

Construct Design and Cloning Guide for Cas9-triggered homologous recombination

Written by Dan Dickinson (ddickins@live.unc.edu) and last updated May 2014.

Reference: Dickinson DJ, Ward JD, Reiner DJ and Goldstein B (2013). Engineering the *Caenorhabditis elegans* genome using Cas9-triggered homologous recombination. *Nature Methods* 10: 1028-1034. DOI: 10.1038/nmeth.2641.

Overview

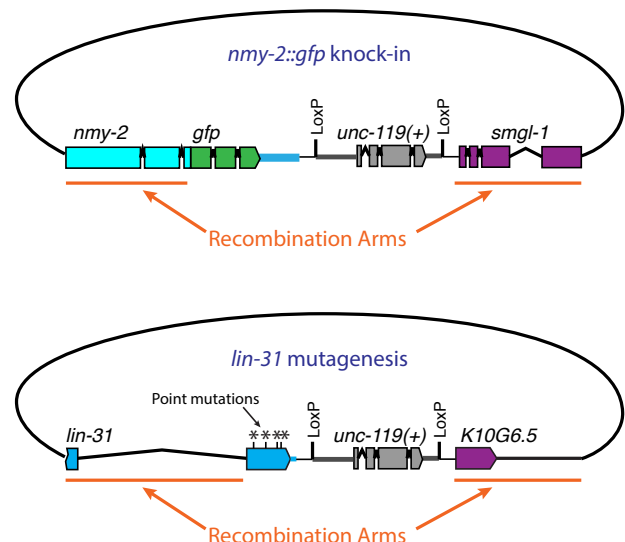
Preparing constructs for Cas9-triggered homologous recombination involves 3 steps:

- 1) Design and construct the homologous repair template. This plasmid contains the desired genome modifications, along with a selectable marker, flanked by unmodified recombination arms.
- 2) Choose the Cas9 target site, where the double strand break that triggers homologous recombination will occur.
- 3) Put the Cas9 targeting sequence into the Cas9-sgRNA plasmid.

1) Homologous repair template

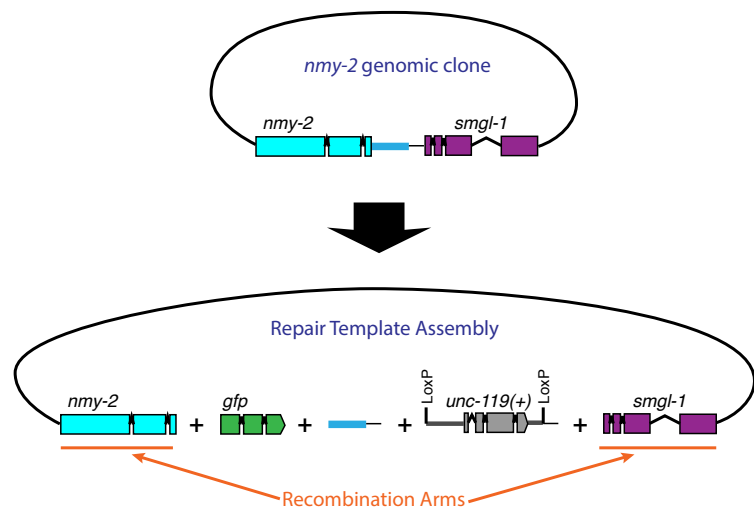
Two examples of homologous repair templates are shown at right. Each contains several pieces: a 5' recombination arm; some modified sequence that depends on the experiment; a selectable marker; and a 3' recombination arm. For more template design ideas, see the end of this document or

<http://wormcas9hr.weebly.com/construct-designs.html>.



