Behavior of homing endonuclease gene drives targeting genes required for viability or female fertility with multiplexed guide RNAs

Georg Oberhofer[a], Tobin Ivy[a], and Bruce A. Hay[a,1]

[a]Department of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125

Edited by Dana Carroll, University of Utah School of Medicine, Salt Lake City, UT, and approved August 17, 2018 (received for review March 28, 2018)

A gene drive method of particular interest for population suppression utilizes homing endonuclease genes (HEGs), wherein a site-specific, nuclease-encoding cassette is copied, in the germline, into a target gene whose loss of function results in loss of viability or fertility in homozygous, but not heterozygous, progeny. Earlier work in *Drosophila* and mosquitoes utilized HEGs consisting of Cas9 and a single guide RNA (gRNA) that together target a specific gene for cleavage. Homing was observed, but resistant alleles immune to cleavage, while retaining wild-type gene function, were also created through nonhomologous end joining. Such alleles prevent drive and population suppression. Targeting a gene for cleavage at multiple positions has been suggested as a strategy to prevent the appearance of resistant alleles. To test this hypothesis, we generated two suppression HEGs in *Drosophila melanogaster* targeting genes required for embryonic viability or fertility, using a HEG consisting of CRISPR/Cas9 and gRNAs designed to cleave each gene at four positions. Rates of target locus cleavage were very high, and multiplexing of gRNAs prevented resistant allele formation. However, germline homing rates were modest, and the HEG cassette was unstable during homing events, resulting in frequent partial copying of HEGs that lacked gRNAs, a dominant marker gene, or Cas9. Finally, in drive experiments, the HEGs failed to spread due to the high fitness load induced in offspring as a result of maternal carryover of Cas9/gRNA complex activity. Alternative design principles are proposed that may mitigate these problems in future gene drive engineering.

gene drive | suppression HEG | CRISPR/Cas9 | gRNA multiplexing

Gene drive occurs when particular alleles are transmitted to viable, fertile progeny at rates greater than those due to Mendelian transmission. The possibility of using gene drive to bring about population suppression has long been of interest. One such strategy creates conditions in which super-Mendelian transmission results in the spread of a fitness cost through the population. Homing endonuclease genes (HEGs) have been proposed as a vehicle for creating and, at the same time, spreading such a fitness cost (1). A HEG encodes a site-specific endonuclease. When the HEG is located within its target site in a heterozygous individual, HEG-induced DNA double-strand breakage at the target site on the wild-type chromosome can lead to repair through homologous recombination, in which the HEG-bearing chromosome is used as the template for repair. This results in the HEG being copied into the broken chromosome in a process referred to as "homing," thereby resulting in an increase in HEG frequency. Naturally occurring HEGs use this transmission distortion mechanism to spread through populations. These HEGs utilize a variety of methods, such as inteins and self-splicing introns, to bring about copying into highly conserved sequences without altering the function of the (invariably essential) genes into which they insert (2). In this way, they are able to spread in a population with little or no cost to those who carry them.

Burt (1) proposed that HEGs could also be used to bring about population suppression by spreading a fitness cost through the population. In this scenario, homing would be limited to the germline and would bring about disruption of genes whose homozygous, but not heterozygous, loss of function resulted in inviability or sterility. Modeling suggests that such drive elements could spread and, as a result, bring about population suppression under a variety of conditions (1, 3–5). However, because homing requires the targeting and cleavage of a specific sequence, its efficacy is sensitive to genomic sequence variation. Variation can occur as preexisting sequence polymorphisms in a population. It can also arise from mutation, and as a result of break repair through nonhomologous end joining (NHEJ), which is error prone (6, 7). Regardless of the mechanism, sequence variants that are not cleaved are resistant to homing and may retain some or complete wild-type gene function. The presence of such resistant alleles can block HEG spread, and thereby prevent population suppression (3, 4, 8–11). Thus, the question of how to bring about high-frequency homing that is gene specific, but insensitive to some level of sequence variation within the gene, is central to the development of HEG-based population suppression technologies.

