Scalable and versatile genome editing using linear DNAs with micro-homology to Cas9 sites in *C. elegans*


*Dept of Molecular Biology and Genetics, Johns Hopkins University, School of Medicine, 725 N. Wolfe Street, Baltimore MD 21205, USA
§ National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

Running title: Scalable genome editing in *C. elegans*

Keywords: CRISPR, Cas9, Genome editing, homology-directed repair, short homology arms, *C. elegans*.

Corresponding author:
Geraldine Seydoux, Dept of Molecular Biology and Genetics, Johns Hopkins University, School of Medicine, PCTB706, 725 N. Wolfe Street, Baltimore MD 21205, USA
(443)-614-4623
gseydoux@jhmi.edu
ABSTRACT

Homology-directed repair (HDR) of double-stranded DNA breaks is a promising method for genome editing, but is thought to be less efficient than error-prone non-homologous end joining (NHEJ) in most cell types. We have investigated HDR of double-strand breaks induced by CRISPR-associated protein 9 (Cas9) in C. elegans. We find that HDR is very robust in the C. elegans germline. Linear repair templates with short (~30-60 bases) homology arms support the integration of base and gene-sized edits with high efficiency, bypassing the need for selection. Based on these findings, we developed a systematic method to mutate, tag or delete any gene in the C. elegans genome, without the use of co-integrated markers or long homology arms. We generated 23 unique edits at 11 genes, including premature stops, whole-gene deletions, and protein fusions to antigenic peptides and GFP. Whole-genome sequencing of five edited strains revealed the presence of passenger variants, but no mutations at predicted off-target sites. The method is scalable for multi-gene editing projects and could be applied to other animals with an accessible germline.
INTRODUCTION

The ultimate goal of genetic engineering is to rewrite the genome with precision and without extraneous modifications (e.g. marker insertion). The remarkable efficiency of CRISPR/Cas9 to induce double-strand breaks at defined locations has led to an explosion of new methods for genome engineering [(CARROLL 2014; SANDER and JOUNG 2014) for review]. *Streptococcus pyogenes* CRISPR-associated protein 9 (Cas9) is an endonuclease that is targeted to a specific DNA sequence by an associated guide RNA (GASIUNAS et al. 2012; JINEK et al. 2012). In animal models, expression of Cas9/sgRNA complexes in zygotes creates double-strand breaks that can be repaired by non-homologous end joining (NHEJ) or homology-dependent repair (HDR). NHEJ is an error-prone process that can create insertions, deletions or mutations at the cut site. HDR, in contrast, is a precise mechanism that uses sequences from a homologous donor molecule to repair the break. If the donor molecule carries edits flanked by sequences homologous to the targeted locus (“homology arms”), the edits will be integrated as part of the repair process. In many systems, HDR is thought to be less efficient than NHEJ, requiring high concentrations of donor molecules or long homology arms to stimulate recombination [(BEUMER et al. 2013; SANDER and JOUNG 2014) for review]. Single-strand oligodeoxynucleotides (ssODNs) can be injected at high concentration, but their relatively small size (~200bp or less) limits the types of edits that can be introduced. Studies in zebrafish embryos have also shown that ssODN-templated HDR is often imprecise, involving at least one error-prone NHEJ-like step (AUER and DEL BENE 2014). Plasmid
donors can accommodate gene-sized edits and longer homology arms, but require cloning and often selection to facilitate the recovery of rare recombinants. The selection marker is integrated along with the edit and must be removed in a subsequent step. The requirement for selection can be bypassed by providing high levels of Cas9 and sgRNA, using genome-integrated transgenes [as reported in Drosophila (PORT et al. 2014)], or using RNA injections [as reported in mice (YANG et al. 2013)]. For gene-sized edits, these approaches, however, are still thought to require the construction of plasmids with long homology arms, which limits scalability. Linear PCR fragments with short (<100bp) homology arms are easier to prepare and have been shown to support HDR in yeast and Drosophila tissue culture cells using selection for a co-integrated marker (BOTTCHER et al. 2014; DiCARLO et al. 2013). Our goal was to determine whether a similar approach could be developed in an animal, but with a high enough frequency to bypass the need for selection so as to generate marker-free, gene-sized edits in a single step.

HDR has been used extensively in C. elegans to introduce edits near double-strand breaks created by Mos excision, TALENs, and most recently CRISPR/Cas9 [(WAAIJERS and BOXEM 2014) for review]. As in other systems, both plasmid donors and ssODNs have been used as repair templates. Using CRISPR/Cas9, plasmid donors with long homology arms (1kb or more) have been used to insert GFP by co-selection for a linked marker or by direct screening for GFP expression (DICKINSON et al. 2013; Kim et al. 2014; TZUR et al. 2013). ssODNs have been used to introduce smaller, base-size edits near Cas9
cuts without selection (ZHANG et al. 2014) or by screening worms co-edited at a second locus (ARRIBERE et al. 2014; ZHANG and GLOTZER 2014). The Arribere et al. study reported that ssODN-templated HDR can be highly local and frequently gives rise to partial conversion events where edits >10 bases away from the cut site are not integrated (ARRIBERE et al. 2014). Whether short homology arms can support gene-size edits has not yet been reported.