General Considerations for Homologous Repair Template Construction

- Start by designing the desired genomic modification and make a sketch of the modified locus. Using a DNA sequence editor, generate a file that contains the sequence of the locus in its desired final state. The <http://wormcas9hr.weebly.com/construct-designs.html> page has example sequence files for several repair template designs.
- We mainly use *unc-119* selection in the lab, but other selectable markers could likely be substituted. If you wish to remove the selectable marker later, flank it with *LoxP* sites (5'-ATAACTTCGTATAGCATACATTATACGAAGTTAT-3').
- **IMPORTANT:** Ensure that the Cas9 target sequence (see below) is NOT present in the homologous repair template. Otherwise, Cas9 will cleave your repair template in addition to the genomic DNA, and recombination efficiency will suffer. If the targeting sequence is part of a coding region then you will need to introduce silent mutations to block cleavage. If possible, the simplest and most effective approach is to mutate the PAM (NGG motif), since this motif is absolutely required for cleavage of a substrate by Cas9. If a PAM mutation is not feasible, introduce at least 5–6 mutations in the target sequence. Mutations closer to the 3' end of the targeting sequence are more likely to prevent cleavage.
- The actual construction of the repair template is done in two steps (see the figure below):
 - PCR-amplify a 3–4 kb genomic region, centered on the desired modification, from N2 genomic DNA. We aim to have ~1.5 kb of unmodified sequence at each end of our homologous repair template to allow homologous recombination. Clone the genomic fragment into any suitable vector (we use the ZeroBlunt TOPO kit from Life Technologies).
 - Build the modifications you want into your cloned genomic fragment. We do this using Gibson Assembly Master Mix from NEB (<https://www.neb.com/products/e2611-gibson-assembly-master-mix>), which allows several DNA fragments to be assembled in a single step. Other cloning methods could be used instead but are likely to be slower and more labor-intensive.



Technical Tips for Gibson Assembly

- In brief, Gibson Assembly works by stitching together PCR fragments that have overlapping ends. You'll produce several PCR fragments for the sequences you want to insert into the genome, along with a large PCR product comprising the vector background and recombination arms (see the figure on the previous page).
- Design primers so that the fragments you want to assemble overlap by at least 30 bp. The NEB protocol says that 20 bp is enough, but we have had better luck with longer overlaps.
- In our experience, the biggest issue with Gibson Assembly is that one often sees a large number of colonies that are just parent vector. Treating all PCR fragments with DpnI, followed by gel extraction, has proven to be an effective way of eliminating this background.
- The efficiency of assembling the desired product appears to drop off exponentially with the number of fragments being assembled. In our hands, 2-3 fragment assemblies almost always work, 4-6 fragment assemblies are tricky, and more than 6 is almost impossible (note that the vector counts as a fragment). If you have trouble assembling your entire construct in one shot, try assembling a few pieces at a time to make larger pieces, and then PCR amplify and assemble the larger pieces.

2) Choose the Cas9 target site

- Identify a 200–250 bp region in which the Cas9 target site should be located. The Cas9 target region and the selectable marker should be at opposite ends of the homologous repair template, so that all desired genomic modifications are located between the Cas9 target site and the selectable marker. For example, for C-terminal GFP insertions, we use a Cas9 target sequence at or near the 3' end of the coding sequence, upstream of the 3'UTR, and place the *unc-119(+)* cassette downstream of the 3'UTR in the repair template. The desired modification (GFP insertion) lies between the Cas9 target site and the *unc-119(+)* marker. For point mutations, the target sequence should be located at or near the 5' end of the region containing the desired mutation(s), and the *unc-119(+)* cassette should be 3' of the desired mutations in the repair template. The desired mutations are between the Cas9 target site and the *unc-119(+)* marker.
- Submit this 200–250 bp genomic sequence to the Zhang lab's CRISPR design tool at <http://crispr.mit.edu>. Make sure you have selected *C. elegans* as the genome for checking specificity.

- The design tool returns a list of potential targeting sequences, ranked in order of predicted specificity. We usually choose the most specific target unless we have a reason to do otherwise. For the targeting sequence you choose, copy and paste the list of potential off-target sites into a Word document or Excel sheet and save this list for future reference.
- The design tool returns target sites of the form **5'N₂₀-NGG-3'**, where N is any base. Target sites in our system must begin with a G in order to be transcribed by the U6 promoter. If the target site you chose already begins with a G, you are done; go on to the next step. If not, lengthen the targeting sequence by adding nucleotides to the 5' end (the opposite end from the NGG) so that the targeting sequence does begin with a G. Your targeting sequence should then conform to the pattern **5'G-N₁₉₋₂₅-NGG-3'**. Note that in our original paper, we tested only targeting sequences of the form **5'G-N₁₉-NGG-3'**; however, new data from Feng Zhang's lab (Hsu et al. *Nature Biotechnology* 2013) suggests that addition of more nucleotides to the 5' end of the targeting sequence does not adversely affect Cas9 activity or specificity.