Germline homing and transmission of naturally occurring and engineered HEGs into artificial target sites, and of engineered HEGs into endogenous sites, have been demonstrated recently in mosquitoes and *Drosophila* (7, 12–22). In particular, a number of recent experiments have demonstrated significant rates of germline homing using HEGs created using the CRISPR/Cas9 endonuclease system, in which the Cas9 endonuclease is targeted to specific sequences through association with an independently expressed guide RNA (gRNA). The gRNA includes

Significance

Homing endonuclease gene (HEG)-based gene drive can bring about population suppression when genes required for viability or fertility are targeted. However, these strategies are vulnerable to failure through mechanisms that create alleles resistant to cleavage but that retain wild-type gene function. We show that resistance allele creation can be prevented through the use of guide RNAs designed to cleave a gene at four target sites. However, homing rates were modest, and the HEGs were unstable during homing. In addition, use of a promoter active in the female germline resulted in levels of HEG carryover that compromised the viability or fertility of HEG-bearing heterozygotes, thereby preventing drive. We propose strategies that can help to overcome these problems in next-generation HEG systems.

Author contributions: G.O. and B.A.H. designed research; G.O. performed research; G.O., T.I., and B.A.H. analyzed data; and G.O., T.I., and B.A.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: haybruce@caltech.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1805278115/-/DCSupplemental.

Published online September 17, 2018.

an ~20-nt protospacer sequence that mediates RNA/DNA base pairing-dependent target selection. Target sequence limitations with Cas9 are very modest; thus, Cas9 and gRNAs can be used to uniquely target most positions in any genome, making them ideal tools for HEG engineering (23–26).

Multiplexing of gRNAs in a Cas9-based HEG so as to target multiple positions within a gene has been suggested as a way to overcome the problem of resistant allele formation (10, 11, 20, 26, 27). The importance of this problem is highlighted by several observations. First, sequence polymorphisms are common in some populations, such as those of mosquitoes (28, 29). Second, germline homing experiments that relied on cleavage at a single Cas9 target site (one gRNA) resulted in the NHEJ-mediated creation of resistant alleles at high rates in the germline, as well as in zygotes as a result of maternal carryover of active Cas9/gRNA complexes (11, 14–17, 20–22). The consequences of this were demonstrated in homing-based population suppression experiments carried out in mosquitoes, in which genes required for female fertility were targeted by a Cas9-based HEG. HEG frequency initially increased, but this was accompanied by the creation of resistant alleles. Over multiple generations, these increased in frequency, while those of the HEG decreased (9, 22). Somewhat similar dynamics were also recently reported in the context of a homing-based sex conversion drive model in *Drosophila* (21).

These observations argue that single gRNA-based HEGs are unlikely to be useful in any real-world context due to resistant allele formation. Instead, the important question is whether the use of multiple gRNAs can prevent the formation of resistant alleles, while still promoting high levels of homing. To address this question, in the context of a suppression HEG that could, in principle, be adapted for a pest invasive species, such as *Drosophila suzukii*, we developed two multi-gRNA HEGs designed to bring about population suppression by homing into a gene required for embryonic viability or female fertility in *Drosophila melanogaster* and explored their behavior. Each HEG included four gRNAs designed to cleave the gene of interest. We determined the homing and cleavage rates of these elements, and characterized the products generated postcleavage on a molecular level. Cleavage rates were high. Homing rates were modest in general, varied greatly between individuals, and were associated with HEG instability. Importantly, cleavage-resistant alleles were not observed. Reduced resistance allele formation was also recently reported in *Drosophila* in a two-guide scenario in which nonessential genes were targeted (20). Finally, in drive experiments, we observed that after an initial increase in allele frequency, the HEGs proceeded to drop out of the populations. Possible reasons for this are discussed.