In this study, we demonstrate that, in C. elegans, short homology arms flanking Cas9 sites support robust and precise HDR regardless of the size of the edit. Based on this finding, we developed a systematic and scalable method to create marker-free mutations, insertions and deletions at any locus. Unlike earlier approaches, our method uses the same 10-day protocol to mutate, tag or delete genes of interest, generates “clean” homozygous mutants with no co-integrated markers or footprints, and can be scaled up for systematic editing of multiple genes.

RESULTS

Insertion of premature stop codons and small protein tags using ssODNs

To test the robustness of HDR using short homology arms, we first established a systematic protocol to generate small insertions/deletions using ssODNs. We designed ssODNs and single guide RNAs (sgRNAs (JINEK et al. 2012) to target Cas9 to 8 different loci in 5 genes using genomic sequence information available on WormBase (sgRNA and ssODNs sequences used for
each experiment are listed in Tables S1, S2 and S3). The ssODNs contained homology arms (43 to 100 bases in length, average ~60 bases) flanking the insertion/deletion placed directly at the Cas9/sgRNA site (Figure 1A, Strategy 1, Figures S1A-B) or up to 27 bases away (Figure 1A, Strategy 2). In Strategy 2, we also included silent mutations in the ssODNs to prevent re-cutting (Figure 1A). We also embedded a restriction site in the insertion to help identification of the edits by PCR (Figures S1 and S2). We co-injected the ssODNs with a plasmid coding for Cas9 and the sgRNA (DICKINSON et al. 2013) (Figure 2 for protocol outline). We also included in the injection mix a plasmid coding for a visible marker. pRF4 is maintained extra-chromosomally and causes a roller phenotype (Rol) in the F1s that inherit it (MELLO et al. 1991). We used this marker to identify injected mothers that incorporated the DNAs in germ cell nuclei as evidenced by the appearance of Rol progeny in their broods, and screened only these “marked broods”. We screened all F1 progeny (Rol and non-Rol) laid within ~24-48 hours after gonad injection. The F1s (singly or in pools of 2) were allowed to lay eggs overnight and processed the next day by PCR and restriction digest. 4 days later, 8-16 F2s derived from positive F1s pools were processed in the same way to confirm germline transmission and isolate animals homozygous for the edit. Edits frequencies were calculated based on the number of positive PCR reactions divided by the total number of F1s screened (Table 1 and Table S1).

Edit frequencies ranged widely (Table 1) and did not appear to correlate with insert size. For example, using the same sgRNAs to target the nos-2 locus, we recovered insertion of a premature stop (12 bp) and a FLAG tag (66 bp) at
similar frequencies (1% and 2.4%) (Exp. 4 and 11, Table 1). In another experiment, we co-injected three ssODNs carrying insertions of two different sizes (42, 42 and 66 nucleotides, Exp. 8) and recovered all three edits. We observed, however, significant variability in sgRNA efficiency. In two cases where we used the same ssODN with two different sgRNAs, we obtained different edit frequencies [0.4 versus 7.3% (Exp. 3/2) and 0 versus 1.7 (Exp. 5/6) Table 1].

Sequencing of PCR amplified regions in F2 worms confirmed correct insertion for 26 out of 35 independent edits. Interestingly, all 9 incorrect edits were obtained using Strategy 2 (25 edits), where the insert is placed at a distance from the sgRNA/Cas9 site. The 9 incorrect edits contained mutations, deletions or insertions around the cut site and/or the insertion, and in one case (Exp. 6, Table 1) a single base change in the tag sequence. In contrast, 13 out of 13 edits obtained using Strategy 1 (insertion directly in the sgRNA site) were error-free. These observations suggest that HDR is more robust when the homology arms directly flank the cut site. Analyses of lines established for a subset of edits (Table 1) confirmed that the tags were expressed as protein fusions of the expected size (Figure 3A and 3B) and in the expected cells (Figure 3C).

To assess the potential for off-target effects, we performed whole-genome sequencing and variant analysis of five edited lines obtained from two different sgRNA/repair ssODN combinations (Exp. 1 and 12, Table 1, Figure S3A) plus two wild-type (N2) populations from which the edited lines were derived. No mutations were observed in the seven (Exp. 1, Table 1) or 13 (Exp.
predicted off-target sites with sequence similarity to their respective sgRNAs (Figures S3B and S3C). We also screened for extraneous insertions of the repair ssODNs or Cas9 and Rol plasmid sequences within the genomes of the edited strains. No insertion events beyond the targeted edit sites were detected. We conclude that ssODNs with short (>60 nt) homology arms can be used to create insertions (largest tested: 66 nt) at or near Cas9 sites without also causing random insertions in the genome or mutations in predicted off-target sites. However, we did observe a number of variants (mostly single nucleotide polymorphisms) unique to the edited strains (Figures S3D and S3E). Potential sources of such “passenger variants” in the edited lines are addressed below (see Discussion).