3) Add the targeting sequence to the Cas9-sgRNA construct

- We use NEB's Q5 Site-Directed Mutagenesis Kit to insert the targeting sequence into our Cas9-sgRNA construct (Addgene #47549). Use forward primer
- **5' -N₁₉₋₂₅GTTTTAGAGCTAGAAATAGCAAGT-3'**, where N₁₉₋₂₅ is replaced by the desired 19-25 bp targeting sequence, and reverse primer **5' -CAAGACATCTCGCAATAGG-3'**. Note that the initial G in the G-N₁₉₋₂₅ sequence is included in the reverse primer.
- **IMPORTANT:** Do not include the PAM (NGG motif) in your primers for the Cas9-sgRNA construct. The NGG motif must be present in the target DNA, but it is not part of the sgRNA.
- We use sequencing primer **GGTGTGAAATACCGCACAGA** to verify correct insertion of the targeting sequence.

Sample Homologous Repair Template Designs

The following pages show sample repair template designs, using the *his-72* locus as an example.

Sample sequence files for repair templates can be found at

<http://wormcas9hr.weebly.com/construct-designs>.

Thanks to Chris Higgins, Mike Nonet and Ari Pani for helpful discussions about construct designs.

The original design described in our publication

Brief description: Basic design for introducing modifications at the 3' end of a gene of interest.

Where Cas9 cuts: At or near the C-terminus of the gene of interest.

Where the selectable marker goes: In the intergenic region between the gene of interest and the next gene 3'.

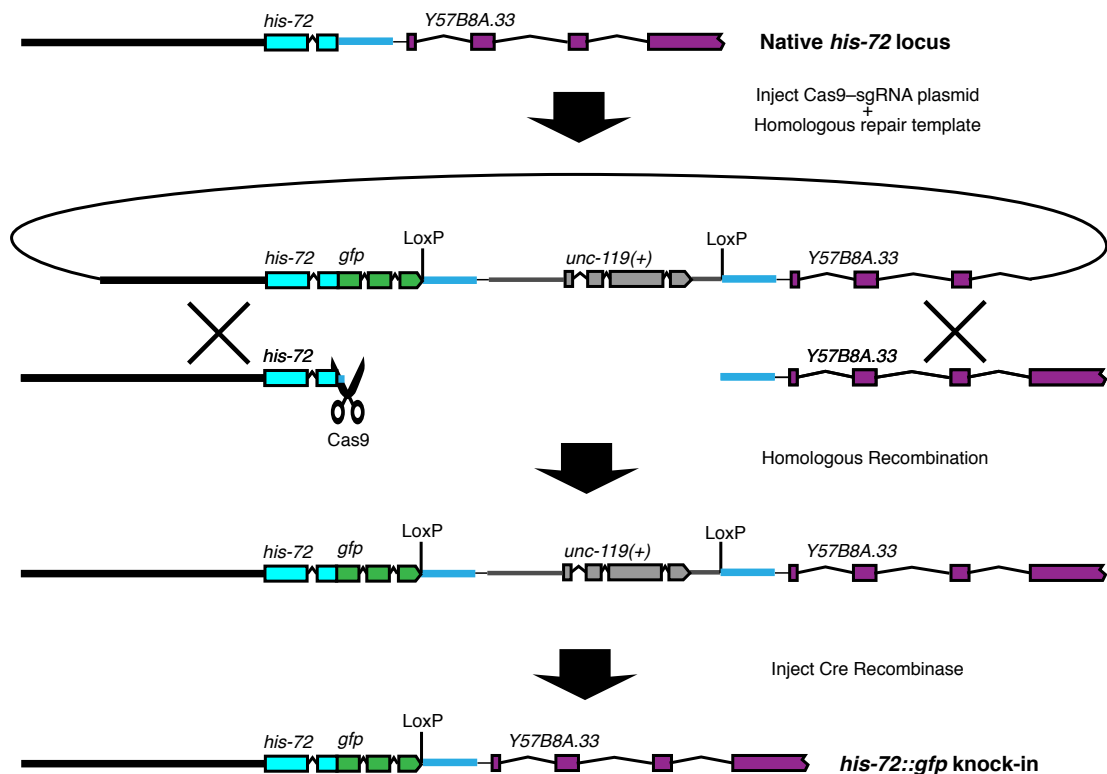
Best choice for: C-terminal tags or modifications

Advantages:

- Simplest strategy, applicable to many kinds of modifications
- Can look for fluorescence (or phenotype) immediately after isolating knock-ins, before injecting Cre.
- Thoroughly tested and very robust strategy; as of December 2013, 20/21 projects attempted in our lab have been successful (18/21 succeeded on the first try).

