Genome-engineering with CRISPR-Cas9 in the mosquito *Aedes aegypti*

Kathryn E. Kistler[1], Leslie B. Vosshall[1,2], Benjamin J. Matthews[1,2] ∗

1 Laboratory of Neurogenetics and Behavior, The Rockefeller University, New York, NY 10065 USA

2 Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10065 USA

∗ **E-mail: Ben.Matthews@rockefeller.edu**

Abstract

The mosquito *Aedes aegypti* is a potent vector of the Chikungunya, yellow fever, and Dengue viruses, which result in hundreds of millions of infections and over 50,000 human deaths per year. Loss-of-function mutagenesis in *Ae. aegypti* has been established with TALENs, ZFNs, and homing endonucleases, which require the engineering of DNA-binding protein domains to generate target specificity for a particular stretch of genomic DNA. Here, we describe the first use of the CRISPR-Cas9 system to generate targeted, site-specific mutations in *Ae. aegypti*. CRISPR-Cas9 relies on RNA-DNA base-pairing to generate targeting specificity, resulting in cheaper, faster, and more flexible genome-editing reagents. We investigate the efficiency of reagent concentrations and compositions, demonstrate the ability of CRISPR-Cas9 to generate several different types of mutations via disparate repair mechanisms, and show that stable germ-line mutations can be readily generated at the vast majority of genomic loci tested. This work offers a detailed exploration into the optimal use of CRISPR-Cas9 in *Ae. aegypti* that should be applicable to non-model organisms previously out of reach of genetic modification.

Introduction

As a primary vector of the serious and sometimes fatal Chikungunya, yellow and Dengue viruses, the mosquito *Aedes aegypti (Ae. aegypti)* is responsible for hundreds of millions of human infections annually [1]. To transmit disease, a female mosquito must first bite an infected individual, and, after a period of viral incubation within the mosquito, bite another human. Female mosquitoes use cues such as odor, carbon dioxide, and temperature to locate a host and obtain a blood-meal [2], used to produce a clutch of approximately 100 eggs. Once a mosquito has developed mature eggs, she uses volatile and contact cues to locate and evaluate a body of water at which to lay her eggs, or oviposit. Our long-term goals involve using genome-engineering techniques coupled with quantitative behavioral analysis to investigate the genetic and neural basis of innate chemosensory behaviors in this important disease vector.

Clustered regularly interspaced palindromic repeats (CRISPR) and CRISPR associated (Cas) genes are components of an adaptive immune system that are found in a wide variety of bacteria and archaea [3]. Beginning in late 2012 [4], the bacterial type II CRISPR-Cas9 system has been adapted as a genome-engineering tool in a wide variety of organisms and *in vitro* preparations, dramatically expanding the ability to introduce specific genome modifications [3]. In particular, the ease of designing and generating these reagents at the bench has opened the door for studies of gene function in non-traditional model organisms.

The genome of *Ae. aegypti* is relatively large and incompletely mapped [5–8], presenting difficulties in recovering mutations generated by traditional forward genetics. *Ae. aegypti* has a recent history of genetic modification, including transposon-mediated transgenesis [9, 10] and loss-of-function gene editing with zinc-finger nucleases (ZFNs)

[2, 11, 12], TAL-effector nucleases (TALENS) [13, 14], and homing endonuclease genes (HEGs) [15]. ZFNs and TALENs are modular DNA-binding proteins tethered to a non-specific FokI DNA nuclease [16], while HEGs are naturally occurring endonucleases that can be reengineered to target novel sequences [17]. Targeting specificity in these classes of genome-engineering reagents is conferred by context-sensitive protein-DNA binding interactions that are not completely understood. It is therefore technically difficult to engineer these proteins to target many sequences of interest.

In this paper, we describe methods for site-directed mutagenesis in *Ae. aegypti* using RNA-guided endonucleases (RGENs) based on the type II CRISPR-Cas9 system. RNA-DNA Watson-Crick base pairing is used to target the double-stranded endonuclease Cas9, derived from *Streptococcus pyogenes*, to specific genome locations where it introduces a double-stranded break. In just two years, the CRISPR-Cas9 system has been adapted for precision genome-engineering in dozens of model organisms from bacteria to primates [3, 18]. While many of these efforts have influenced our work, two studies in the vinegar fly *Drosophila melanogaster* [19] and the zebrafish *Danio rerio* [20] were particularly important in guiding our early attempts to adapt CRISPR-Cas9 to the mosquito *Ae. aegypti*.

A detailed bench manual with step-by-step guidance for designing, generating, and testing these reagents is available as a supplement to this paper. This work is not an exhaustive exploration of the parameters controlling the efficiency of CRISPR-Cas9 mutagenesis in *Ae. aegypti*, but is instead a practical guide for generating mosquito mutants. Given the proven flexibility of the CRISPR-Cas9 system, we believe that the protocols and procedures outlined here and by numerous other laboratories will continue to be optimized and modified for use in many organisms for which precision genome-engineering has not yet been employed.

Materials and Methods

A detailed bench manual is available as Supplemental File S1.

Microinjection and insect rearing

Microinjection into *Ae. aegypti* embryos was performed according to standard protocols [21] at the Insect Transformation Facility at the University of Maryland. Embryos were hatched 3 days post-injection and reared to pupal or adult stages according to previously described rearing procedures [2, 12]. Liverpool IB12 (LVP-IB12) mosquitoes [5] were used both as the injection strain and as the wild-type strain for out-crossing and were obtained through the MR4 as part of the BEI Resources Repository, NIAID, NIH. This *Ae. aegypti* LVP-IB12, MRA-735 strain, was originally deposited by M.Q. Benedict. Female mosquitoes were provided with a mouse or human blood source for egg production. All laboratory blood-feeding procedures with mice and humans were approved and monitored by The Rockefeller University Institutional Animal Care and Use Committee and Institutional Review Board, protocols 11487 and LVO-0652, respectively. All humans gave their informed consent to participate in mosquito blood-feeding procedures.

Cas9 mRNA and protein

Cas9 mRNA was transcribed from vector pMLM3613 (plasmid 42251, AddGene) [20] using mMessage mMachine T7 Ultra Transcription kit (AM1345, Life Technologies). Recombinant Cas9 protein was obtained commercially (CP01, PNA Bio).

sgRNA design and construction

sgRNAs were designed by manually searching genomic regions for the presence of protospacer-adjacent motifs (PAMs) with the sequence NGG, where N is any nucleotide. We required that sgRNA sequences be 17-20 bp in length, excluding the PAM, and contain one or two 5' terminal guanines to facilitate transcription by T7 RNA polymerase. sgRNA sequences were checked for potential off-target binding using the following two web tools: `http://zifit.partners.org/ZiFiT/` and `http://crispr.mit.edu`. To minimize the potential for off-target mutagenesis, we avoided sgRNA sequences with closely related binding sites, defined as fewer than 3 mismatched bases to reduce the possibility of cutting, whenever possible. See Table 1 for sgRNA sequences and closest matches.

Linear double-stranded DNA templates for specific sgRNAs were produced by a template-free polymerase chain reaction (PCR) with two partially overlapping oligos (primers sgRNA-R and sgRNA-F; Table 2) [19]. In early experiments, sgRNA sequences were cloned into pDR274 [20] and linearized with PmeI (R0560, NEB). In both cases sgRNAs were produced using the MegaScript T7 *in vitro* transcription kit (AM1334, Life Technologies) overnight at 37°C. Transcribed sgRNA was purified with MegaClear columns (AM1908, Life Technologies).

Extraction of genomic DNA

Genomic DNA was extracted from individual or pools of mosquitoes using either the DNEasy Blood and Tissue Kit (69581, Qiagen) or a 96-well plate extraction protocol [22].

Sequencing and analysis of CRISPR-Cas9 induced mutations

A two-step PCR protocol was used to amplify amplicons surrounding the putative CRISPR-Cas9 cut site from genomic DNA of G_0 mosquitoes (those that were injected as embryos and allowed to develop to pupal or adult stages) or G_1 individuals (the progeny of G_0 individuals crossed to LVP-IB12). First, genome-specific primers were designed with the following tails:

- Forward: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-GENESPECIFIC-FORWARD-3'
- Reverse: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-GENESPECIFIC-REVERSE-3'

20 cycles of PCR was performed using KOD Hotstart polymerase (71086-3, EMD Millipore). The product of the first PCR was used as a template for a second round of PCR with Nextera XT Indexed Primers (item FC-131-1002, Illumina). PCR cycles were kept to a minimum to reduce the non-linearity of PCR amplification that can occur at higher cycle numbers. Samples were purified using Ampure XP magnetic beads at 0.6x volume to separate amplicons from primers, were examined on an Agilent Bioanalyzer to verify purity and sizing, and pooled in roughly equimolar amounts. To ensure high sequencing quality, final amplicon pools were quantitated by qPCR to determine the precise molarity of amplicons containing the adapters need for clustering and sequencing (KK4844, KAPA BioSystems). Samples were sequenced on an Illumina MiSeq following the manufacturer's instructions.

Following sequencing, reads were aligned to reference sequences of the PCR amplicons and examined for the presence of insertions, deletions, or other polymorphisms. Scripts developed for the analysis of this data, including those used to run the alignments and mutation quantitation for all figures, are available at `https://github.com/bnmtthws/crispr_indel`. Other software packages used include GMAP/GSNAP [23]

http://research-pub.gene.com/gmap/, GATK [24] https://www.broadinstitute.org/gatk/, and pysamstats https://github.com/alimanfoo/pysamstats.

Donor construction for homology-directed repair

Single-stranded DNA oligodeoxynucleotides (ssODNs) were ordered as 200 bp 'Ultramers' from Integrated DNA Technologies (IDT). Plasmids used as double-stranded DNA donors were constructed by PCR-amplifying homology arms from LVP-IB12 genomic DNA and cloning with In-Fusion HD cloning (Clontech) into one of two base plasmids, PSL1180polyUBdsRED (AddGene #49327) and pSL1180-HR-PUbECFP (AddGene #47917). Annotated sequences of oligonucleotides and plasmids used for homology-directed repair are available as Supplemental File S2.

Molecular genotyping of stable germ line alleles by PCR

To verify the presence of exogenous sequences inserted by homology-directed-repair or the presence of insertions and deletions, PCR amplicons surrounding the putative cut site were generated from genomic DNA (see Table 2 for primer sequences). Following purification, amplicons were Sanger sequenced (Genewiz), or used as a template for a restriction digest using the enzyme BamHI (R0136, New England Biolabs [NEB]) or PacI (R0547, NEB).

Genotyping stable germ line alleles by fluorescence

Larvae or pupae were immobilized on a piece of moist filter paper and examined under a dissection microscope (SMZ1500, Nikon) with a fluorescent light source and ECFP and dsRed filtersets.