**Insertion of GFP using PCR fragments with short homology arms**

To test whether short homology arms can support the integration of larger edits, we used PCR fragments as repair templates, as is standard in yeast (HORECKA and DAVIS 2014), except that we did not use a selection marker. We first attempted to insert green fluorescent protein (GFP - 864 bases) using an sgRNA used in Exp. 6 to insert the small protein tag TetraCys (18 bases) at the C-terminus of K08F4.2. We amplified GFP with PCR primers designed to contain 59/59 nt homology arms that extended from the cut site (Strategy 1, Exp. 13, Table 1) or arms designed to position GFP precisely before the stop codon, 27 bases away from the cut site (Exp. 14, Table 1). In the latter, we included in the repair template mutations in the sgRNA pairing sequence to prevent re-cutting
(Strategy 2 in Figure 1A, Tables 1 and S3). We screened F1s laid over a 48-hour period after injection, in pools of eight worms, and identified edits at an estimated minimum frequency of 4% (Exp. 13) and 0.9% (Exp. 14). Remarkably, these frequencies were comparable to the frequency (1.7%) observed for the insertion of the much smaller TetraCys tag using the same sgRNA.

To test whether this approach is robust and can be used at other loci, we designed sgRNAs to target the C-termini of seven other genes. Where possible, we used sgRNA sites that overlapped the stop codon (Strategy 1 in Figure 1A). Alternatively, we chose sgRNAs close (<30 bases) to the stop codon and used silent mutations to prevent re-cutting (Strategy 2 in Figure 1A). We obtained GFP insertions for five of the seven genes attempted at estimated minimum frequencies ranging from 0.4 to 1.4% (Exp. 15-21, Table 1). 24 F2s derived from positive F1s were screened singly by PCR or by visual inspection for GFP fluorescence to recover and propagate the edits. All derived strains (each started from a single homozygous F2/3) showed stable GFP expression in the expected pattern (Figure 3D, Table S5).

In one of the two experiments that failed, only one brood (84 F1s) was screened, which may have been too few (Exp. 21). In the other failed experiment, a large number of F1s were screened (402 F1s), raising the possibility that the sgRNA may have been inefficient (Exp. 20). We found that it is possible to use two sgRNAs with overlapping recognition sites in the same experiment (Exp. 15, 17 and 19), which could help avoid low efficiency sgRNAs.
We conclude that small homology arms are sufficient to support GFP insertion at most loci, provided that an efficient sgRNA site can be identified within ~30 bases of the desired site of insertion. We have not yet tested whether insertions could be created at an even greater distance from the sgRNA site.

Precise gene-sized deletions using ssODNs

Using TALENs, Lo et al., 2013 reported the isolation of small precise deletions (<100 bases) templated by ssODNs. To test whether ssODNs could also be used to create gene-size deletions, we attempted to delete an entire ORF, using an ssODN with 67 and 57 base-homology arms that fused the START and STOP codons of K08F4.2. We co-injected this ssODN with four sgRNAs with cut sites at the 5’ and 3’ ends of the K08F4.2 ORF (Exp. 22, Table 1 and Figure S1). Unlike in the insertion experiments described above, both homology arms of the ssODN were separated from the cut sites by 10 to 31 bases (Strategy 3 in Figure 1A, Figure S1C). We obtained 22 deletions (frequency 3.3%). The deletions, however, were imprecise as evidenced by their varied sizes and sequencing results (Table 1), suggestive of NHEJ repair. NHEJ repair of two cuts separated by 53 bp was reported previously (Cho et al. 2013). We confirmed that large deletions can be created directly by NHEJ alone, using sgRNA pairs targeting the 5’ and 3’ ends of mbk-2 and the swan-1/swan-2 operon (Exp. 23 and 24, Table 1). Sequencing of the deletion breakpoints revealed small insertion/deletions consistent with error-prone NHEJ (data not shown).
To obtain a deletion with a precise fusion point, we modified the design of the ssODN targeting K08F4.2 to contain 1) 80-51 base homology arms that precisely flanked the sgRNA cut sites and 2) a restriction site inserted at the cut site (Strategy 4 in Figure 1A, Figure S1D). This time, we obtained 22 correctly sized-deletions, 20 of which contained the edited restriction site (3.8%, Exp. 25, Table 1). We repeated the same experiment omitting the sgRNA on one side of the deletion, and failed to obtain any deletions (Exp. 26, Table 1 and Figure S1E). We conclude that large deletions with precise breakpoints (including insertions at the fusion point) can be obtained using two sgRNAs that targets the ends of the deletion and one ssODN with homology arms that closely flank the sgRNA sites (Strategy 4).

