: Rosalind Lee, Victor Ambros, Bob Horvitz, Marty Chalfie, John Sulston, Jamie Fleenor.

L4624+ Map of lin-4 segment and flanking restriction sites

(BssHII) SacI AleI NotI SpeI AgeI NcoI NheI PmlI Sma/XmaI PstI EcoRI EcoRV HindIII ClaI HincII/SalI

lin-4

KpnI/Asp718I EcoO109I ApaI/PspOMI (BssHII)

Some Reporters

Note that most of these are provided as single plasmids
that carry the reporter downstream of a standard promoter.
These are all based on the same "general vector plan" as
the original vector kit vectors. Thus you should be able
to switch reporters in any given construct with relative
ease by just exchanging the promoter and/or reporter
segments between plasmids.

L4901 myo3-EGFP-u'

   The gfp coding regions used in most vertebrate assays
is a version that has codons changed to produce a
humanized codon bias. This reporter works in C. elegans,
although somewhat less well than the standard C. elegans-
optimized vectors that are in the kit. Nevertheless, it
is occaisionally useful to have a vector that will either
a. Not recombine with a co-injected second GFP fusion at
high frequency, or
b. Be easily exchanged into vertebrate expression
vectors.
The gfp isoform in this vector is "EGFP": F64LS65T

Thanks for information, etc: Brian Seed, Marty Chalfie,
Jamie Fleenor.

L7136 pPD187.24 pie1-SPGFP-pie'
L7152 SPhased-GFC in pCDNA3

   Here is another gfp reporter with only minimal nucleic
acid homology to any of the standard gfp reporters.
Again the encoded protein is still gfp, just with a very
different use of codons and intron sequences in a fully
"synthetic" gene. This vector has introns and a codon
bias similar to that in C. elegans as well as signals in
the introns that may act to constrain nucleosome
positining and thereby (possibly very slightly) decrease
the degree of silencing observed. The introns in this
vector are longer than in the previously-standard C.
elegans vectors. Thus the reporter works in both worms

and mammalian cells (it would be expected also to work in yeast).
The gfp isoform in this reporter is "S65C" (similar to that in the majority of optimized vectors in older kits).

L7136 carries this gfp driven by the pie-1 promoter for exchange with standard C. elegans reporters.
L7152 carries the reporter in the standard mammalian expression vector pCDNA3 (CMV promoter, bovine growth hormone polyA).

Thanks for information, etc: Fred Tan, Rosa Alcazar, Jamie Fleenor, Daniel Blanchard, Ed Trifonov, Jon Widom, Andrew Travers, Horace Drew, Chen Ming Fan. The pie-1 vector is based on a number of very nice germline expression vectors originally made by the Seydoux lab.

L5743 myo3-gfC-smg5-l'
Knowledge of the pathway for degradation of mRNAs with aberrant 3' ends has been very useful for generating regulated expression systems. Previous vector kits have contains vectors with an altered 3' UTR which destabilizes an arbitrary mRNA by about 20-fold in a wild-type (Smg+) genetic background. One means to then control the expression of resulting clones was by using as a host for injection a temperature sensitive mutant in the gene Smg-1. At the permissive temperature, these animals are Smg+ and the transgene is mostly off. At the non-permissive temperature the Smg system is essentially inactive and activity of the transgene is restored. An alternative method for controlling the expression of the same aberrantly-terminated transcripts involves the directed expression of coding regions for components of the SMG machinery. Working in a genetic background mutant for the critical SMG component (in this case SMG-5), corresponding expression constructs should be expressed at high levels. Targeted expression of the smg component can then be used for a modicum of tissue-specific suppression. SMG-5 has a relatively short coding region, hence its choice for this vector. Note that this construct encodes a functional gfp::smg-5 fusion protein. If needed, the gfp segment can be removed by cleavage with AgeI and religation. Alternatively several sites just upstream of the smg-5 coding region can be used to transfer just the smg-5 region to an alternative context.

L5784 myo3p-rfp2iV105AS197T
L5785 unc119p-rfp2iV105AS197T
L5832 myo3-iA-RFP2iV105A-l'
There are now better red fluorescent vectors than these
but we have not yet put them into C. elegans contexts.
For the interim, here are a trio of vectors based on the
dsRED reporter. They each have a pair of introns
inserted into the dsRED coding region, but we do not know
if this improves expression. The V105A mutation has been
reported in other systems to improve efficacy as a
reporter, which the S197T mutation has been shown to
provide a "historical perspective" on expression in that
newly synthesized protein is particularly green, while
older protein acquires a red fluorescence. I stress that
our success with dsRED constructs in C. elegans has been
limited (okay in differentiated tissues, very poor in
embryos and other dynamic stages).