Results

HEG Design. We selected two recessive target genes for our HEG constructs, *yellow-g* (*yg*) and *deformed* (*Dfd*). The ortholog of *yellow-g* has been used previously as a target gene for a suppression HEG in mosquitoes (16). The *yellow-g* gene is somatically expressed in follicle cells of the ovary and encodes a protein required for egg shell formation. Loss of *yellow-g* results in female sterility due to collapse of laid eggs (30). The *Dfd* gene encodes a homeobox (HOX) family transcription factor expressed in the early embryo head anlagen, and is essential for posterior head morphogenesis, with loss of *Dfd* resulting in embryonic lethality (31). Each HEG included Cas9 and four gRNAs (Fig. 1*A*). Cas9 expression was driven by a *nanos* regulatory element, which drives germline expression in males and females (32). This transgene was flanked by gypsy insulators to ensure germline-specific expression. The gRNA expression constructs consisted of two tandem pairs, in which one gRNA of each pair was expressed under the control of the *Drosophila* U6:3 regulatory sequences and the other was expressed under the control of U6:1 regulatory sequences (33). To increase gRNA expression levels, the gRNA scaffolds were modified to eliminate cryptic termination sequences, as detailed previously (34). One U6:3-gRNA/U6:1-gRNA tandem was placed to the left of Cas9, while the other was placed to the right. Homology arms of ~1 kb were included to the left and right of the two outermost gRNAs to promote homologous recombination. Finally, the HEG contained a *3xP3-td-tomato* dominant marker (Fig. 1*A*).

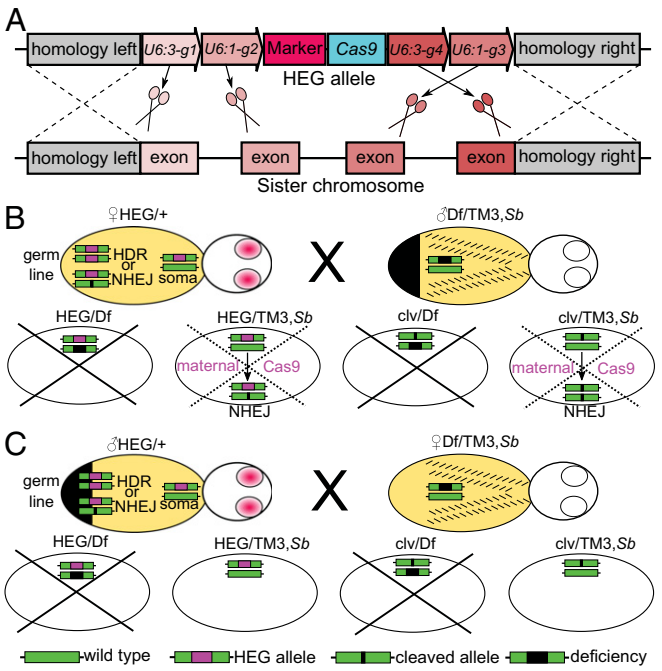

Fig. 1. HEG design and possible cross outcomes for the HEG targeting *Dfd*. (*A*) Design of the CRISPR-based HEG construct. The HEG element consists of two outer homology arms to mediate HDR with the wild-type allele on the homologous chromosome. Between the arms are a set of four U6 (alternating U6:3 and U6:1) driven gRNAs, Cas9 driven by germline-specific *nanos* regulatory elements, and a *3xP3-td-tomato* marker. Cross outcomes are shown for HEG heterozygotes for the *Dfd*-HEG crossed to a Df line, with the HEG coming from a female (*B*) or male (*C*). (*Upper*) Shown are parents of the crosses and genotypes of progeny (ovals). In the germline, the wild-type allele was either converted to the HEG allele by HDR or to a cleaved allele (clv) via NHEJ or incomplete homing. Parental genotypes were HEG heterozygotes (HEG/+) and heterozygotes for a stock carrying a Df that removes *Dfd* or *yellow-g* and the third chromosome balancer TM3 (Df/TM3,*Sb*). (*Lower*) Genotypes of the offspring were HEG allele over Df (HEG/Df), HEG allele over balancer (HEG/TM3,*Sb*), cleaved allele over Df (clv/Df), cleaved allele over balancer (clv/TM3,*Sb*), noncleaved allele over Df (+/Df, not shown), and noncleaved allele over balancer (+/TM3,*Sb*; not shown). Genotypes of interest in the embryos are depicted as ovals, and lethal genotypes (or sterile female genotypes in the case of *yellow-g*) are crossed out (dashed crosses if due to maternal carryover). (*B*) When the HEG was transmitted through the female germline, all embryos had Cas9 and gRNA deposited during oogenesis, leaving all embryos with active HEG components. (*C*) No paternal carryover of Cas9 to the embryos was observed. All of the offspring that inherited a wild-type allele from the Df strain (from the balancer TM3,*Sb*) were viable in the case of the *Dfd*-HEG or fertile with the *yg*-HEG.