Results

Outline of CRISPR and injection components

The bacterial type-II CRISPR system has been adapted in many organisms to generate RNA-guided endonucleases, or RGENs, that are targeted to specific regions of the genome by RNA-DNA base-pairing [3]. The core of this adapted CRISPR/Cas system is comprised of two components: 1) a synthetic single guide RNA (sgRNA), which is a small RNA containing 17-20 bases of complementarity to a specific genomic sequence, and 2) the Cas9 nuclease derived from *Streptococcus pyogenes* (SpCas9). SpCas9 forms a complex with the sgRNA and induces double-stranded DNA breaks at regions of the genome that fulfill two criteria: 1) contain a sequence complementary to the recognition site of the sgRNA that is 2) directly adjacent to a protospacer-adjacent motif (PAM). The PAM sequence for SpCas9 takes the form of NGG. These motifs occur approximately once every 17 bp in the *Ae. aegypti* genome, making it possible to target essentially any genomic loci of interest with the CRISPR-Cas9 system.

To generate stable mutations that can be transmitted through the germline, CRISPR-Cas9 reagents must be introduced pre-blastoderm stage embryos composed of a syncytium of nuclei prior to cellularization (Figure 1). This early developmental stage offers access of genome-editing reagents to the nuclei of both somatic and germ-line cells. In brief, embryos are microinjected 4-8 hours after egg-laying, and allowed to develop for 3 days before being hatched in a deoxygenated hatching solution [21]. Following hatching, pupae are collected for sequencing or allowed to emerge as adults and outcrossed to wild-type LVP-IB12 males or females, as appropriate. Genomic DNA from injected embryos is examined for modification rates. Following blood-feeding, G_1 eggs are collected from these outcrosses to screen for germ-line transmission of stable

mutations.

When faced with a double-stranded DNA break, DNA repair machinery can resolve this break in one of two ways: non-homologous end-joining, which can result in small insertions and deletions, or, less frequently, homology-directed repair, which uses exogenous sequence containing regions of homology surrounding the cut site as a template for repair. Additionally, cutting with multiple sgRNAs can result in large deletions between the two cut sites. In this paper, we will discuss stable germ-line transmission of all three types of these alleles in *Ae. aegypti*.

Identifying optimal injection mixes for CRISPR-Cas9 mutagenesis

The insertions and deletions resulting from non-homologous end-joining act as a signature for evaluating the efficiency of reagents such as ZFNs, TALENs, and CRISPR-Cas9. Quantifying the level of these insertions and deletions can give an indication of the activity of a particular sgRNA/Cas9 combination. A variety of techniques have been described to evaluate the cutting efficiency of Cas9/sgRNA RNA-guided nucleases. These include enzymatic detectors of polymorphisms such as Surveyor or T7 Endonuclease I [25], high-resolution melting point analysis (HRMA) [26], Sanger sequencing [27] or deep sequencing [28]. Each of these techniques evaluates the level of polymorphism in a short PCR-generated amplicon surrounding the sgRNA target site. With the exception of deep-sequencing, these approaches provide only semi-quantitative estimates of the effectiveness of mutagenesis in each sample. In addition, the *Ae. aegypti* genome is highly polymorphic, and so techniques that are based on detecting sequence mismatches, such as HRMA, T7E1, or Surveyor nuclease assays, are prone to false positives from polymorphisms present in our wild-type strains.

For these reasons, we opted to prepare and deep sequence barcoded PCR amplicons surrounding putative CRISPR-Cas9 cut sites from small pools of injected animals to accurately determine the rates of cutting at different genomic loci. Briefly, sequencing libraries are prepared using a two-step PCR process that incorporates adapter and barcode sequences necessary for Illumina sequencing (Figure 2A). Many distinct barcoded amplicons can be pooled together using Illumina-supplied Nextera XT primers. We estimate that 10,000-100,000 reads are ample for this analysis, meaning that at current prices, sequencing of amplicons from 3 sgRNAs per gene, for 10 distinct genes can be run for a cost of approximately $70 per gene (MiSeq v3 reagents, 150-cycle flowcell, item MS-102-3001).

Following sequencing, reads were aligned to a reference sequence using the GSNAP short read aligner [23], and insertions and deletions were quantified using the python package pysamstats. This procedure results in data on the number of polymorphisms, including insertions and deletions, found in reads that span each nucleotide of a reference sequence (Figure 2B). Notably, in injections containing a sgRNA and Cas9, a pattern of elevated insertions and deletions can be observed with a peak 3 bp 5' of the beginning of the PAM, exactly the position at which Cas9 is known to make a double-stranded break (Figure 2C). Importantly, there was high concordance in the mutagenesis rates seen between multiple biological replicates generated from different pools of pupae from a single injection (Figure 2C), suggesting that there is minimal variability in mutagenesis frequency in individual injected animals.

We next varied the delivery method and concentration of Cas9 and the concentration of a given sgRNA to determine an optimal injection mix composition that would yield high levels of somatic and germline mutations. We first attempted mutagenesis with plasmid DNA constructed to express Cas9 under the control of an *Ae. aegypti* polyubiquitin

(PUb) promoter that can effectively drive TALEN half-sites when microinjected into embryos [13,29]. However, injection mixes containing PUb-Cas9 plasmid and a validated sgRNA did not induce detectable rates of insertions or deletions (data not shown).

The failure of Cas9 expressed by plasmid DNA might be due to the temporal dynamics of transcription and translation, so we explored two additional methods of Cas9 delivery: mRNA produced by *in vitro* transcription and recombinant SpCas9 protein. When included at 500 ng/µL, both Cas9 mRNA and protein induced detectable mutagenesis at two distinct guide RNA sites. However, recombinant Cas9 protein induced mutagenesis at rates 5-10x higher than Cas9 mRNA (Figure 2D). To test whether the concentration of sgRNA or Cas9 mRNA or protein was limiting in these earlier injections, we tried four additional injection mixes (Figure 2E), with a single validated sgRNA, and determined that mixes containing 400 ng/µL Cas9 recombinant protein induced the highest rates of mutagenesis. Increasing sgRNA concentration did not dramatically increase mutagenesis rates.

Identifying active sgRNAs for a given genomic target

We reasoned that sgRNAs that show higher somatic cut rates in injected animals would be more likely to result in stable germ-line transformation. We first designed 3 different sgRNAs against 6 different genes using publicly-accessible design tools to minimize the potential for off-target mutagenesis (Figure 3A). These varied in length between 17-20 bp [30], and in addition to being adjacent to a PAM sequence, were required to begin with GG or G to facilitate *in vitro* transcription of sgRNA.

To test the efficiency of these sgRNAs, we performed a series of 6 small test injections (145-168 embryos each) into *Ae. aegypti* embryos (Figure 3B). Each injection

mix was comprised of recombinant Cas9 protein at 333 ng/µL and a pool of three sgRNAs (40 ng/µL each), each targeted against a different gene. Survival rates were very high for these injections (ranging from 46.1%-63.3%, as compared to a 18.6% average survival in previous experiments with 500 ng/µL Cas9 protein or mRNA), and we attribute this marked increase in survival to the reduction in Cas9 (or overall injection mix) concentration. Surviving embryos were reared to pupal stages and collected for PCR amplicon preparation and sequencing analysis (Figure 3B).

All 18 sgRNAs induced detectable levels of mutagenesis, although the rates of cutting were highly variable within and between different genomic targets (Figure 3C). Thus, in addition to potential effects of chromatin state or other genome accessibility at a particular genomic locus, there appear to be sequence- or context-dependent effects on sgRNA efficiency that are not yet fully understood. Designing and testing 3 sgRNAs resulted in the identification of at least one highly active sgRNA for the 6 genomic targets tested here. The variability observed between different sgRNAs targeting the same locus underscores the benefits of testing multiple sgRNAs per gene before undertaking large-scale mutagenesis injections.

Germ-line transmission of mutant alleles

We next examined whether somatic mutagenesis detected in adults reared from injected embryos (G_0 animals) could result in transmission of stable mutant alleles through the germline to their offspring (G_1 animals). We designed an sgRNA near the 5' end of *Aaeg-wtrw*, a single-exon gene identified through bioinformatics which bears homology to the *D. melanogaster* TRP channel *water witch (wtrw)* [31]. We also included 300 ng/µL of a 200 bp ssODN donor to act as a template for homology-directed repair. The

ssODN comprised homology arms of 87-90 bases on either side of the Cas9 cut site, with an insert containing stop codons in all three frames of translation as well as an exogenous restriction enzyme recognition site. Successful integration of this template would result in a truncated protein of 91 amino acids rather than a full-length protein of 908 amino acids (Figure 4A).

636 embryos were injected with a mixture of 200 ng/µL Cas9 mRNA and 12.5 ng/µL sgRNA. Note that these injections were performed prior to the optimization of injection mixes described above. To determine the activity of our injected reagents, we performed amplicon sequencing on 6 pools of 5-6 adult G_0 animals after they were outcrossed and allowed to lay eggs (Figure 4B). After alignment and analysis of reads aligning to the reference sequence, we found that these samples contained a mean maximum somatic insertion or deletion rate of 24.87% centered on the Cas9 cut site (Figure 4C), indicating a high rate of somatic mosaicism at the targeted locus in animals injected with this cocktail. As a control, we sequenced amplicons at the *Aaeg-wtrw* locus from animals injected with an injection mix containing another sgRNA targeting a different region of the genome. These samples contained no detectable insertions or deletions at the *Aaeg-wtrw* locus (Figure 4C). Across the six samples, a mean of 0.71% of reads that aligned to the reference sequence contained sequences corresponding to the exogenous sequence in the single-stranded DNA donor (Figure 4D). This indicated that the ssODN template could drive homology-directed repair in somatic tissue, though at a much lower frequency than insertions or deletions mediated by non-homologous end-joining.

To determine whether these mutations were stably transmitted through the germline, we sequenced PCR amplicons derived from pools each containing 5 male and 5 female G_1 offspring that were previously mated together and from which we had already

collected F2 eggs. Analysis of resulting insertions and deletions using Genome Analysis Toolkit [GATK; [24]] revealed that there were at least 117 mutant chromosomes spread across 50 pools, meaning that the G_1 mutation rate was at least 117/620, or 18.9% (Figure 4E). Four of these alleles corresponded perfectly to the sequence of the ssODN, meaning that our rate of stable germline transmission of alleles generated by homology-driven repair recombination is at least 0.6% (Figure 4E). Notably, these numbers correlate well with those derived from G_0 animals (Figure 4E), suggesting that assays of somatic mutagenesis can be useful in predicting the efficiency of germ-line mutagenesis.