Disadvantages:

- Not well suited to N-terminal fusion tags or to modifications far from the 3' end of the gene.



N-terminal “broken GFP” tag with selectable marker in GFP intron

Brief description: The *unc-119(+)* selectable marker is placed within an intron of GFP. The example here is for an N-terminal tag, but this approach would also work for a C-terminal tag.

Where Cas9 cuts: At or near the N-terminus of the gene of interest.

Where the selectable marker goes: In a synthetic intron within the GFP coding sequence.

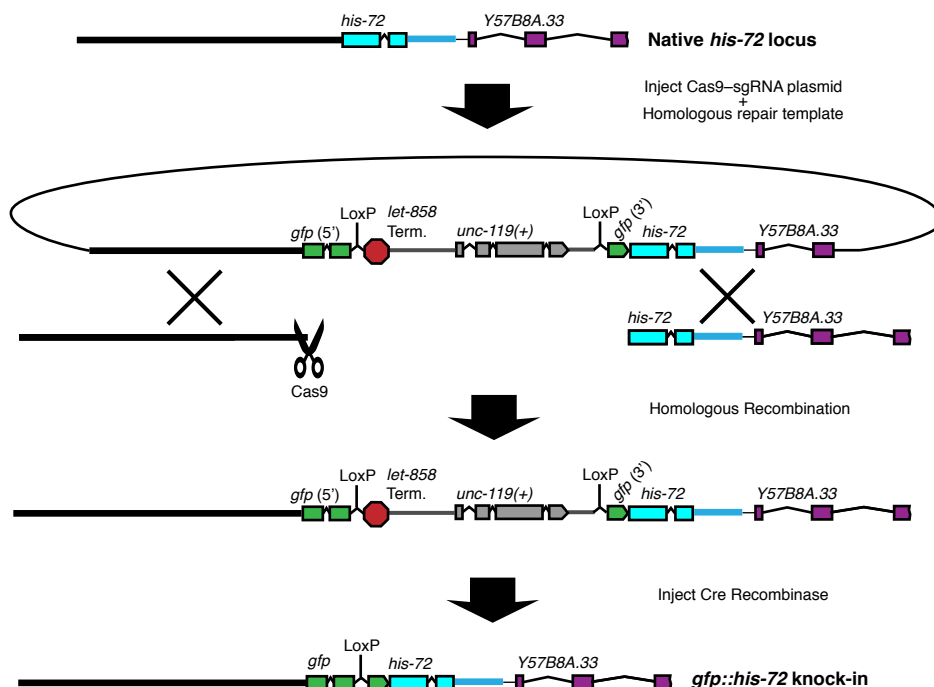
Best choice for: N-terminal tags on large, essential genes.

Advantages:

- This design is useful when the initial knock-in must be isolated as a heterozygote. Fluorescence is only observed after excision of the selectable marker, so after Cre injection one can screen by looking for GFP-positive animals.
- After selectable marker excision, no exogenous sequence is left behind except for insertion of the GFP.
- The broken GFP cassette containing the selectable marker can be amplified as a single PCR product, which simplifies cloning the repair template.
- We have used this strategy in our lab to N-terminally tag several endogenous genes. It works robustly as long as the gene of interest is expressed at high enough levels for GFP fluorescence to be visible.

Disadvantages:

- This strategy relies on being able to see GFP fluorescence from the knock-in on a dissecting microscope, so it may not work if your gene is expressed at low levels.
- GFP fluorescence is only observed after excision of the selectable marker. Although this makes screening much easier when starting from a heterozygous knock-in, it also means that two injection steps are required to generate a tagged protein. You won't know whether your knock-in is bright enough to be useful until after the second injection step.
- Because the initial insertion behaves as a null, recombinants will have to be isolated and maintained as heterozygotes.



N-terminal tag with selectable marker in synthetic intron

Brief description: GFP and a selectable marker are inserted at the 5' end of the gene of interest to make an N-terminal tag.

Where Cas9 cuts: At or near the N-terminus of the gene of interest.

Where the selectable marker goes: In a synthetic intron between the GFP and the N-terminus of the gene of interest.