**Edit frequency is highest in marked broods, fluctuates from brood to brood, and does not increase with longer homology arms.**

In all the experiments described above, we only screened the broods of injected mothers that segregated Roller animals (“marked broods”) for genome edits. Because the Rol marker is on a different plasmid than Cas9/sgRNA, we investigated whether non-marked broods might also contain edits. We repeated Exp. 25 to generate a precise deletion in K08F4.2 and screened F1s from 16 marked and 16 non-marked broods in pools of 1-8. We identified 12 deletions from 295 F1s from marked broods (est. 4% efficiency) and 0 deletions from 332 F1s from non-marked broods. We conclude that screening only marked broods enriches for broods with edits, as expected.
To determine the optimal length for homology arms, we repeated Exp. 13 using the same PCR fragment (GFP flanked by 59/59 bp homology arms) and with new PCR fragments with shorter and longer arms (Figure 4). We separated the F1s laid in the first and second 24 hours after injection and screened each F1 directly for GFP expression by live fluorescence microscopy. We found most GFP+ edits among F1s laid on the second day after injections. 33/33 bp homology arms gave the highest frequency of GFP+ edits and 15/19 bp arms gave the lowest (12.8% and 0.1% on the second day). Longer arms did not increase, and in fact appeared to decrease, edit frequency (Figure 4). Overall edit frequency for 59/59 bp homology arms (4.6%) was comparable to that found by PCR screening (4%, Table 1, Exp. 13), strongly suggesting that all edits express GFP already in the first generation.

Edit frequency varied greatly between broods. For example, on the second day after injection using 33/33 homology arms, ~40% of broods yielded no edits and 20% of broods gave 20-60% edits (“jackpot broods”, Table S4). We also observed jackpot broods when using ssODNs (example shown in Figure S2), as also reported by Arribere et al., 2014. In the one experiment using ssODNs where we separated F1s laid on the first and second day (Exp. 2, Table 1), we did not notice a difference in edit frequency between the two egg-laying periods (data not shown).

We conclude that GFP edits can be obtained with homology arms as short as 33 bp, that longer arms do not increase edits frequencies, and that edits are distributed unevenly between broods.
DISCUSSION

In this study, we demonstrate that linear templates with short homology arms (~30-60 bp) support robust integration of both small and large (gene-size) edits in *C. elegans*. The efficiency of HDR (typical range: 0.4 - 7% edits across all F1s, and as high as 60% in jackpot broods) is high enough to bypass the need for selection and is not affected significantly by edit size. HDR efficiency is affected, however, by the distance between the homology arms and the cut site(s). We obtained best results when both homology arms were homologous to sequences immediately flanking the cut site (Strategy 1). To create a small deletion or an insertion at a distance from the cut site, it is possible to use repair templates with one homology arm flanking the cut and the other at a distance, but this approach yielded more imprecise edits (Strategy 2, maximum distance tested 27 bp). This strategy was also used by Lo et al., 2013 to create a 77 bp deletion using a single, TALEN-induced cut. We were not able, however, to generate a 1.6 kb deletion using this single-cut strategy, possibly because the distance was too large between the cut site and the distant homology arm (Exp. 26). We were able to make such a deletion, however, using a two-cut strategy (Strategy 4, Exp. 25). When using two cuts, we found again that it is critical that the homology arms in the ssODN extend close to the cut sites. Separation of both homology arms from the cut sites by as few as 10-31 bp favors an error-prone NHEJ-like mechanism where the repair template is not used (Exp. 22). In fact, it is possible to create large imprecise deletions by NHEJ using two cuts and no repair template (Strategy 3 – largest deletion attempted: 6kb). We conclude
that, for precise edits, it is best to design repair templates with homology arms that extend as close to the cut site(s) as possible at least on one side of the cut. Since it is possible to use multiple cut sites and to simultaneously insert and delete sequences, this requirement still allows for many different types of edits.

The edit efficiencies we report here are within the range reported for similar experiments in *Drosophila* and zebrafish using ssODNs or PCR fragments (and plasmids) with long (>800 bp) homology arms (Auer and Del Bene 2014) (Beumer and Carroll 2014). In zebrafish embryos, the majority of repair events are imprecise and involve an NHEJ-like mechanism on at least one side of the edit (Auer and Del Bene 2014). The practice in *C. elegans* of injecting the repair template and Cas9/sgRNA plasmid directly in the meiotic (oogenic) germline syncytium may help favor HDR. It is likely that most edits were generated in the oogenic germline of the injected hermaphrodites, since most edits were heterozygous in the F1 generation and were transmitted to the F2 generation in the expected 1:2:1 ratio. We speculate that high rates of HDR could also be obtained in other animals by targeting meiotic germ cells instead of embryonic cells where NHEJ dominates (Auer and Del Bene 2014).