We hatched F2 eggs from a single family containing an allele generated by homology-directed repair. Sequencing of single-pair crosses allowed us to isolate a stable mutant line that can be genotyped by a simple restriction digest with the enzyme site introduced by the ssODN targeting event (Figure 4F). To enable future phenotypic characterization, this line was outcrossed for 8 generations to wild-type mosquitoes to increase genetic diversity in the mutant strain and to reduce the possibility of retaining off-target mutations in this population of animals.

Deletions induced by multiplexed sgRNAs

Double-stranded breaks induced at multiple sgRNA sites can induce large deletions between the two cut sites in *D. melanogaster* [32]. We performed a series of five injections into small numbers of embryos using sgRNAs targeting 3 different genes. All but one of these sgRNAs (AAEL000926-sgRNA4) were previously validated as having high activity in somatic mutagenesis assays. We also included ssODN donors with arms on either side of the two breaks and an exogenous sequence containing stop codons in all 3 frames of translation as well as a restriction enzyme recognition

site (Figure 5A). We injected small numbers of embryos and collected G_1 embryos from all female G_0 animals, which were hatched and screened individually (Figure 5B). From these individual families, we screened male pupae in pools via PCR, identifying potential mutations through size-based genotyping on an agarose gel. Individual females were outcrossed, blood-fed, and allowed to lay eggs. PCR amplicons were generated from these animals individually and examined for size-shifted bands on agarose gels (Figure 5C). Candidate alleles were characterized by Sanger sequencing of gel-purified fragments (Figure 5C). We found a wide range of mutant transmission rates in female G_1 animals derived from single G_0 individuals where genotyping male pupae showed the offspring of these G_0 animals to contain at least one mutant G_1 (Figure 5D).

Although we did not fully characterize each potential mutant allele, we did note several different types of mutations in those that were characterized. These ranged from simple deletions, to homology-directed repair from the ssODN donor, to more complex modifications, including polymorphisms, inversions, and duplications. This indicates that induction of multiple double-stranded breaks is highly mutagenic. We also note that we were successful in obtaining germ-line mutations at high rates in all 3 injections (Figure 5E), making this a cost-effective and efficient way to generate loss-of-function mutant alleles. The relatively large size of deletions generated by this method makes sized-based molecular genotyping straightforward and greatly facilitates the out-crossing and homozygosing of mutant lines.

Integration and transmission of large fluorescent cassettes

Finally, we asked whether CRISPR-Cas9 could be used to introduce longer cassettes of exogenous sequence via homology-dependent repair. In previous genome-engineering

experiments in *Ae. aegypti*, zinc-finger nucleases were used to introduce large cassettes from a plasmid DNA donor with homology arms of at least 800 bp on either side [2, 11]. The insertion of a fluorescent cassette simultaneously creates a null mutant by interrupting the protein-coding sequence and inserts the visible fluorescent reporter. We performed injections with Cas9 protein, a validated sgRNA, and a plasmid donor containing homology arms of 799 bp and 1486 bp. This plasmid contains a cassette comprising the constitutive PUb promoter [29] driving the expression of a fluorescent reporter (Figures 6A-B). These arms were cloned from the strain of mosquito into which injections were performed to maximize homology, and were designed to avoid repetitive sequences such as transposable elements. Following injection, individual female G_0 animals were outcrossed to wild-type mosquitoes and G_1 eggs were collected (Figure 6A). Past experiences in our laboratory indicate that successful homology-directed repair occurs primarily, if not exclusively, in female G_0 animals in *Ae. aegypti*. We therefore discarded G_0 males and restricted our screening to G_1 families generated from females.

G_1 individuals were screened under a fluorescence dissecting microscope as larvae at 3-5 days post-hatching. The expression of the PUb promoter is clearly visible in larvae and pupae (Figure 6C and 7B). Fluorescent individuals were collected and reared to adulthood and crossed again to wild-type animals to establish stable lines. To verify directed insertion of our cassette as opposed to non-targeted integration at another genomic locus, we designed PCR primers spanning both homology arms (Figure 6D). It is critical that these primers are designed outside each arm and that bands obtained are sequenced to verify junctions between genomic and exogenous sequence on each end of the insertion.

Lines containing verified targeted insertions were out-crossed to wild-type mosquitoes

for 8 generations, at which point a homozygous line was established by mating heterozygous males and females and selecting putative homozygous individuals from their offspring. Because homozygous individuals show increased levels of fluorescence relative to heterozygotes, the generation of homozygous lines is straightforward. Putative homozygous mosquitoes were separated by sex and used to establish single-pair matings, and the genotype of these single-pair matings were verified by PCR across the integration site of the cassette. In this reaction, PCR primers span the entire cassette, resulting in a large shift in size in homozygous mutant individuals (Figure 6E).

We generated verified targeted insertions in two genomic loci with the *Ae. aegypti* PUb promoter driving the expression of ECFP (Figures 6, 7A) or dsRed (Figure 7B). Attempts at two additional loci failed to produce directed insertion of the donor cassette (Figures 7C-D). We conclude that, similar to the rates of ssODN integration, homology-directed repair from with large plasmid donors occurs at a relatively low frequency compared to other forms of CRISPR-Cas9-mediated genome modification. A drastic variance in the efficiency noted in two injections (Figure 7A) suggests that higher rates can be achieved through simple modifications of injection mix component concentration, perhaps at the expense of embryo survival. The presence of non-directed targeting events underscores the necessity of verifying all lines generated by this technique by PCR or other molecular methods (Figure 6).

Taken together, these data show that the CRISPR-Cas9 system can be been successfully adapted for genome-engineering in the mosquito *Ae. aegypti*. Following the generation of a double-stranded break by an RNA-guided Cas9 nuclease, a variety of mutant alleles can be recovered, including frame-shift mutations caused by insertions or deletions, deletion of a region between two sgRNA target sites, and integration of exogenous sequences from a single-stranded oligonucleotide or a double-stranded plas-

mid DNA donor. We recommend these methods to any laboratory wishing to generate loss-of-function mutations in *Ae. aegypti*.

Discussion and Conclusions

We have demonstrated that the CRISPR-Cas9 system is a highly effective and efficient tool for precision genome-editing in the mosquito *Ae. aegypti*. Compared to the relatively low throughput and high cost of ZFN- and TALEN-mediated mutagenesis, the ease and efficiency of designing and producing CRISPR-Cas9 reagents in the laboratory has allowed us to generate stable and precise loss-of-function mutations in over 10 genes to date. This protocol provides a step-by-step manual to mutagenesis in *Ae. aegypti* and also provides general principles that will be useful when translated to other species.

Optimal injection mix

We recommend using recombinant Cas9 protein for its reproducibility and observed increased rates of mutagenesis and embryo survival. Recombinant Cas9 protein is likely to be much more stable than mRNA, both at the bench and in injected embryos. Additionally, it is likely that sgRNA and Cas9 protein can form a complex in the absence of any other factors, allowing them to generate pre-formed RGENs prior to injection. This can serve to both stabilize sgRNA/Cas9 complexes [33] and ensure that mutagenesis can occur immediately, rather than waiting for Cas9 mRNA to be translated in the embryo. The specific concentrations suggested here represent a good trade-off between survival and efficiency in our hands. However, it is entirely possible that further modifications to this protocol could result in significant increases in certain types of repair.

For CRISPR-Cas9 injections into *Ae. aegypti*, we currently recommend the following

injection mix. This mix may also be a good starting point for microinjections into other insect embryos.

- 300 ng/µL recombinant Cas9 protein
- 40 ng/µL sgRNA (each)
- (optional) donor DNA
    - 200 ng/µL single stranded ssODN **or**
    - 500 ng/µL double stranded plasmid DNA

Designing active sgRNAs

We observed significant variability in the effectiveness of different sgRNAs, even between sgRNAs targeted to a small genomic region (Figure 3). As in other organisms and cell lines [30, 34], we observed success with sgRNAs ranging in length from 17-20 bp, and recommend considering sgRNAs in this size range when screening for the most active target sites. A single genomic target (AAEL001123) has proven resistant to mutagenesis with 6 different sgRNAs, perhaps reflecting an underlying chromatin-state [35]. Aside from this one example, we were able to design effective sgRNAs to all other examined genes. Systematic examination of sgRNA efficiency in *D. melanogaster* suggests that the GC content of the 6 nucleotides adjacent to the PAM contributes to the efficiency of a given sgRNA [34]. We will incorporate these suggestions as well as those generated from further investigations into the sequence or context-dependency of sgRNA activity into our future sgRNA design. However, given the ease of testing sgRNA efficiency *in vivo*, we still recommend the design and testing of multiple sgRNAs targeting a given gene before committing to large-scale injections.

Off-target effects

Off-target effects are a concern with any genome-editing technology. We currently adopt several approaches to try and address these concerns in our experiments. First, we check for the sgRNA specificity using publicly available bioinformatic tools [20, 36, 37], selecting the most specific sgRNAs within the region we wish to target. Second, we design several sgRNAs for a given genomic target. For genes of particular interest, we can generate mutant alleles with multiple sgRNAs and examine their phenotype in transallelic combination, reducing the likelihood that these independent lines will share the same off-target mutations. Third, we have successfully used truncated sgRNAs of less than 20 bp in length, which have been shown in cell culture to reduce the likelihood of off-target modifications [30]. Finally, we subject all lines to at least 8 generations of out-crossing to wild-type mosquitoes, which should reduce the co-inheritance of all but the most tightly linked off-target mutations. While we believe these guidelines reduce the likelihood of significant levels of off-target mutagenesis, we recognize the need for continued efforts to improve and to verify the specificity of all precision genome-engineering technologies.

Variants on the traditional CRISPR-Cas9 reagents designed to improve specificity include the use of paired Cas9 'nickases' to generate adjacent single-stranded DNA breaks [38–41]. These dual 'nicks' are proposed to be less mutagenic than a single double-stranded break. This approach should decrease the chance that the binding of a single sgRNA at an off-target site produces a mutation. Another approach fuses a nuclease-dead Cas9 mutant (dCas9) to the FokI nuclease. As in ZFN and TALENs, FokI requires paired binding, within a certain distance, on opposite strands of a target, and this technique aims to combine the specificity of TALENs and ZFNs with the ease-of-targeting of CRISPR-Cas9 [42, 43]. Though we have not investigated these

CRISPR-Cas9 variants in our experiments, we believe that their benefits are likely to translate to *Ae. aegypti* and other organisms in the future.

Enhancing the efficiency of homology-directed repair

In our experiments, insertions and deletions mediated by non-homologous end-joining occur at much higher frequency than by homology-directed repair. This is similar to what has been observed in *Ae. aegypti* with other genome-editing tools, such as ZFNs [2, 11], and in other organisms such as *D. melanogaster* [44].