Homing and Cleavage Rates of HEG Strains. Homing and cleavage, followed by inaccurate repair through end-joining pathways, result in the creation of loss-of-function alleles of the target genes. We scored for the presence of these events by crossing HEG heterozygous males and females (HEG/+) to flies heterozygous for a deficiency (Df) that removes the target locus and a version of the TM3,*Sb* balancer chromosome carrying the dominant marker *Sb* (Df/TM3,*Sb*). By comparing the frequency of specific outcomes in progeny that carried the Df chromosome versus the TM3,*Sb* balancer chromosome, we were able to distinguish between homing

and cleavage events. The presence of homing events, or the original HEG chromosome, was inferred by the presence of the *td-tomato* dominant marker. Cleavage and inaccurate repair without homing were inferred based on the presence of a loss-of-function phenotype in the absence of the *td-tomato* marker (Fig. 1 *B* and *C*). To get an overview of the frequency of each event type and interindividual variability, we set up 25 single fly crosses of each sex for both of the HEG constructs. Homing rates were calculated from the relative proportion of HEG/TM3,*Sb* to [+ or cleaved (clv)]/TM3,*Sb* progeny. Cleavage rates were calculated based on the frequency of surviving or fertile +/Df individuals, which represent either resistant alleles or noncleaved alleles. As discussed below, molecular analysis showed that no completely resistant (lacking target sites for all four gRNAs) alleles were formed (details and formulas are provided in *Methods*, and counts are provided in Dataset S1).

When the *Dfd*-HEG came from a mother, most progeny died, presumably due to maternal carryover and subsequent cleavage of *Dfd*. With this low number of surviving progeny, it was not possible to calculate meaningful homing rates (counts are provided in Dataset S1). For the remaining crosses, results are shown in Fig. 2 *A*–*C*. Interestingly, homing rates for both HEGs showed high variability between individual single fly crosses, ranging from 0 to 83%, while cleavage rates were consistently very high [average homing and cleavage rates as mean (SD)]: ♂*Dfd*-HEG/+ homing = 0.33(0.21), cleavage = 0.99(0.02); ♂*yg*-HEG/+ homing = 0.19(0.17), cleavage = 1(0.00); and ♀*yg*-HEG/+ homing = 0.26(0.20), cleavage = 0.89(0.31).

Fitness Costs from Germline Carry over. In order for suppression HEGs to spread, cleavage and homing must be restricted to the adult germline. Otherwise, heterozygotes and/or their progeny will show loss-of-function phenotypes, which, by definition, include a fitness cost. Maternal carryover of Cas9/gRNA complexes into non–Cas9-bearing individuals has been shown to result in high-frequency cleavage of target sequences in *Drosophila* (17, 20, 35). Similar effects can also be inferred from the results of other work in *Drosophila* and mosquitoes that utilized promoters driving Cas9 in the female germline (14–17). The frequency of cleavage in the zygote due to expression in the male germline is low to nonexistent (20, 36). Here, we address the significance of carryover in the context of HEGs designed to induce a fitness cost.