The myo-3 promoter is expressed in body wall; unc-119 in
the nervous system.

L6420 sur5p-gfClacI5i-l'
L6637 ∏(polylinker)-gfClacI5i-l'
L6660 pes10-gfClacI5i-l'
L6709 pes10-yfplacI[6i]-l'
L6711 pes10-cfplacI[6i]-l'

L6705 pes10-tetR-yfp6i-l'
L6708 pes10-tetR-cfp6i-l'
Aidyl Gonzalez-Serrichio and Paul Sternberg introduced
the C. elegans world to the use of fluorescently-tagged
DNA binding proteins as very specific reporters. We have
produced modified versions of gfp-tagged lac Repressor
(lacI) which contains C. elegans introns (in gfp and
lacI) and which have restriction patterns that allow

facile insertion into standard C. elegans promoter
vectors.

To get this to work as a technique, you will need a
second construct with arrayed copies of the relevant
operator sequences (lac or tet). We obtained these
plasmids from Ivy Lau and David Sherrat (Oxford). The
various constructs above are expressed in diverse tissues
(sur-5p) or early somatic blastomeres (pes-10). L6637
carries just the polylinker sequence.

Thanks for information, etc: Luis Pantalena Filho, Jamie
Fleenor, Doug Koshland, Roger Tsien, Andrew Belmont,
Aidyl Gonzalez-Serrichio and Paul Sternberg.

Some promoters

We use a somewhat careful (although by no means perfect)
procedure to make promoter vectors. This involves taking
the coding region into the fist exon (but not including or
mutating the ATG) and placing a polylinker and 3'UTR
downstream. GFP is often present in the promoter vectors,
but in each case there is a bank of restriction enzymes
before the coding region starts. This allows you to insert
your own coding region without worrying about reading frame
or a few extra amino acids attached to the N terminus of
your protein. Although such vectors lack sequences
internal to the gene (introns, etc), they can in many cases
recapitulate the expected expression pattern (certainly not
in all cases, however as gene silencing and internal
signals tend to confound matters).

The following promoter vectors are included in the 2005
kit. Note that you can use them in several ways:
1. Replace the gfp with a coding region of choice.
2. Place a coding region of choice either upstream or
downstream of gfp (taking care to maintain the coding
region). Most of the following vectors have numerous
restriction sites for such manipulations.
3. Just inject these plasmids (the ones that encode gfp)
and enjoy the light show.

L3704 unc37-gfC-∏-1'
L5494 dpy30 5'-gfC-1'
L5767 sur5p-gfC-1'

These vectors express ~everywhere, which is to say that not all tissues or cells express in every transgenic animal, but lots of cells of many different types express within a population.

L5716 vit2-rde4-l'
Adult gut... note this is an rde-4 construct (expression rescues rde-4; in some cases the rde-4 rescue may be non-autonomous, which may be useful for certain applications.

L5711 unc25p-gfC-l'
L5876 unc30-gfC-l'
Both promoters express in inhibitory motor neurons (with some expression in other areas). unc-25 expression preceeds unc-30 expression.

L5840 rol6p-GFP3i-l'
L7315 col10p-gfC-u'
Temporally-controlled expression in subsets of the hypodermis.

Thanks to: James Meir, David Miller, David Hsu, Barbara Meyer, Meera Sundaram, John Yochem, Min Han, Stuart Kim, Pam Hoppe, Bob Waterston, Tom Blumenthal, Peg MacMorris, Yishi Jin, Steve McIntyre, John White, Nicol Thompson, Vector Ambros, Randall French, Craig Mello, Jim Kramer.

A few pointers on using the vectors.

I will assume that you have a specific of even vague experimental idea that you think might be facilitated by one or more vectors in the various "Fire Lab Kits". With several hundred vectors in the kits (of which we know some are more useful than others) the question of "where to look" for an appropriate reagent is certainly valid. Here is some general advice that mirrors how we often design experimental DNA constructs.

1. Browse through all of the various kit documentation once very superficially to see what types of vectors are available. You don't need to be paying any attention to restriction enzymes, specific vector names, details of the reporters, etc, just the general types of constructs that are there.