Several approaches have been developed to increase rates of homology-directed repair. These include injections in the background of a DNA ligase 4 mutation [45, 46] or schemes that linearize a double-stranded donor template *in vivo*. Finally, many laboratories working with *D. melanogaster* have developed transgenic strains that express Cas9 protein under ubiquitous or germ-line promoters, improving the efficiency of mutagenesis generally and homology-directed repair specifically [34, 44]. Other areas of interest include alternatives to homology-directed repair, such as ObLiGaRe [47] and homology-independent double-stranded break repair pathways [48]. It remains to be seen whether transgenic Cas9 delivery or alternative integration approaches can be effectively implemented in *Ae. aegypti*.

We note that we have observed a single round of injection that resulted in high ($>$30%) rates of homology-directed repair but extremely low survival (Figure 7A). This suggests that we might achieve improvements in insertion efficiency by continuing to examine the effects of modulating the concentrations of the three components of the injection mix. If the rates of homology-directed-repair can be sufficiently improved, CRISPR-Cas9 coupled with transgene insertion via homology-directed repair will likely prove to be a versatile tool to tag gene products and introduce transgenes into specific

genomic loci, enabling studies of specific neural circuits and other subsets of cells.

Conclusions

Precision genome-engineering in mosquitoes holds great promise for studies on the genetic basis of behavior [2,11,12] and for genetic strategies to control vector population or disease competence [49]. Ongoing efforts to increase the specificity and efficiency of these technologies is critical to their adaptation as routine techniques, and we believe that the protocols outlined here have met those criteria for the generation of loss-of-function mutations in the mosquito *Ae. aegypti*. Reagents based on the CRISPR-Cas9 platform have been used successfully in organisms from bacteria to primates. This suggests that the techniques described here can likely be adapted to many other non-model organisms, so long as efficient methods of introducing the reagents into the germline and screening for mutations can be developed.

Acknowledgments

We thank Rob Harrell and the Insect Transformation Facility (University of Maryland) for expert mosquito embryo microinjection, Gloria Gordon and Libby Meija for mosquito rearing assistance, and members of the Vosshall laboratory for helpful comments and critiques. This work was supported in part by contract HHSN272200900039C from the National Institute of Allergy and Infectious Diseases and grant UL1 TR000043 from the National Center for Advancing Translational Sciences (NCATS, National Institutes of Health (NIH) Clinical and Translational Science Award (CTSA) program). B.J.M. was a Jane Coffin Childs Postdoctoral Fellow and L.B.V is an investigator of the Howard Hughes Medical Institute.

References

1. Bhatt, S, Gething, P. W, Brady, O. J, Messina, J. P, Farlow, A. W, Moyes, C. L, Drake, J. M, Brownstein, J. S, Hoen, A. G, Sankoh, O, Myers, M. F, George, D. B, Jaenisch, T, Wint, G. R. W, Simmons, C. P, Scott, T. W, Farrar, J. J, & Hay, S. I. (2013) The global distribution and burden of dengue. *Nature* **496**, 504–507.

2. McMeniman, C. J, Corfas, R. A, Matthews, B. J, Ritchie, S. A, & Vosshall, L. B. (2014) Multimodal integration of carbon dioxide and other sensory cues drives mosquito attraction to humans. *Cell* **156**, 1060–1071.

3. Doudna, J. A & Charpentier, E. (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096.

4. Jinek, M, Chylinski, K, Fonfara, I, Hauer, M, Doudna, J. A, & Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821.

5. Nene, V, Wortman, J. R, Lawson, D, Haas, B, Kodira, C, Tu, Z. J, Loftus, B, Xi, Z, Megy, K, Grabherr, M, Ren, Q, Zdobnov, E. M, Lobo, N. F, Campbell, K. S, Brown, S. E, Bonaldo, M. F, Zhu, J, Sinkins, S. P, Hogenkamp, D. G, Amedeo, P, Arensburger, P, Atkinson, P. W, Bidwell, S, Biedler, J, Birney, E, Bruggner, R. V, Costas, J, Coy, M. R, Crabtree, J, Crawford, M, Debruyn, B, Decaprio, D, Eiglmeier, K, Eisenstadt, E, El-Dorry, H, Gelbart, W. M, Gomes, S. L, Hammond, M, Hannick, L. I, Hogan, J. R, Holmes, M. H, Jaffe, D, Johnston, J. S, Kennedy, R. C, Koo, H, Kravitz, S, Kriventseva, E. V, Kulp, D, Labutti, K, Lee, E, Li, S, Lovin, D. D, Mao, C, Mauceli, E, Menck, C. F. M, Miller, J. R, Montgomery, P, Mori, A, Nascimento, A. L, Naveira, H. F, Nusbaum, C, O'leary, S, Orvis, J, Pertea, M,

Quesneville, H, Reidenbach, K. R, Rogers, Y.-H, Roth, C. W, Schneider, J. R, Schatz, M, Shumway, M, Stanke, M, Stinson, E. O, Tubio, J. M. C, Vanzee, J. P, Verjovski-Almeida, S, Werner, D, White, O, Wyder, S, Zeng, Q, Zhao, Q, Zhao, Y, Hill, C. A, Raikhel, A. S, Soares, M. B, Knudson, D. L, Lee, N. H, Galagan, J, Salzberg, S. L, Paulsen, I. T, Dimopoulos, G, Collins, F. H, Birren, B, Fraser-Liggett, C. M, & Severson, D. W. (2007) Genome sequence of Aedes aegypti, a major arbovirus vector. *Science* **316**, 1718–1723.

6. Timoshevskiy, V. A, Severson, D. W, Debruyn, B. S, Black, W. C, Sharakhov, I. V, & Sharakhova, M. V. (2013) An integrated linkage, chromosome, and genome map for the yellow fever mosquito Aedes aegypti. *PLoS Neglected Tropical Diseases* **7**, e2052.

7. Timoshevskiy, V. A, Kinney, N. A, Debruyn, B. S, Mao, C, Tu, Z, Severson, D. W, Sharakhov, I. V, & Sharakhova, M. V. (2014) Genomic composition and evolution of Aedes aegypti chromosomes revealed by the analysis of physically mapped supercontigs. *BMC Biology* **12**, 27.

8. Juneja, P, Osei-Poku, J, Ho, Y. S, Ariani, C. V, Palmer, W. J, Pain, A, & Jiggins, F. M. (2014) Assembly of the Genome of the Disease Vector Aedes aegypti onto a Genetic Linkage Map Allows Mapping of Genes Affecting Disease Transmission. *PLoS Neglected Tropical Diseases* **8**, e2652.

9. Coates, C. J, Jasinskiene, N, Miyashiro, L, & James, A. A. (1998) Mariner transposition and transformation of the yellow fever mosquito, Aedes aegypti. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 3748–3751.

10. Lobo, N. F, Hua-Van, A, Li, X, Nolen, B. M, & Fraser, M. J. (2002) Germ line transformation of the yellow fever mosquito, Aedes aegypti, mediated by transpositional insertion of a piggyBac vector. *Insect Molecular Biology* **11**, 133–139.

11. Liesch, J, Bellani, L. L, & Vosshall, L. B. (2013) Functional and genetic characterization of neuropeptide Y-like receptors in Aedes aegypti. *PLoS Neglected Tropical Diseases* **7**, e2486.

12. DeGennaro, M, McBride, C. S, Seeholzer, L, Nakagawa, T, Dennis, E. J, Goldman, C, Jasinskiene, N, James, A. A, & Vosshall, L. B. (2013) orco mutant mosquitoes lose strong preference for humans and are not repelled by volatile DEET. *Nature* **498**, 487–491.

13. Aryan, A, Anderson, M. A. E, Myles, K. M, & Adelman, Z. N. (2013) TALEN-based gene disruption in the dengue vector Aedes aegypti. *PLoS ONE* **8**, e60082.

14. Aryan, A, Myles, K. M, & Adelman, Z. N. (2014) Targeted genome editing in Aedes aegypti using TALENs. *Methods* **69**, 38–45.

15. Aryan, A, Anderson, M. A. E, Myles, K. M, & Adelman, Z. N. (2013) Germline excision of transgenes in Aedes aegypti by homing endonucleases. *Scientific Reports* **3**, 1603.

16. Carroll, D. (2014) Genome Engineering with Targetable Nucleases. *Annual Review of Biochemistry* **83**, 409–439.

17. Stoddard, B. L. (2014) Homing endonucleases from mobile group I introns: discovery to genome engineering. *Mobile DNA* **5**, 7.

18. Peng, Y, Clark, K. J, Campbell, J. M, Panetta, M. R, Guo, Y, & Ekker, S. C. (2014) Making designer mutants in model organisms. *Development* **141**, 4042–4054.

19. Bassett, A. R, Tibbit, C, Ponting, C. P, & Liu, J.-L. (2013) Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. *Cell Reports* **4**, 220–228.

20. Hwang, W. Y, Fu, Y, Reyon, D, Maeder, M. L, Tsai, S. Q, Sander, J. D, Peterson, R. T, Yeh, J.-R. J, & Joung, J. K. (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology* **31**, 227–229.

21. Lobo, N. F, Clayton, J. R, Fraser, M. J, Kafatos, F. C, & Collins, F. H. (2006) High efficiency germ-line transformation of mosquitoes. *Nature Protocols* **1**, 1312–1317.

22. Holleley, C & Sutcliffe, A. (2012) *Methods in Anopheles Research*, Chapter 3.12. (Centers for Disease Control).

23. Wu, T. D & Nacu, S. (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873–881.

24. DePristo, M. A, Banks, E, Poplin, R, Garimella, K. V, Maguire, J. R, Hartl, C, Philippakis, A. A, del Angel, G, Rivas, M. A, Hanna, M, McKenna, A, Fennell, T. J, Kernytsky, A. M, Sivachenko, A. Y, Cibulskis, K, Gabriel, S. B, Altshuler, D, & Daly, M. J. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**, 491–498.

25. Reyon, D, Tsai, S. Q, Khayter, C, Foden, J. A, Sander, J. D, & Joung, J. K. (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nature Biotechnology* **30**, 460–465.

26. Dahlem, T. J, Hoshijima, K, Jurynec, M. J, Gunther, D, Starker, C. G, Locke, A. S, Weis, A. M, Voytas, D. F, & Grunwald, D. J. (2012) Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. *PLoS Genetics* **8**, e1002861.

27. Brinkman, E. K, Chen, T, Amendola, M, & van Steensel, B. (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research* **42**, e168.

28. Gagnon, J. A, Valen, E, Thyme, S. B, Huang, P, Ahkmetova, L, Pauli, A, Montague, T. G, Zimmerman, S, Richter, C, & Schier, A. F. (2014) Efficient Mutagenesis by Cas9 Protein-Mediated Oligonucleotide Insertion and Large-Scale Assessment of Single-Guide RNAs. *PLoS ONE* **9**, e98186.