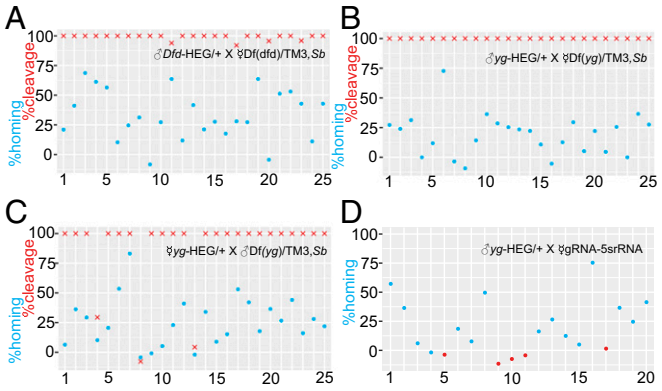

Fig. 2. (*A*–*C*) Homing rates and cleavage rates of HEG-bearing heterozygotes crossed to a Df line. Homing rates are denoted as blue dots, and cleavage rates are denoted as red crosses on the *y* axis, with the number of each individual cross indicated on the *x* axis. Homing rates were highly variable between crosses, whereas cleavage rates were 100% for most of the crosses. (*D*) Test crosses to determine Cas9 activity in the *yg*-HEG flies. Shown are the homing rates of heterozygous *yg*-HEG males crossed to a gRNA strain targeting the 5s rRNA that results in sterility in all offspring when crossed to a strain with functional germline-specific Cas9. Homing rates varied, as with those from crosses to the Df lines. In five of 20 crosses, the Cas9-carrying progeny were fertile (red dots) and no homing occurred, indicating that Cas9 in the HEG strain had lost functionality, even though the HEG cassette dominant *td-tomato* marker was still present. (Note that since homing is calculated based on deviation of progeny genotype ratios from those of Mendelian transmission, homing rates of less than zero can occur. These likely reflect cases in which no homing occurred and genotypes due to Mendelian segregation were not balanced.)

For the *Dfd*-HEG, we were unable to determine the extent of male carryover-dependent cleavage in the zygote because progeny genotypes that include a homed or cleaved paternal allele and a cleaved maternal allele die during embryogenesis. For the *yg*-HEG, we determined the frequency of carryover-dependent cleavage from the germline of male *yg*-HEG/+ individuals crossed to Df/TM3,*Sb* females by determining the frequency of TM3,*Sb*-bearing female progeny (*yg*-HEG/TM3,*Sb*; +/TM3,*Sb*; and clv/TM3,*Sb*) that were sterile. None should be sterile if carryover-dependent cleavage did not occur. One hundred forty-one TM3, *Sb* female progeny of the first five crosses shown in Fig. 2*B* were tested. Of these, only three (2.1%) did not produce offspring, suggesting that carryover-dependent cleavage is minimal. Based on this observation, we assume below that carryover-dependent cleavage in the zygote is specific to females.

For the *Dfd*-HEG, we estimated fitness cost indirectly, under the assumption that crosses between HEG/+ and Df(*Dfd*)/TM3, *Sb* adults would, in the absence of carryover into the embryo, produce equal numbers of progeny regardless of which parent carries the HEG for the *Dfd* locus. With this assumption, and in the absence of carryover, the number of TM3,*Sb*-bearing progeny (HEG/TM3,*Sb* and +/TM3,*Sb*) should be similar, regardless of whether the HEG came from the mother or father. In actuality, when the HEG came from the mother, the total number of adult TM3,*Sb*-bearing progeny from all 25 crosses shown in Fig. 2 was 97 and large numbers of dead eggs were observed. In contrast, when the HEG came from the male, 1,218 TM3,*Sb* progeny were obtained, suggesting a maternal carryover-dependent fitness cost of 92%. In the case of the *yg*-HEG, we calculated the fitness cost due to maternal carryover directly, as the number of sterile females (HEG/TM3,*Sb* and clv/TM3,*Sb*; $n = 405$) divided by the sum of all females with the balancer chromosome (HEG/TM3,*Sb*; clv/TM3,*Sb*; and +/TM3,*Sb*; $n = 489$), suggesting a maternal carryover-dependent fitness cost of 82.8%. Together, these results support earlier observations (17, 20) and demonstrate that when using the *nanos* regulatory sequences, maternal carryover-dependent cleavage of paternal alleles in the zygote is common and fitness costs to HEG-bearing zygotes are high. The consequences of these fitness costs for gene drive are explored further below.