We sequenced the genomes of five edited lines made using two different sgRNA/ssODN combinations and observed no mutations at predicted off-target sites, as also reported by (Chiu et al. 2013). We also observed no insertions of the ssODNs outside of the targeted loci, confirming specificity. However, we did observe several variants (from one to 18 per strain, predominantly SNPs) in the edited strains that were not detected in the wild-type populations (N2) used for
injections. The source of these variants is unclear. One possibility is that the
variants observed in the edited strains were derived from rare alleles present in
the wild-type population that became fixed during edit isolation. Each edited line
was founded by a single hermaphrodite, and underwent at least one additional
round of clonal isolation. These putative rare alleles may have been lost from the
wild-type strains during propagation, or remained present but at a frequency
below the level of detection. Both the wild-type and edited strains were passaged
several times between the day of injection and the time of harvest for sequencing,
providing an opportunity for genetic drift. The proposed model of fixation of rare
parental variants is most consistent with the data. We observed the same
variants in independent edited lines obtained from the same injected
hermaphrodite (Figure S3: AP-1 and AP-3), and different variants in edited lines
derived from different injected hermaphrodites (Figure S3: AP-1/AP-3 vs. AP-2 vs.
YW-1/YW-2). Alternatively or additionally, some of the variants could have been
caused by a sequence non-specific mutagenic effect of Cas9 and/or the ssODNs.
Although further analyses will be needed to distinguish between these
possibilities, our findings so far indicate that 1) passenger mutations can become
fixed in the edited lines and that 2) it is advisable to isolate at least two
independent edits (from different injected mothers) to avoid possible background
effects.
Based on these observations, we created a simple method for genome editing in *C. elegans*. Our approach differs in several respects from previous methods [(WAAIJERS and BOXEM 2014)]. First, our method is versatile, allowing users to follow 4 different strategies (Figure 1A) and the same protocol (Figure 2) to tag, mutate or delete their gene of interest. Second, our method does not use co-integrated markers, and thus generates marker-free edits and does not require a specific genetic background or time-consuming selection schemes. Because injected DNAs form stable extrachromosomal arrays in *C. elegans* (STINCHCOMB et al. 1985), selection-based approaches must also include counter-selection against such arrays (CHEN et al. 2013). The selection marker integrates with the edit and must be removed in an additional step using flanking recombination sites, which can leave a footprint (DICKINSON et al. 2013). Third, by relying on direct screening of F1s, edits are identified five days after injection, compared to two weeks or more when using selection markers, although PCR screening requires more hands-on time. Fourth, genome edits are identified in heterozygous animals in the first generation after editing, ensuring the recovery of both viable and lethal alleles (asterisks in Table 1). Finally, the use of short homology arms does not reduce edit frequency even for gene-sized insertions like GFP. PCR fragments with longer arms (up to 500 bp) do not exhibit higher edit rates (Figure 4). Also, using plasmids with 1kb homology arms and the same marker plasmid we used here (pRF4-Rol), (KIM et al. 2014) reported 1-10% GFP edits among Rol F1s. Similarly, using 60bp homology arms, we obtained 3-11% GFP edits among Rol F1s (Table S1). Importantly, we also found significant
numbers of GFP edits among unmarked F1s, especially in the second day after injection (9%-13%, Figure 4 and Table S4). Unmarked F1s are more numerous than marked F1s and therefore require fewer injections to generate. The ability to recover edits in unmarked F1s also reduces exposure to the Cas9/sgRNA plasmid, which likely is co-inherited with the pRF4 marker plasmid in the Rol F1s (Mello et al. 1991).

The use of short homology arms also offers several technical advantages. ssODNs and PCR fragments require no cloning, making our approach scalable. We successfully designed repair templates targeting 17 unique sequences in 11 different genes using genome sequence information available in WormBase, suggesting that, even when relying on micro-homology, polymorphisms that could interfere with HDR are not an issue at least in the common C. elegans lab strain Bristol (N2). The oligo-based design of the templates also facilitates the incorporation of helpful modifications in the homology arms and/or the insertion. These include restriction sites to facilitate screening and mutations in sgRNA sites to prevent re-cutting after editing. This is particularly useful when using multiple sgRNAs to target a single site to reduce the chance of choosing inefficient sgRNAs, since it is advisable to mutate each site in the repair template to prevent re-cutting (Kim et al. 2014). Short homology arms also greatly simplify PCR screening by making it possible to use primers close to the insert without risking amplification of non-integrated templates that might persist in F1s. Finally, when making GFP fusions, short homology arms can avoid the inclusion of
promoter sequences, which could drive expression directly from the template. In this way, GFP edits can be identified directly by visual inspection of F1 animals.

When using PCR to identify edits, the most labor-intensive part of the protocol is the handling and processing of the F1s. We used two strategies to minimize this step. First, we used a dominant episomal marker (pRF4) to identify successfully injected mothers and screened only their progeny. This approach cut by ~50% the number of F1s that need to be screened. Second, we processed the F1s in pools. We used pools of two F1s for micro-edits that require restriction digestion, and pools of eight F1s for larger edits that can be detected directly by PCR. 100 to 200 pools can be processed in two days and edits are easily isolated from the pool in the next generation by individual screening of 8 to 24 F2s. Our observations indicate that edits are distributed highly unevenly among broods, with ~15% of injected hermaphrodites generating broods with 20% or more edited F1s (“jackpot broods”, Table S4). Identification of these jackpot broods before F1 screening would substantially reduce workload. Recently developed Co-CRISPR methods should make it possible to identify jackpot broods by selecting for broods containing edits at a second locus with a visible phenotype (ARRIBERE et al. 2014; KIM et al. 2014; ZHANG and GLOTZER 2014). Particularly promising is the recent report of Arribere et al., 2014 who showed that single-base edits can be recovered in as high as 80% of F1s selected for editing at a second locus with a dominant phenotype.