29. Anderson, M. A. E, Gross, T. L, Myles, K. M, & Adelman, Z. N. (2010) Validation of novel promoter sequences derived from two endogenous ubiquitin genes in transgenic Aedes aegypti. *Insect Molecular Biology* **19**, 441–449.

30. Fu, Y, Sander, J. D, Reyon, D, Cascio, V. M, & Joung, J. K. (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnology* **32**, 279–284.

31. Liu, L, Li, Y, Wang, R, Yin, C, Dong, Q, Hing, H, Kim, C, & Welsh, M. J. (2007) Drosophila hygrosensation requires the TRP channels water witch and nanchung. *Nature* **450**, 294–298.

32. Ren, X, Sun, J, Housden, B. E, Hu, Y, Roesel, C, Lin, S, Liu, L.-P, Yang, Z, Mao, D, Sun, L, Wu, Q, Ji, J.-Y, Xi, J, Mohr, S. E, Xu, J, Perrimon, N, & Ni, J.-Q. (2013) Optimized gene editing technology for Drosophila melanogaster using

germ line-specific Cas9. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 19012–19017.

33. Jinek, M, Jiang, F, Taylor, D. W, Sternberg, S. H, Kaya, E, Ma, E, Anders, C, Hauer, M, Zhou, K, Lin, S, Kaplan, M, Iavarone, A. T, Charpentier, E, Nogales, E, & Doudna, J. A. (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997.

34. Ren, X, Yang, Z, Xu, J, Sun, J, Mao, D, Hu, Y, Yang, S.-J, Qiao, H.-H, Wang, X, Hu, Q, Deng, P, Liu, L.-P, Ji, J.-Y, Li, J. B, & Ni, J.-Q. (2014) Enhanced Specificity and Efficiency of the CRISPR/Cas9 System with Optimized sgRNA Parameters in Drosophila. *Cell Reports* **9**, 1151–1162.

35. Wu, X, Scott, D. A, Kriz, A. J, Chiu, A. C, Hsu, P. D, Dadon, D. B, Cheng, A. W, Trevino, A. E, Konermann, S, Chen, S, Jaenisch, R, Zhang, F, & Sharp, P. A. (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature Biotechnology* **32**, 670–676.

36. Hsu, P. D, Scott, D. A, Weinstein, J. A, Ran, F. A, Konermann, S, Agarwala, V, Li, Y, Fine, E. J, Wu, X, Shalem, O, Cradick, T. J, Marraffini, L. A, Bao, G, & Zhang, F. (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology* **31**, 827–832.

37. Sander, J. D, Maeder, M. L, Reyon, D, Voytas, D. F, Joung, J. K, & Dobbs, D. (2010) ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Research* **38**, W462–8.

38. Mali, P, Aach, J, Stranges, P. B, Esvelt, K. M, Moosburner, M, Kosuri, S, Yang, L, & Church, G. M. (2013) CAS9 transcriptional activators for target specificity

screening and paired nickases for cooperative genome engineering. *Nature Biotechnology* **31**, 833–838.

39. Ren, X, Yang, Z, Mao, D, Chang, Z, Qiao, H.-H, Wang, X, Sun, J, Hu, Q, Cui, Y, Liu, L.-P, Ji, J.-Y, Xu, J, & Ni, J.-Q. (2014) Performance of the Cas9 nickase system in Drosophila melanogaster. *G3* **4**, 1955–1962.

40. Ran, F. A, Hsu, P. D, Lin, C.-Y, Gootenberg, J. S, Konermann, S, Trevino, A. E, Scott, D. A, Inoue, A, Matoba, S, Zhang, Y, & Zhang, F. (2013) Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154**, 1380–1389.

41. Shen, B, Zhang, W, Zhang, J, Zhou, J, Wang, J, Chen, L, Wang, L, Hodgkins, A, Iyer, V, Huang, X, & Skarnes, W. C. (2014) Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nature Methods* **11**, 399–402.

42. Tsai, S. Q, Wyvekens, N, Khayter, C, Foden, J. A, Thapar, V, Reyon, D, Goodwin, M. J, Aryee, M. J, & Joung, J. K. (2014) Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature Biotechnology* **32**, 569–576.

43. Guilinger, J. P, Thompson, D. B, & Liu, D. R. (2014) Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nature Biotechnology* **32**, 577–582.

44. Gratz, S. J, Ukken, F. P, Rubinstein, C. D, Thiede, G, Donohue, L. K, Cummings, A. M, & O'Connor-Giles, K. M. (2014) Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. *Genetics* **196**, 961–971.

45. Beumer, K. J, Trautman, J. K, Mukherjee, K, & Carroll, D. (2013) Donor DNA Utilization during Gene Targeting with Zinc-finger Nucleases. *G3* **3**, 657–664.

46. Beumer, K. J, Trautman, J. K, Christian, M, Dahlem, T. J, Lake, C. M, Hawley, R. S, Grunwald, D. J, Voytas, D. F, & Carroll, D. (2013) Comparing zinc finger nucleases and transcription activator-like effector nucleases for gene targeting in Drosophila. *G3* **3**, 1717–1725.

47. Maresca, M, Lin, V. G, Guo, N, & Yang, Y. (2013) Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome Research* **23**, 539–546.

48. Auer, T. O, Duroure, K, De Cian, A, Concordet, J.-P, & Del Bene, F. (2014) Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Research* **24**, 142–153.

49. Alphey, L. (2014) Genetic control of mosquitoes. *Annual Review of Entomology* **59**, 205–224.

Tables and Figures

Table 1. Sequences of sgRNAs used in this study. Potential off-target sites were evaluated by two publicly available tools. ZiFiT (`http://zifit.partners.org`) was used to determine the number of potential off-target sites in the *Ae. aegypti* genome with 3 or fewer mismatched bases (mm). The number of distinct sites with the given number of mismatches is noted in parentheses. The CRISPR Design Tool (`http://crispr.mit.edu`) uses an algorithm to evaluate potential off-target sites in the context of the *Ae. aegypti* genome and returns a specificity score for sgRNAs of length 20 bp. Scores above 50 are considered "high quality". $*$ indicates sgRNA targets of 17-19bp that cannot be tested with the CRISPR Design Tool

sgRNA name	length	target sequence	PAM	ZiFit off-targets	CRISPR Design Tool score
AAEL000582-sgRNA1	17bp	GGTTGGCAGTTGAGTCC	CGG	2mm (3)	$*$
AAEL010779-sgRNA1	17bp	GGTGACGCGCCTTGGTT	CGG	2mm (1)	$*$
AAEL010779-sgRNA2	20 bp	GGCGGTGACGCGCCTTGGTT	CGG	3mm (1)	97
AAEL010779-sgRNA3	20 bp	GGAACTGACTATTTTCGTCT	CGG	3mm (6)	81
AAEL004091-sgRNA1	20 bp	GAAGTATTTAATTCCATGAA	TGG	3mm (5)	77
AAEL004091-sgRNA2	20 bp	GATAACCGTCGCTTCTGTCG	TGG	3mm (1)	98
AAEL004091-sgRNA3	20 bp	GGTGGTGACGTTTCTCATAT	CGG	3mm (1)	96
AAEL000926-sgRNA1	20 bp	GGTATTTCGGGTGTCGCGAA	AGG	3mm (1)	96
AAEL000926-sgRNA2	20 bp	GGAACGCTCCAAGGTGCAGT	TGG	3mm (2)	94
AAEL000926-sgRNA3	20 bp	GGCAGAAGCCAACTGCACCT	TGG	$>$3mm	95
AAEL000926-sgRNA4	19bp	GGCGTTGAATTTTTAATCT	TGG	3mm (8)	$*$
AAEL014228-sgRNA1	17bp	GGGCAAATCGTAACCGC	TGG	2mm (5)	$*$
AAEL014228-sgRNA2	20 bp	GGATGACTTAACTTTAGTTT	CGG	2mm (1)	83
AAEL014228-sgRNA3	20 bp	GGAGTAGCGTTATACTATTG	TGG	$>$3mm	96
AAEL002575-sgRNA1	20 bp	GGCGCGAAAATCCTCCTGAT	TGG	$>$3mm	97
AAEL002575-sgRNA2	20 bp	GGGCGAACGTTTCCACGATG	AGG	$>$3mm	94
AAEL002575-sgRNA3	20 bp	GGAAGGGAACGTACTTGATG	GGG	3mm (3)	89
AAEL013647-sgRNA1	20 bp	GGCAAACCGTGGCTGCCGGT	CGG	3mm (1)	93
AAEL013647-sgRNA2	20 bp	GGAATCCTCGGTGCATGGTT	TGG	3mm (1)	96
AAEL013647-sgRNA3	20 bp	GACAACAGGTGAGAATTCAC	TGG	$>$3mm	97
wtrw-sgRNA1	20 bp	GGATTTGTTAGAGAGCCTAC	TGG	3mm (1)	94