Molecular Characterization of Cleavage Events. Because most progeny of *Dfd*-HEG–bearing heterozygous mothers died during embryogenesis as a result of carryover-dependent cleavage, we were unable to further characterize these individuals. However, since *yellow-g* is only required for eggshell formation in somatic ovarian cells of adult females, female clv/Df(*yg*) offspring of the ♀*yg*-HEG/+ X ♂Df(*yg*)/TM3,*Sb* cross and the ♂*yg*-HEG/+ X ♀Df(*yg*)/TM3,*Sb* cross are viable, but sterile. Genomic DNA was isolated from 10 sterile female cross progeny (two flies each from five different vials) where the HEG came from a female (F1–10) and from 10 where it came from a male (M1–10). All flies analyzed lacked the *td-tomato* marker, indicating that a complete homing event had not occurred. However, these females were sterile, suggesting their genotype was clv/Df. The target region was amplified by PCR and sequenced. Since the Df-bearing chromosome lacks the *yellow-g* locus, all products must come from a cleaved or partially homed locus.

We found a wide range of different cleavage events (Fig. 3). Nine flies showed deletions between the four gRNAs. In this category, five flies had a deletion between the outer gRNA1 and gRNA4 (F1, F3, F8, M5 with additional bases deleted, and M7).

GENETICS

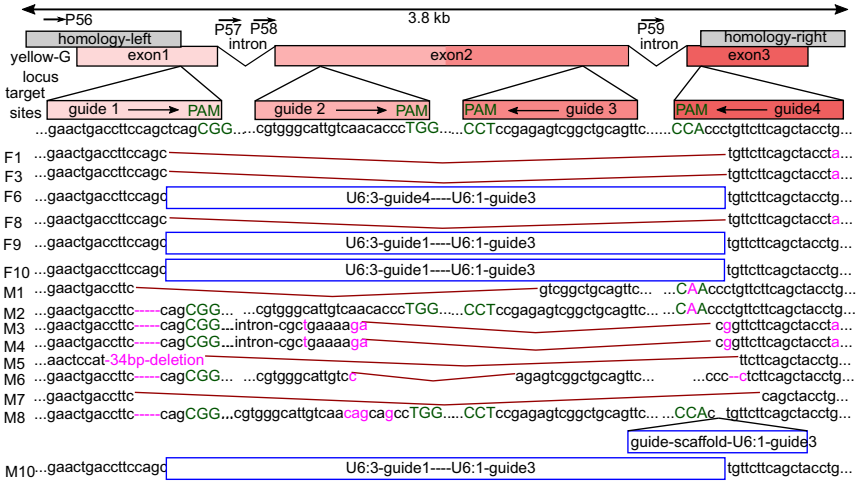

Fig. 3. Cleavage events. Sanger sequencing results from individual sterile progeny of the clv/Df(*yg*) genotype, in which progeny come from ♀*yg*-HEG/+ × ♂Df(*yg*)/TM3,*Sb* (F1–F10) or from ♂*yg*-HEG/+ × ♀Df(*yg*)/TM3,*Sb* (M1–M10). Flies were taken in pairs from individual crosses (i.e., F1 and F2 coming from one vial, F3 and F4 coming from another vial, and so on). PAM (green), mutations and small deletions (magenta), large deletions between gRNA target sites (brown lines) and HEG cassette remnants due to incomplete homing (blue) are shown. The 5-bp deletion in M2–M4, M6, and M8 was likely caused by microhomology-mediated end joining repair of the CAG microhomology on either side of the Cas9-induced double-strand breakage (DSB). The figure is not drawn to scale. Sequencing primer binding sites are ∼700 bp apart and indicated as small arrows (P56–P59).

One had a deletion between gRNA1 and gRNA3 (M1 with PAM of gRNA4 mutated). Two had a deletion between the intron located between gRNA1s and gRNA2 and gRNA4 (M3, M4, coming from the same parents), and one had a deletion between gRNA2 and gRNA3 (M6). Finally, one fly had only a small deletion at the gRNA1 target site (M2, which also had a mutated PAM for gRNA4).