In summary, we have found that short homology arms stimulate HDR at high enough efficiency in the C. elegans germline to create marker-free, gene-
sized edits in a single step. The scalability of our method should make it possible to produce precise ORF deletions and reporter (e.g. GFP) fusions for every gene in *C. elegans*, a first for an animal model. There is no reason, *a priori*, to think that a similar approach could not be applied to other organisms with an accessible germline, thereby expanding the versatility and applicability of this exciting new era of genome engineering.

**MATERIALS AND METHODS**

**Protocol**

We provide a detailed protocol in supplementary materials.

**Whole-genome sequencing**

Libraries were constructed on the Mondrian SP+ (Nugen) and sequenced on the HiSeq 2500 (Illumina). For each library, a minimum of $4.4 \times 10^7$ 50bp reads (22-fold genome coverage) were aligned to the reference genome WS220 (www.wormbase.org) using BFAST software (HOMER *et al.* 2009). Potential off-target sites were predicted using CRISPR Design Tool (crispr.mit.edu). Mutation screening was by visual inspection of the aligned data at predicted sites and flanking sequences (+/-35bp). Potential insertion mutations were detected using split-end alignment (SMITH 2011).
**Western Blotting**

Transgenic worms were lysed by freeze-thaw lysis in 1XM9 with 2.5% SDS. For embryonic lysates, 50ul packed embryos were resuspend in lysis buffer (2% SDS, 10% glycerol, 65mM Tris-HCl pH7.5, protease inhibitors). Embryos were lysed using a Misonix Sonicator 3000 with total of 30 seconds sonication (15 seconds on, 45 seconds off at power level 2). Samples were run on a polyacrylamide gel and transferred overnight to a PVDF membrane. The membrane was blocked with 5% nonfat milk in PBS-Tween (Tw), washed and probed with the following antibodies: anti-V5 HRP (R961-25, Invitrogen, 1/1000) for 2 hours at RT, and anti-FLAG HRP (2044-S, Cell Signaling Technology, 1/1000).

**Immunofluorescence**

For staining, embryos were freeze-cracked on poly-L-lysine-coated slides and fixed in −20°C methanol for 15 min and −20°C acetone for 10 min. Samples were blocked in PBS-Triton-BSA for 30 minutes, and stained with anti-FLAG M2 (1/500, Sigma F1804), and anti-PGL-1 (K76, 1/15) overnight at 4°C. Primary antibody was detected using appropriate fluorescent secondary antibodies, mounted and imaged. N2 worms were used as negative control.

**In gel TetraCys Tag detection**

Transgenic worms were transferred in PBS containing protease inhibitors and freeze-thawed to lyse in NP40 buffer containing protease inhibitors. The lysate
was processed using the Lumio Green detection kit (Invitrogen) following manufacturer instructions, run on polyacrylamide gel and imaged.

ACKNOWLEDGMENTS: We thank Craig Mello for sharing results prior to publication, Mario de Bono for reagents, Bram Lambrus for sgRNA cloning assistance, Dominique Rasoloson for strain archival, and Andrew Folkmann and Dan Dickinson for discussion and comments on the manuscript. This work was supported by NIH grant R01HD37047, by the Intramural Research Program of the NIH, The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). G. Seydoux is an investigator of the Howard Hughes Medical Institute.

References


A
1. Insertion in sgRNA recognition sequence

\[\text{HDR}\]

2. Insertion outside sgRNA recognition sequence

\[\text{HDR}\]

3. Imprecise deletion between two outs

\[\text{NHEJ}\]

Repair template not integrated, imprecise deletion

4. Precise deletion/insertion between two outs

\[\text{RE}\]

Repair template integrated, precise deletion

B
1. Screen for small insertion or point mutation

\[\text{RE}\]

IPCR

\[\text{RE}\]

RE digest

2. Screen for large insertion

\[\text{RE}\]

IPCR

3. Screen for deletion

\[\text{IPCR}\]

4. Screen for precise deletion

\[\text{IPCR}\]

RE digest
Figure 1  Strategies for HDR and PCR screening

(A) Donor design strategies:

Black lines represent genomic DNA. Red lines represent inserted sequence and blue lines are homology arms on template DNA. Scissors denote the sgRNA recognition sequence (pairing site + PAM). Red crosses indicate mutations in the repair template in sequences corresponding to the sgRNA recognition sequence. HDR – homology-direct repair. NHEJ – non-homologous end-joining.

See text for strategy descriptions. Note that for Strategies 1 and 2, only insertions are shown in the schematics, but it is possible to simultaneously insert and delete sequences as described in Figure S1. In Strategy 3, the homology arms in the donor template do not extend to the cuts and as a result, NHEJ is preferred over HDR and the donor template is not used. This strategy can be used without a template to generate an imprecise deletion between two distant sgRNA cut sites.

(B) PCR screening strategies:


1. Amplify region with primers flanking the insertion and digest with a restriction enzyme whose site is embedded in the insertion. See the protocol for sequences coding for antigenic tags that contain convenient restriction sites. This strategy works best with small F1 pools (2 F1s).
2. Amplify with a primer specific for the insertion and a primer specific for the gene of interest. This approach could also be used for small insertions, but in practice, when using ssODNs to generate small insertions, we have observed non-specific products likely due to perduring ssODNs in the F1s.