Table 2. PCR primer sequences used in this study

primer name	sequence (5' - 3')
sgRNA-R	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTT–GCTATTTCTAGCTCTAAAAC
sgRNA-F	GAAATTAATACGACTCACTATA **- sgRNA target sequence -** GTTTTAGAGCTAGAAATAGC
AAEL000582-1-F-miseq	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTACAACTGTAGACACATGTTCCG
AAEL000582-1-R-miseq	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTTTCTGGCGTCCATTCAATC
AAEL010779-12-F1-miseq	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTAGCAATGTCGGTAGTATGGTAGT
AAEL010779-12-R12-miseq	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTACCACAGCCTAGTATGTTCATCA
AAEL010779-3-F1-miseq	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTTGCTAACTACCATACTACCG
AAEL010779-3-R1-miseq	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGATTCAAGGAAAACGCATCG
AAEL004091-1-F1-miseq	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTCGCTTCTGTCGTGGATTGAAAAG
AAEL004091-1-R1F2-miseq	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTGCAACTTTGAGAATACAGCACATTC
AAEL004091-2-F1-miseq	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCACAATAGTGCAGCTGACTCGCGTTC
AAEL004091-2-R1-miseq	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTACTTCTCATAGATGGTGCGCCTTTTAC
AAEL004091-3-F1-miseq	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAATGTGCTGTATTCTCAAAGTTGCAC
AAEL004091-3-R1-miseq	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTACTCTTCTCGTTGAAACTCACAATC
AAEL000926-1-F1-miseq	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTTCTGCCGGAGCGAAGGATTTGTTC
AAEL000926-1-R1-miseq	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAAACTCACAATCACCGGTGTTTGATC
AAEL000926-23-F13R2-miseq	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGCGTTGAATTTTTAATCTTGG
AAEL000926-23-R1-miseq	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGAACTATTGCTACAGTATTCCAGGAAC
AAEL014228-12-F1-miseq	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCGCGATCCTATCATTATTAGTTTCGACC
AAEL014228-12-R2-miseq	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGCACACTGGAGAATAGAAAAATGATC
AAEL014228-3-F1-miseq	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTCGAAACTAATAATGATAGGATCGCG
AAEL014228-3-R1-miseq	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAGTGCAAACCAACTGTTGATTGAAC
AAEL002575-1-F2-miseq	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTAAGAAGATCGCGTAACTGCCGATACAG
AAEL002575-1-R2-miseq	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCGGTGGAAAGTTTATAAATCGCAAG
AAEL002575-2-F12-miseq	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGATCAGGGGGTTAGACGTGAAAC
AAEL002575-2-R2-miseq	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCTATGCCATCTTCACCCATTTTCTAG
AAEL002575-3-F2-miseq	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCGGGCAAGTTCTTCGGCGTAGTGC
AAEL002575-3-R2-miseq	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAGATTGATGTGGCTGCTGATCCTCATC
AAEL013647-1-F1-miseq	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAACGCAGAAACACTCTTTACATACTTC
AAEL013647-1-R1-miseq	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAACCCTCAGAAGGCAAAAGTTTAAATC
AAEL013647-2-F1-miseq	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAACCCTCAGAAGGCAAAAGTTTAAATC
AAEL013647-2-R1-miseq	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTAGACAAACACCGTACGAGAAGTAAATAG
AAEL013647-3-F1-miseq	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTTTGGCCAATATTGATATGCTGGAAG
AAEL013647-3-R1-miseq	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTACTGTTGTAAGTTATCCCGTATGTTG
wtrw-1-F-miseq	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCGACTATCATCAGGAAGAC
wtrw-1-R-miseq	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAATCTATGGCATTTGACCGTACC
wtrw-1-F-BamHI	GACAAAGATCAAAACCCAACAAGTTCCG
wtrw-1-R-BamHI	GTAATAAATTCACACACTCGACGTTCCC
AAEL010779-deletion-F	CTACCACAGCCTAGTATGTTCATCA
AAEL010779-deletion-R	TTCATCGTGCCATTGCTCG
AAEL000926-deletion-F	CCGTCGTTTGCGGAGTTGAGCC
AAEL000926-deletion-R	CTTCCCACCGACAGGACAAACACCAC
AAEL014228-deletion-F	CGCACACTGGAGAATAGAAAAATGATC
AAEL014228-deletion-R	CAGTGCAAACCAACTGTTGATTGAAC
AAEL002575-deletion-F	CGCAAGAAAACTCGACATTGAAGTTG
AAEL002575-deletion-R	CTTGCGGGCAAGTTCTTCGGCGTAG
AAEL000582-ECFP-Larm-F	GTGAGGGTGGTGTCGAATTAACTCTT
AAEL000582-ECFP-Larm-R	CAAGTGACGTCAACCCTTCTAAATCG
AAEL000582-ECFP-Rarm-F	CTGTACAAGAGATCTCGACCCAAGAA
AAEL000582-ECFP-Rarm-R	CCAGCTCAAAGTCCAAAAACGAAACC
AAEL000582-ECFP-both-F	GTGAGGGTGGTGTCGAATTAACTCTT
AAEL000582-ECFP-both-R	GTTAGGTCAGAGGTATCCCTGAACAT

Figure 1.pdf

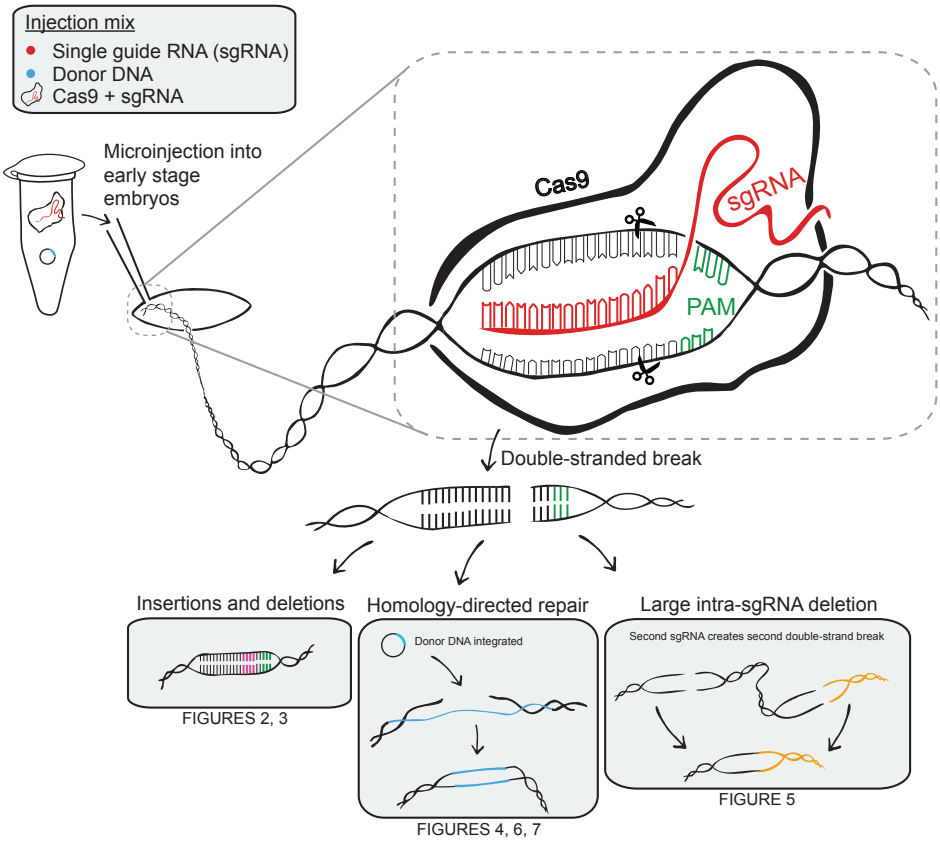


Figure 1. CRISPR-Cas9 genome-engineering system as adapted to the mosquito *Ae. aegypti*.
An sgRNA, donor DNA, and the nuclease Cas9 are microinjected into *Ae. aegypti* embryos. A 17-20 bp sequence of the sgRNA binds to the complementary site in the genome, targeting Cas9 and generating a double-stranded break. These breaks can be repaired in a number of ways, including the following: insertions or deletions resulting from error-prone non-homologous end-joining, insertion of exogenous DNA sequences through homology-directed repair, and generating large deletions between multiple sgRNA/Cas9 double-stranded breaks.

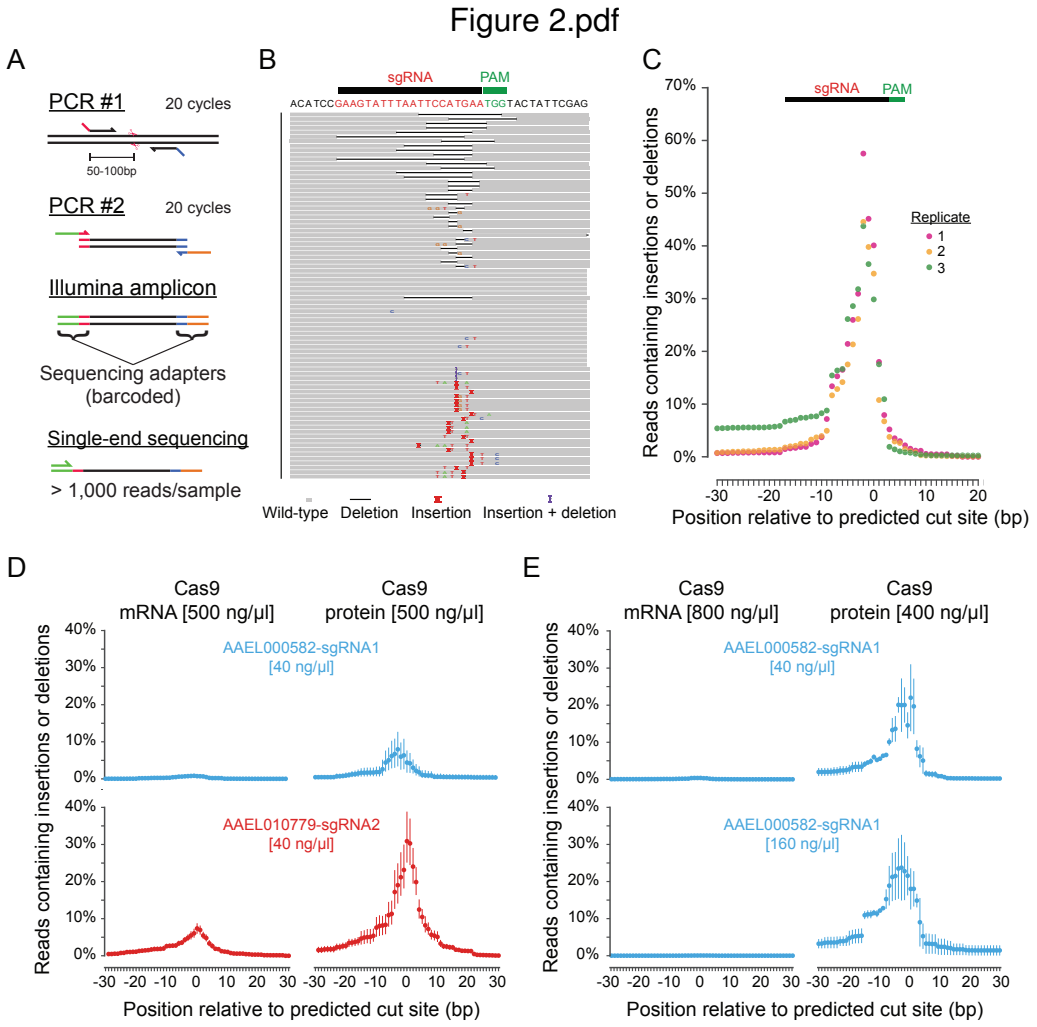


Figure 2. Deep sequencing approach to quantifying CRISPR-Cas9 efficiency
A) Schematic of strategy to generate barcoded PCR amplicons for sequencing. In this example (B-C) amplicons were generated from adults reared from embryos injected with Cas9 and an sgRNA (AAEL004091-sgRNA1) using primers AAEL004091-1-F and AAEL004091-1-R. **B)** Visualization of a subset of alignments to the amplicon reference sequence (top) reveals a heterogeneity of insertions, deletions, and other changes. **C)** Quantification of three replicate libraries, plotted as a percentage of reads aligned to a given base that contain an insertion or deletion at that base, shows reproducibility of mutation rates between individual pools of pupae derived from the same injection. Data are represented as a percentage of reads aligned to a given base that contain an insertion or deletion at that base, and is plotted as mean (circle) and 95% confidence intervals (line). **D)** Summary of sequencing data from animals injected with two different sgRNAs in combination with Cas9 mRNA (left) or Cas9 recombinant protein (right), both at 500 ng/µL. **E)** Increasing Cas9 mRNA concentration to 800 ng/µL (left) did not improve cut rate in conjunction with an sgRNA at 40 or 160 ng/µL. Reducing Cas9 protein concentration to 400 ng/µL (right) increased insertion or deletion rate slightly when combined with an sgRNA at 40 ng/µL. Increasing sgRNA concentration to 160 ng/µL did not dramatically affect the insertion or deletion rate.