For the remaining 10 flies analyzed, we detected incomplete or partial homing events (*SI Appendix*, Fig. S3). Three of these flies had half of the left and half of the right guide cassette inserted between gRNA1 and gRNA4 cut sites (F9, F10, and M10). One had the complete right guide cassette inserted between gRNA1 and gRNA4 (F6). One had half of the right guide cassette inserted into the cut site of gRNA4 (M8). For the remaining five flies, we were not able to generate PCR products that spanned the two outer cut sites. However, we could amplify a part of the Cas9 sequence. Since these flies lacked the *td-tomato* marker but retained other parts of the HEG cassette (at least some of Cas9), they were also considered to represent incomplete homing events (F2, F4, F5, F7, and M9). Together, these results show that the consequences of Cas9-mediated cleavage are diverse and include a significant frequency of incomplete homing events.

HEG Cassette Stability. Given the high frequency of incomplete homing events, wherein the *td-tomato* marker was lost but Cas9 was still present, we were interested to determine if incomplete homing events also resulted in events in which Cas9 was lost but the *td-tomato* marker was retained. To test this possibility, we isolated males that were *td-tomato*–positive (*yg*-HEG–bearing) at generation 10 from a drive experiment (discussed below), in which the *yg*-HEG had been introduced into a wild-type population at a frequency of 25%. These males were outcrossed to *w*- females to ensure heterozygosity for the HEG. Twenty male heterozygous progeny from these crosses were mated to females of a Cas9 tester strain that expresses, under the control of U6:3 regulatory sequences, a gRNA designed to bring about cleavage of the 5s ribosomal RNA repeats (gRNA-5srRNA) (Fig. 2*D*). When gRNA-5srRNA females are crossed to males expressing Cas9 under the control of *nanos* regulatory sequences, all offspring that carry both gRNA-5srRNA and Cas9 should be sterile, allowing sterility to be used as a test for the presence of functional Cas9. Progeny from the above 20 crosses were scored for the *td-tomato* marker to determine homing rates, and all gRNA-5srRNA/*td-tomato*–positive females were outcrossed to *w*- males to score for fertility. The results are shown in Fig. 2*D* (counts are provided in Dataset S1). Females were fertile in five of the 20 crosses, indicating that Cas9 function had been lost. As expected, no homing was observed in these crosses. The rates of homing for the remaining crosses varied within a range similar to the one observed in crosses to the Df-bearing lines (Fig. 2 *A*–*C*).

Characterization of Escapers. As discussed above and in Fig. 2, the cleavage rates of target loci in the male and female germline of HEG/+ adults was very high for both *Dfd*-HEG and *yg*-HEG. However, in rare cases, the cross between HEG/+ and Df/TM3, *Sb* adults resulted in progeny that carried the Df chromosome (+/Df or clv/Df), but were viable (*Dfd*) or fertile (*yellow-g*). To determine whether these individuals simply carry alleles that escaped cleavage or alleles that are resistant to cleavage, we isolated genomic DNA and carried out PCR and sequencing of the target region for all of them (Table 1 and *SI Appendix*, Fig. S2). For the *yg*-HEG, 24 female escapers (fertile) were isolated from three crosses in which the HEG came from the female (F4, F13, and F8). The six escapers from cross F4 had a mutated PAM at the gRNA4 target site, the 11 escapers from cross F13 had a mutation at base 1 of the gRNA4 target site, and the seven escapers from cross F8 had an intact target gRNA4 target site. All other target sites were intact in all escapers.

For *Dfd*-HEG, we found four escapers from two crosses where the HEG came from a female (F4 and F11) and in five individuals from four crosses where it came from a male (M11, M17, M20, and M23). Escapers from crosses F4 and F11 had a 3-bp in-frame deletion 3 bp upstream of the PAM of the gRNA4 target site. All other target sites in the remaining escapers were intact. To summarize, escapers did not carry resistant mutations at all four sites. This, coupled with the low frequency of escapers, as well as related work with a two-gRNA HEG system (20), shows that multiplexing can dramatically reduce the frequency of resistant alleles. However, our work also shows that multiplexing alone is not (yet) sufficient to completely prevent evasion of cleavage at intact sites. If one gRNA and its corresponding site are, for some reason, used to the exclusion of others, resistance could still arise. Possible explanations for the observed behavior are discussed below.