3. Amplify with primers flanking the deletion. This approach is suitable for pools of 8 F1s since the smaller deletion band is favored in the PCR.

4. Amplify with primers flanking the deletion and digest with restriction enzyme whose site is embedded at the junction. This digest is used to distinguish HDR events from NHEJ events.
Day 1: Inject N2 hermaphrodites and clone each to separate plate.

Day 4: Identify broods with Rol worms, and clone Rol and non-Rol F1 worms in pools.
Day 5: Process pools to identify edits by PCR.

Day 8: Clone F-2s from positive pools.
Day 9: Process F2s to identify homozygous edits.
Figure 2  Experimental outline

On Day 1, ~40 young adult hermaphrodites are injected and allowed to self-fertilize (one injected hermaphrodite per plate). The injections deliver the Cas9/sgRNA and pRF4 plasmids directly into the syncytial oogenic germline (sperm are formed at an earlier stage of development and stored away from the site of injection). The Cas9/sgRNA plasmid is presumed to be expressed shortly after injection in oogenic germ cells, as it contains promoters predicted to be active in that tissue (Dickinson et al. 2013). The pRF4 plasmid encodes a mutated collagen which, when expressed in F1s, causes the worms to roll (Rol phenotype). This marker is used to identify broods derived from mothers that were successfully injected as evidenced by their transmission of the pRF4 plasmid to the next generation. On Day 4 after injection, Rol and non-Rol F1s from plates with Rol animals (“marked broods”) are screened directly for GFP expression or transferred to new plates in pools of 2 or 8 for PCR screening. On Day 5 (after F1s have laid eggs on plates), F1 pools are screened by PCR for the desired genome edits (light red). On Day 8, 8-24 F2s from each positive F1 pool are transferred singly to new plates and allowed to self-fertilize. On day 9, F2s are screened by PCR for homozygous genome edits (dark red).
Figure 3 Expression of tagged proteins

(A) Western blot using whole worm or embryonic lysates and commercially available anti-V5 and anti-FLAG antibodies.

(B) TC::K08F4.2 detection using Lumio Green. Arrows shows a band of the expected size for two independent TetraCys (TC) edits of K08F4.2; the band is absent in N2 (wild-type) animals and also in a third independent edit of K08F4.2 that contains an inactivating mutation in the TC tag (C to Y).

(C) Immunofluorescence image of a fixed embryo stained with DAPI and anti-FLAG antibody. FLAG-NOS-2 is present in the two primordial germ cells, Z2 and Z3, reflecting the wild type distribution of NOS-2 at this stage of development.

(D) Fluorescence pictures of live embryos (deps-1, lin-15), germline (mex-6, fbf-2) and whole animals (mes-2, K08F4.2) expressing the indicated GFP fusion proteins. Dashed lines outline the gonad boundary.
Figure 4  Effect of arm length on edit frequency for a GFP PCR fragment