Figure 3.pdf

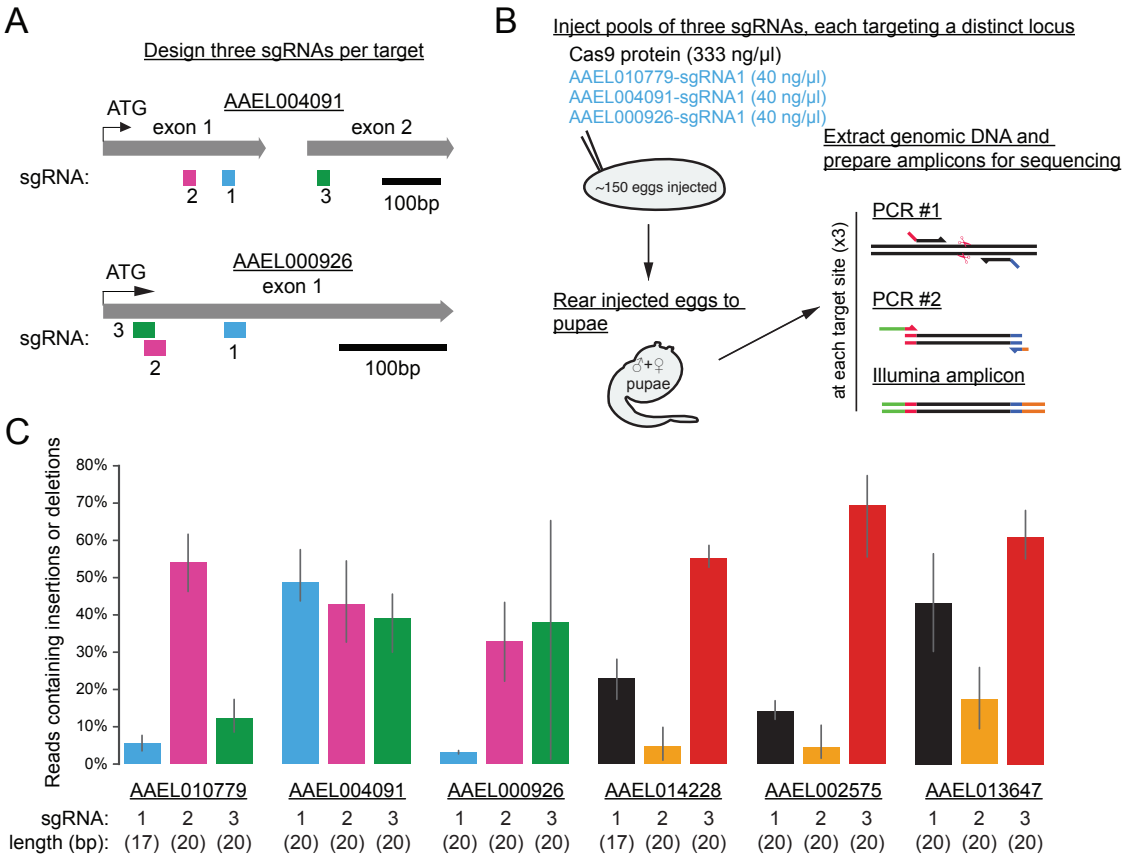


Figure 3. Identifying active sgRNAs
A) Schematic of two target genes in the *Ae. aegypti* genome: AAEL004091 and AAEL000926. Three sgRNAs were designed against the first 500 bp of each gene. **B)** Schematic of the workflow for a small injection (approx. 150 embryos) of Cas9 protein and a pool of three sgRNAs against three distinct target genes. After rearing to pupal stages to synchronize developmental time, genomic DNA was extracted and sequencing amplicons were prepared against each of the three targets. **C)** Sequencing results from 6 small injections (sgRNA sequences can be found in Table 1).

Figure 4.pdf

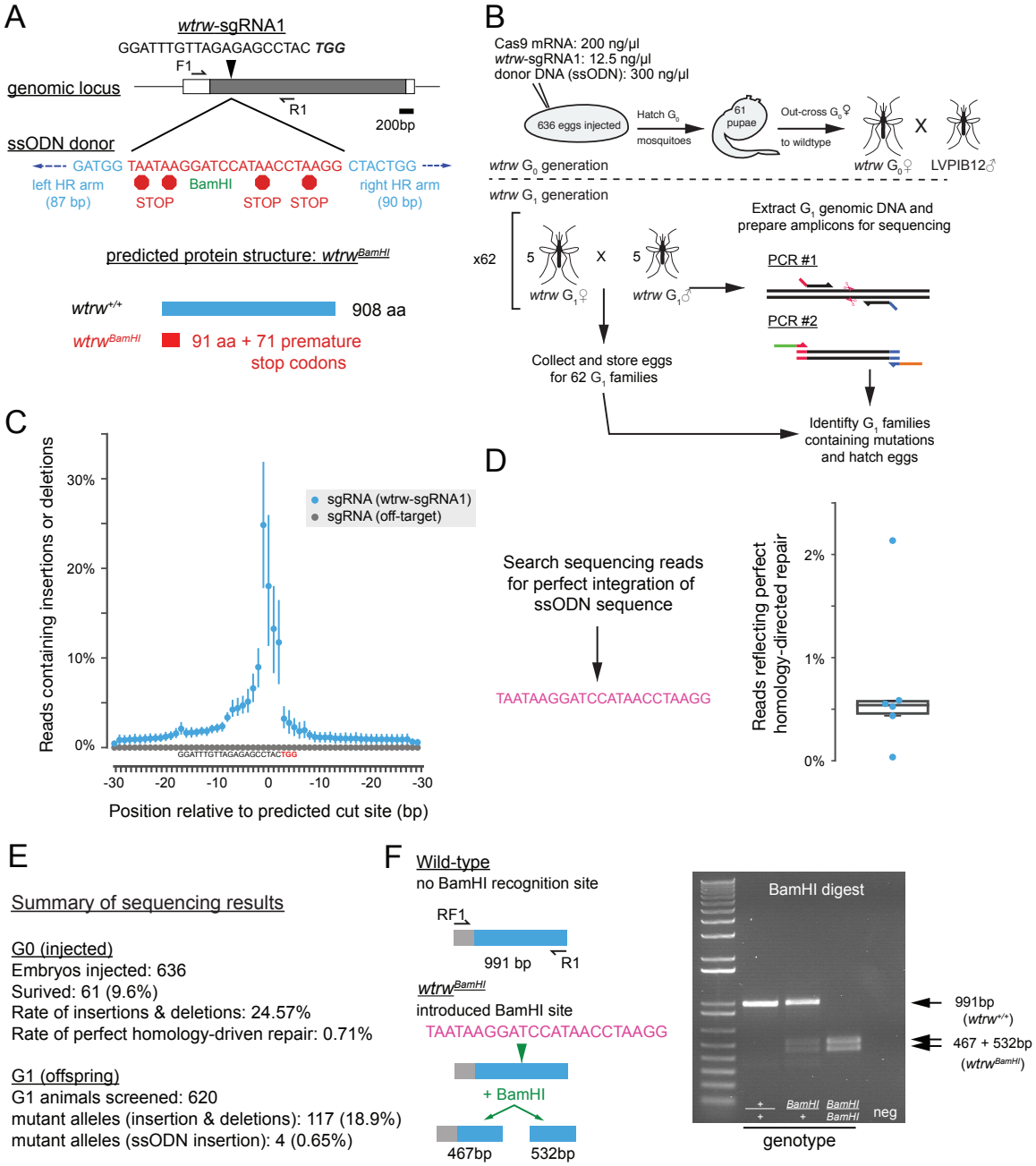
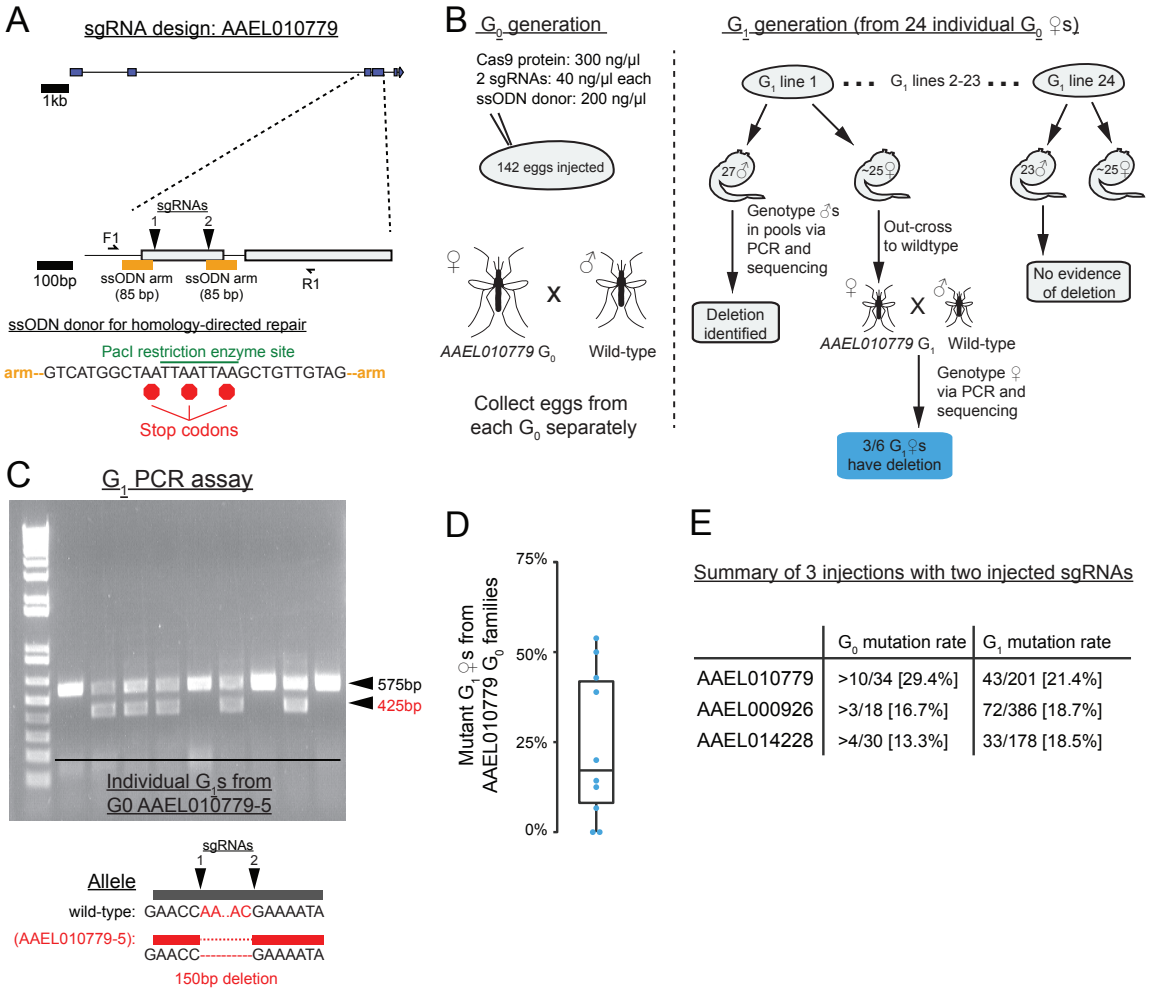