2. Study the documentation of the "cannonical" vector
L3691 in the 1997 vector kit. Thie vector has many of the
features that are shared by other vectors in the various
kits. Also some (but not all) of the restriction sites in
this vectors are shared by many of the vectors in the kits,
so that it becomes straightforward to use some of the same
sites over and over again to do certain exchanges of
promoters, reporters, localization tags and 3' ends.
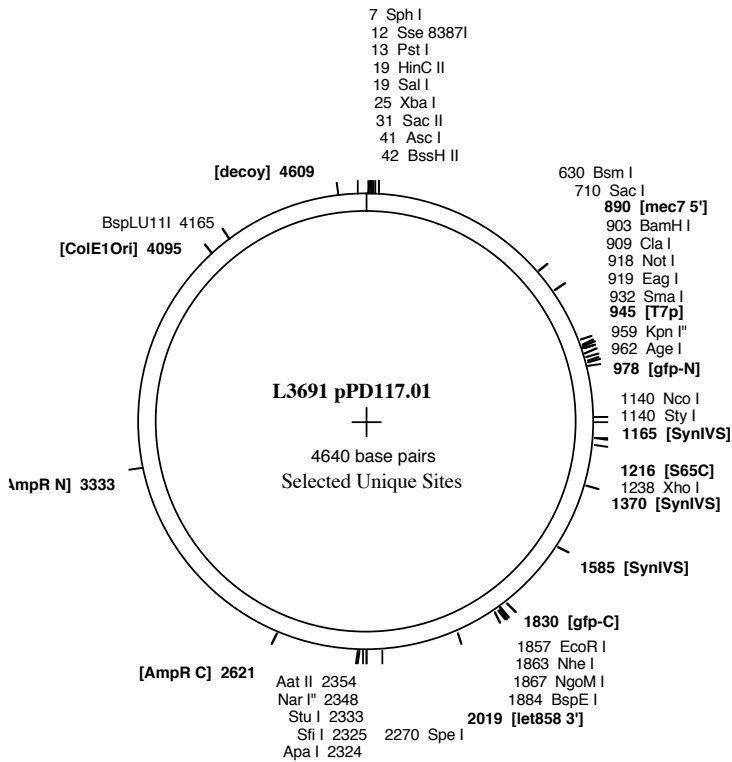
Note at the top of this map a multi-cloning site upstream
of the main promoter, a second multi cloning site (starting
in this case with BamHI) that begins after the promoter but
before the translational start, then (in this case), gfp as
a cannonical coding region, ready to be appended to or
replaced by your coding region of choice. At each end of
gfp are several unique restriction sites for placing a
coding region in phase with GFP. Last but not least, this
vector has a cannonical 3'UTR (from the 'generic' gene let-
858) and a final multi-cloning site after the let-858 3'
segment. Of course not all of the relevant sites are
shared by all vectors but many are certainly present in
large groups of vectors.

For this vector (or any other of vector), it is the
sequence (and not any pre-packaged map) that is the
definitive representation of what you are working with.

For the L3691 vector, there is a full description in the
1997 documentation of the relevant reading frames for the
gfp-upstream and gfp-downstream multi-cloning sites. The
maps in earlier documentation (and in this document) are
still imprecise for several reasons: sequences have been
slightly revised in some cases and the wonderful program
that is used to generate some of the maps (DNA strider)
uses the first base in a given feature definition as the
"position" of the feature. Since features are spread out
over groups of bases (or even hundreds of bases), this can
be very imprecise. A more precise description (still, sad
to say, not 100% accurate) can be found using the archaic
(but still useful) hypercard-based program "DNA Janitor"
which is available from the Fire Lab website. A compiled
version of the vector database with each of the features
that can be easily identified as well as unique restriction
sites can be found on our website as well.

If you have a vector sequence that you really want to know
about in terms of specific features, consider submitting
the sequence to BLAST (NCBI or Sanger) or BLAT (UCSC) to
identify individual C. elegans and other elements.

Not all manipulations need to be done with unique cutting
restriction sites. DNA Janitor and DNA Strider have easy
ways to find sites that are unique just within a region of
interest, allowing a fragment of interest to be swapped
between plasmids. We have also had considerable luck in
performing ligations which incorporate three fragments (as
long as each has non-matching overhanging ends).

How do you know you have what you want? Our first check is
always restriction digestion. Initially an enzyme or
enzyme pair that gives a fairly complex pattern (3-4 bands)
that tell you that you are on the correct track. Nice
adjuncts to this are a few extra digestions (principally
with enzymes used in the actual cloning, since there are
occaisions where the original restriction site is not
properly regenerated as expected). If you have used PCR
products in cloning, it is advisable to sequence the entire
product to avoid errors. It is imperative to sequence the
ends of the PCR product that are contributed by synthetic
primer sequences as such sequences are by far the most
error-prone. Certainly consider that sequences that must
encode proteins (ie where the reading frame is essential)
are more critical than are promoters, 3' ends or linkers.
We consider it essential to assemble for each plasmid that
is constructed a precise "mock" of the expected sequence.

This provides a clear means to check both sequence and restriction digestion patterns to ensure that the clone is as expected.

As an endnote to this section, I would stress the utility of distributing plasmids to the C. elegans community through a service such as Addgene. This doesn't cost anything to the distributing lab, is minimally costly to the using lab, and provides rapid accessibility to many experimental tools.