Graph showing the percent of GFP+ F1s among marked broods derived from hermaphrodites injected with Cas9/sgRNA APs5 as in Exp. 13, except that the GFP PCR fragment contained homology arms of different lengths. Exact sizes were 15/19, 33/33, 59/59, 99/101 and 504/502 bp. The number of GFP+/ total F1s scored is indicated on each bar and the number of broods scored for each experiment are shown in parentheses below. The data shown here for the ~30 bp and ~60 bp homology arms are also shown broken down by broods in Table S4.
<table>
<thead>
<tr>
<th>Exp</th>
<th>Strategy (Fig.1A)</th>
<th>Edit</th>
<th>Gene</th>
<th>sgRNAs</th>
<th># P0</th>
<th># PCR</th>
<th># F1s</th>
<th># edits</th>
<th>% edits</th>
<th># of precise edits / # of sequenced</th>
<th>Verification of expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Insertion - STOP</td>
<td>K08F4.2</td>
<td>APa4-2</td>
<td>18</td>
<td>277</td>
<td>554</td>
<td>41</td>
<td>7.4</td>
<td>5/5</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Insertion/deletion - STOP</td>
<td>K08F4.2</td>
<td>APs1</td>
<td>3</td>
<td>95</td>
<td>178</td>
<td>13</td>
<td>7.3</td>
<td>2/2</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Insertion/deletion - STOP</td>
<td>K08F4.2</td>
<td>APs4</td>
<td>6</td>
<td>141</td>
<td>267</td>
<td>1</td>
<td>0.4</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Insertion/deletion - STOP</td>
<td>nos-2</td>
<td>SL225; SL232</td>
<td>3</td>
<td>194</td>
<td>267</td>
<td>3</td>
<td>1.1</td>
<td>3/3</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>TetraCys at C-term</td>
<td>K08F4.2</td>
<td>APs6</td>
<td>11</td>
<td>188</td>
<td>376</td>
<td>0</td>
<td>0.0</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>TetraCys at C-term</td>
<td>K08F4.2</td>
<td>APs5</td>
<td>5</td>
<td>143</td>
<td>286</td>
<td>5</td>
<td>1.7</td>
<td>2/5</td>
<td>In gel detection</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Myc, 3X Flag, V5, His, HA at C-term</td>
<td>K08F4.2</td>
<td>APs5</td>
<td>9</td>
<td>179</td>
<td>316</td>
<td>3</td>
<td>0.9</td>
<td>1/3</td>
<td>WB/IF</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>V5, 3X Flag, OLLAS at N-term</td>
<td>mex-5</td>
<td>JS129; JS130</td>
<td>11</td>
<td>235</td>
<td>412</td>
<td>12</td>
<td>2.9</td>
<td>3/4</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Myc, V5 at N-term</td>
<td>swan-1</td>
<td>CSD35</td>
<td>1</td>
<td>19</td>
<td>19</td>
<td>5</td>
<td>26.3</td>
<td>3/3</td>
<td>WB/IF</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3X Flag at C-term</td>
<td>pgl-1</td>
<td>TL001; TL002</td>
<td>6</td>
<td>134</td>
<td>134</td>
<td>1</td>
<td>0.8</td>
<td>0/1</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>3X Flag at N-term</td>
<td>nos-2</td>
<td>SL225; SL232</td>
<td>11</td>
<td>370</td>
<td>781</td>
<td>19</td>
<td>2.4</td>
<td>6/8</td>
<td>WB/IF</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>V5 at N-term</td>
<td>mbk-2</td>
<td>HS516</td>
<td>2</td>
<td>48</td>
<td>48</td>
<td>1</td>
<td>2.1</td>
<td>1/1</td>
<td>WB/IF</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>GFP at sgRNA site near C-term</td>
<td>K08F4.2</td>
<td>APs5</td>
<td>12</td>
<td>108</td>
<td>741</td>
<td>30</td>
<td>4.0</td>
<td>4/4</td>
<td>GFP expression</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>GFP at C-term</td>
<td>K08F4.2</td>
<td>APs5</td>
<td>20</td>
<td>179</td>
<td>1134</td>
<td>10</td>
<td>0.9</td>
<td>-</td>
<td>GFP expression</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>GFP at C-term</td>
<td>fbf-2</td>
<td>APs12; APa13</td>
<td>17</td>
<td>134</td>
<td>832</td>
<td>3</td>
<td>0.4</td>
<td>-</td>
<td>GFP expression</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>GFP at C-term</td>
<td>mes-2</td>
<td>sg16</td>
<td>18</td>
<td>150</td>
<td>738</td>
<td>5</td>
<td>0.7</td>
<td>-</td>
<td>GFP expression</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>GFP at C-term</td>
<td>lin-15b</td>
<td>sg23; sg25</td>
<td>5</td>
<td>51</td>
<td>336</td>
<td>2</td>
<td>0.6</td>
<td>-</td>
<td>GFP expression</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>GFP at C-term</td>
<td>deps-1</td>
<td>sg6; sg21</td>
<td>11</td>
<td>83</td>
<td>487</td>
<td>7</td>
<td>1.4</td>
<td>-</td>
<td>GFP expression</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>GFP at C-term</td>
<td>mex-6</td>
<td>sg3; sg18</td>
<td>10</td>
<td>90</td>
<td>666</td>
<td>8</td>
<td>1.2</td>
<td>-</td>
<td>GFP expression</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>GFP at C-term</td>
<td>glh-1</td>
<td>sg4</td>
<td>5</td>
<td>54</td>
<td>402</td>
<td>0</td>
<td>0.0</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>GFP at C-term</td>
<td>htp-3</td>
<td>sg5</td>
<td>1</td>
<td>12</td>
<td>84</td>
<td>0</td>
<td>0.0</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Deletion</th>
</tr>
</thead>
<tbody>
<tr>
<td>22 3</td>
</tr>
<tr>
<td>23 3</td>
</tr>
<tr>
<td>24 3</td>
</tr>
<tr>
<td>25 4</td>
</tr>
<tr>
<td>26 1</td>
</tr>
</tbody>
</table>
# P0s: number of injected hermaphrodites whose broods were screened.

# PCRs: total number of F1 pools that were screened by PCR.

# F1s: number of F1s screened.

# edits: number of positive PCRs.

% edits: number of positive PCRs divided by the number of F1s screened. The assumption is that each positive pool contains only one positive F1 animal. This may be an underestimate for the GFP experiments where each pool contained 8 F1s. Indeed in some pools, we observed a higher frequency of edits among F2s than expected if the pool contained only one edited F1.

Verification of expression: WB: western blot, IF: immunofluorescence, GFP expression: GFP fluorescence in live animals in a pattern expected for the targeted ORF. See Figure 3.

*: edits are maternal effect lethal (V5-tagged MEX-5, and deletion in mbk-2 locus). Interestingly, 2/2 independent V5-MEX-5 edits are maternal-effect lethal as is a mex-5(0) mutant (SCHUBERT et al. 2000). In contrast the FLAG-MEX-5 and OLLAS-MEX-5 edits were viable. The ability to generate fusions to different tags in the same experiment will be useful to determine the best tagging strategy for each gene and avoid those tags that interfere with normal protein function.