Figure 4. Germ-line transmission of mutant alleles
A) Schematic of the *Ae. aegypti* wtrw locus detailing the sgRNA binding site wtrw-sgRNA1, the sequence of a 200 bp single-stranded oligonucleotide (ssODN) donor, and the predicted protein structure of a successfully modified locus. **B)** Schematic of injection performed to isolate mutations. Sequencing libraries were generated from both G_0 and G_1 animals after egg-laying. **C)** Summary sequencing data from G_0 adults reveals robust levels of insertion and deletion in injections of wtrw-sgRNA1 and Cas9 mRNA. **D)** Exogenous ssODN sequence used to as a query for the unix tool grep to identify sequencing reads containing perfect homology-directed repair. A mean of 0.71% of reads contained this insertion. **E)** Summary statistics of sequencing reads from G_0 and G_1 individuals. **F)** Wild-type amplicons contain no BamHI recognition site while mutant alleles are cut into bands of 467 and 532 bp.

Figure 5.pdf



	G_0 mutation rate	G_1 mutation rate
AAEL010779	>10/34 [29.4%]	43/201 [21.4%]
AAEL000926	>3/18 [16.7%]	72/386 [18.7%]
AAEL014228	>4/30 [13.3%]	33/178 [18.5%]

Figure 5. Deletions generated by multiple sgRNAs
A) Schematic of the AAEL010779 genomic locus detailing the design of two sgRNAs and a ssODN donor. **B)** Injection strategy to identify deletion events in G_1 animals with the injection of a small (125-150) number of embryos. **C)** Example agarose gel of 9 G_1 females that are the offspring of a single G_0 female. All individuals contain the wild-type band (black arrowhead). Sanger sequencing verified that the smaller band (red arrowhead) present in 5/9 heterozygous individuals is the result of a clean deletion between the two sgRNA cut sites. **D)** Plot of the proportion of mutant AAEL010779 G_1 females from the 10 G_0 families identified as containing at least one mutant allele. **E)** Summary data from 3 injections of this type show high rates of G_1 mutagenesis across all 3 injections.

Figure 6.pdf

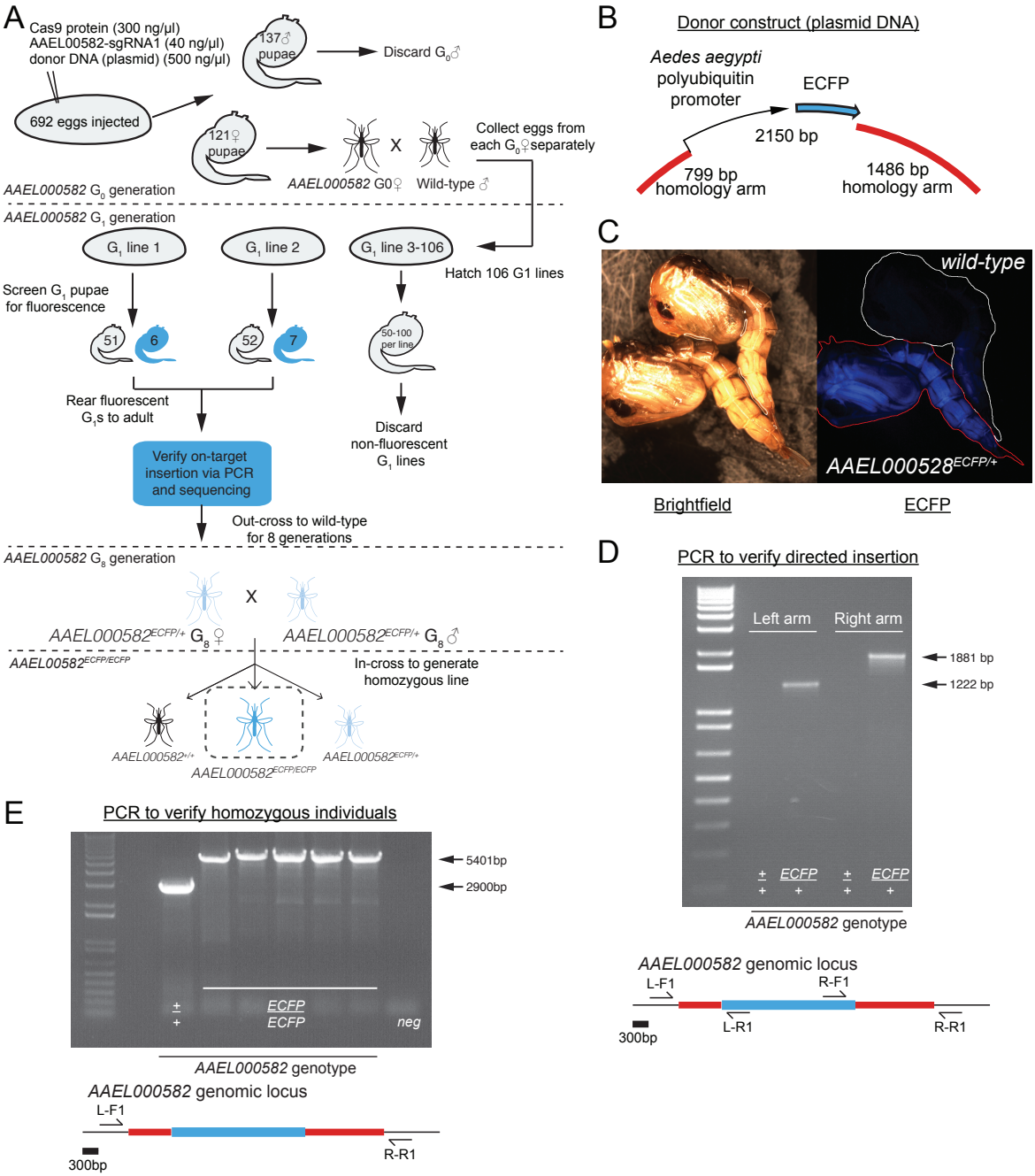


Figure 6. Insertion of fluorescent cassettes by homology-directed repair
A) Schematic of injection and screening strategy to obtain alleles with an insertion of a fluorescent cassette. Blue pupae represent animals with ECFP expression. **B)** Design of the plasmid donor with homology arms of 799 and 1486 bp surrounding a 2150 bp payload of the PUb promoter driving ubiquitous expression of EFCP. **C)** Brightfield and ECFP fluorescence images of two pupae: wild-type (top) and AAEL000582ECFP/+ (bottom). PUb-ECFP is clearly visible in both larval and pupal stages. **D)** PCR strategy to verify directed insertion of the PUb-ECFP cassette. **E)** PCR strategy to identify homozygous individuals.

Figure 7.pdf

A

AAEL000528-sgRNA1: injection 1

Embryos injected: 1037
Survival: 42 (4.1%)
Screened (♀ G_0 only): 13
Fluorescent: 4 (30.8%)
z: 4 (30.8%)

Cas9 protein: 500 ng/µl
sgRNA: 40 ng/µl
dsDNA donor: 500 ng/µl

HR arm size: 799 bp (L), 1486 bp (R)
Payload size: 2150 bp

AAEL000528-sgRNA1: injection 2

Embryos injected: 692
Survival: 258 (37.3%)
Screened (♀ G_0 only): 101
Fluorescent: 2 (2.0%)
Directed insertions: 2 (2.0%)

Cas9 protein: 300 ng/µl
sgRNA: 40 ng/µl
dsDNA donor: 500 ng/µl

HR arm size: 799 bp (L), 1486 bp (R)
Payload size: 2150 bp

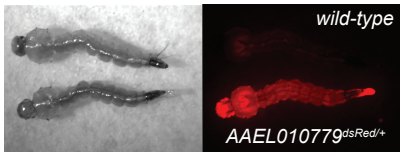
B

AAEL010779-sgRNA2

Embryos injected: 1016
Survival: 518 (51.0%)
Screened (♀ G_0 only): 106
Fluorescent: 2 (1.9%)
Directed insertions: 1 (0.9%)

Cas9 protein: 333 ng/µl
sgRNA: 40 ng/µl
dsDNA donor: 500 ng/µl

HR arm size: 2049 bp (L), 2058 bp (R)
Payload size: 2321 bp



Brightfield dsRed

C

AAEL004091-sgRNA1

Embryos injected: 1071
Survival: 153 (14.3%)
Screened (♀ G_0 only): 85
Fluorescent: 3 (3.5%)
Directed insertions: 0 (0%)

Cas9 protein: 300 ng/µl
sgRNA: 40 ng/µl
dsDNA donor: 500 ng/µl

HR arm size: 766 bp (L), 1153 bp (R)
Payload size: 2321 bp (R)

D

AAEL014228-sgRNA3

Embryos injected: 1074
Survival: 227(21.1%)
Screened (♀ G_0 only): 41
Fluorescent: 0 (0%)
Directed insertions: 0 (0%)

Cas9 protein: 300 ng/µl
sgRNA: 40 ng/µl
dsDNA donor: 500 ng/µl

HR arm size: 1005 bp (L), 1373 bp (R)
Payload size: 2321 bp (R)

Figure 7. Summary of injections to insert fluorescent cassettes by homology-directed repair
Table of injection statistics for: **A)** Two successful injections to generate fluorescent insertions in AAEL000582. **B)** One successful injection to generate a fluorescent insertion in AAEL010779. At right, brightfield and dsRed images of a wild-type (top) and AAEL010779dsRed larva (bottom). **C)** One unsuccessful injection attempt to generate a fluorescent insertion in AAEL004091. **D)** One unsuccessful injection attempt to generate a fluorescent insertion in AAEL014228.