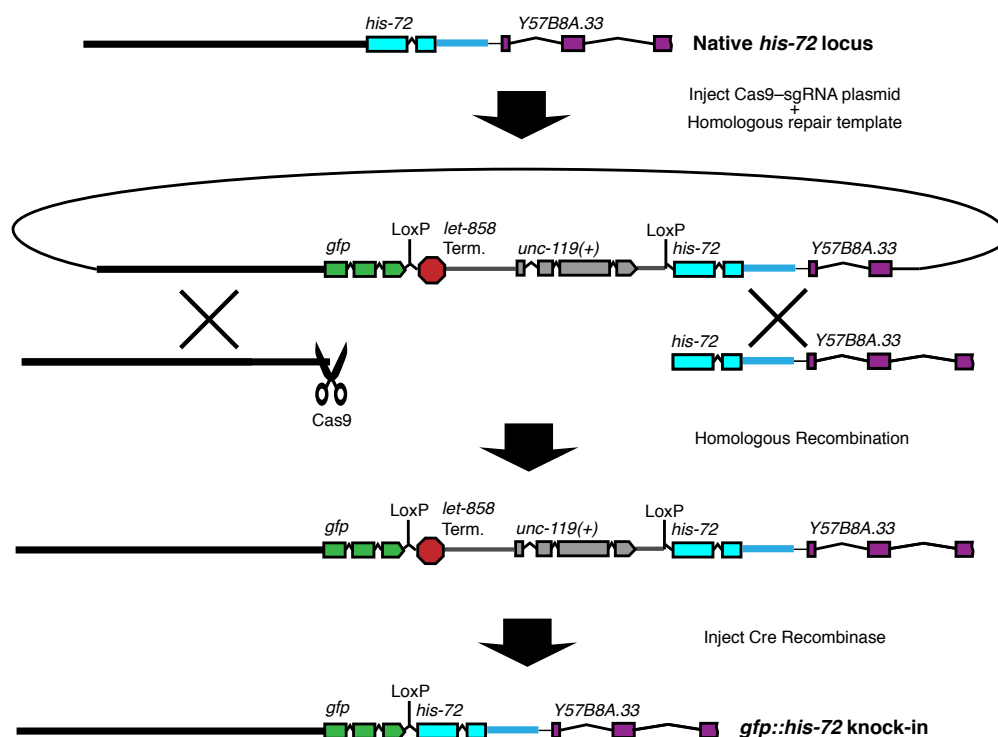
Best choice for: N-terminal tags on non-essential genes, especially when you are interested in assaying promoter activity.

Advantages:

- This design kills two birds with one stone: The initial insertion behaves as a null mutant expressing a promoter::GFP fusion (i.e., a gene replacement). This is converted into a translational fusion after Cre injection.
- After selectable marker excision, no exogenous sequence is left behind except for insertion of the GFP.

Disadvantages:

- Two injection steps are required to generate a tagged protein (because the first step produces a promoter fusion; see below for a design that generates a translational fusion in a single step).
- Because the initial insertion behaves as a null, recombinants will have to be isolated and maintained as heterozygotes if the gene of interest is essential. Cre-mediated excision of the selectable marker from a heterozygous strain can be very challenging. Therefore, for essential genes we recommend using one of the designs below instead.
- We are still in the process of testing this strategy in our lab. If you try it and find that it works (or doesn't work), please let us know.



Whole-gene replacement

Brief description: The genomic DNA is cut twice by Cas9, and the entire coding region is replaced.

Where Cas9 cuts: At or near both ends of the coding sequence.

Where the selectable marker goes: In the intergenic region between the gene of interest and the next gene 3'.

Best choice for: Point mutations or N-terminal tags on relatively small genes.

Advantages:

- This is the quickest and most straightforward way to get an N-terminal tag or to make point mutations far from the 3' end of the gene.
- Particularly convenient if you want to examine the localization or function of a GFP-tagged wild-type protein and also several mutants or variants, since the same basic design and Cas9-sgRNA plasmids can be used to produce the wild-type and mutant versions.

Disadvantages:

- Two Cas9-sgRNA plasmids are required. It is possible that needing two cuts to the genome will reduce the efficiency somewhat, although we don't anticipate this being a serious issue.
- This approach requires cloning the entire coding region of the gene, which could become prohibitive for large genes.
- We are still in the process of testing this strategy in our lab. If you try it and find that it works (or doesn't work), please let us know.