Gene Drive Outcomes. To assess the potential of our HEG constructs to suppress populations, we set up several drive experiments. As the seed for generation 0, we introduced *w*- females that had been mated to HEG/+ males and *w*- females mated to *w*- males. These two groups of mated females were mixed at a 1:1 ratio, corresponding to a HEG-allele introduction frequency of 25%. Four replicates were carried out for each drive experiment. In each generation, adults were allowed to lay eggs for 2 d and then removed. After progeny had matured to adulthood, we

Table 1. Escaper and resistant gRNA target sites

Cross	Total	No. of escapers	gRNA1	gRNA2	gRNA3	gRNA4
♀Dfd-HEG/+ × ♂Df(Dfd)	101	4	+	+	+	F4 (1) 3-bp del F11 (3) 3-bp del
♂Dfd-HEG/+ × ♀Df(Dfd)	1,223	5	+	+	+	M11 (2) + M17 (1) + M20 (1) + M23 (1) +
♀yG-HEG/+ × ♂Df(yG)	2,344	24	+	+	+	F4 (6) (PAM) F8 (7) + F13 (11) bp 1 of target
♂yG-HEG/+ × ♀Df(yG)	2,714	0				

Shown are the total number of progeny and number of escapers from all 100 single fly crosses to the Df lines. Intact gRNA target sites are indicated with a "+" symbol. When a mutation was observed, the cross and the number of individuals (in parentheses) are indicated, and a short description of the mutation is provided. del, deletion.

scored for the presence of the *td-tomato* marker, seeded the next round with 200 flies, and repeated the cycle. As a gene drive control, we used inactive HEG constructs that were inserted at the target locus and expressed the dominant marker and gRNAs but lacked Cas9.

The dynamics of HEG behavior are illustrated in Fig. 4. In the case of *Dfd*, the HEG increased in frequency for one generation but decreased continuously thereafter. By generation 9, the *Dfd*-HEG had been completely lost from all replicates. In contrast, the inactive *Dfd*-HEG remained in the population. In the case of *yellow-g*, the HEG increased in frequency for two generations. It then decreased in frequency, as with the *Dfd*-HEG, albeit at a slower rate. By generation 12, the HEG had been lost completely from one replicate and had fallen below the introgression frequency in the others. The control version of the *yg*-HEG increased in frequency slightly above the introgression frequency up until generation 7 before returning to around 25% at generation 12. The observed frequencies of both control HEGs remains higher than the average expected for a recessive lethal or female sterile allele (Fig. 4 *C* and *D*). The basis for this behavior is unclear and requires further exploration since these elements carry only gRNAs and a dominant marker (*flightin-td-tomato* rather than the *3xP3-td-tomato* used in the complete HEGs) inserted into the target locus.

We generated a deterministic population frequency model of our two HEGs, incorporating the observed cleavage and homing rates, as well as fitness costs due to maternal carryover-dependent cleavage (details are provided in *Methods*). Models for both HEGs fit the data well. In the case of *Dfd*, the model predicted an increase in frequency for one generation followed by a rapid decrease, as was observed (Fig. 4*A*). In the case of *yellow-g*, the model predicted a two-generation increase in the frequency of HEG-bearing individuals before a slow decline, as was also observed in the drive experiments (Fig. 4*B*). These observations suggest that the variables studied and quantified in this work, with respect to homing, cleavage, and carryover, are sufficient to capture major aspects of HEG behavior.

Discussion

Homing-based gene drive requires the targeting of specific sequences for cleavage, and is therefore sensitive to existing sequence variation or novel sequences created by error-prone repair pathways. Homing-based gene drive using Cas9 and a single gRNA has been demonstrated in *Drosophila* (14, 17, 20, 21), mosquitoes (9, 15, 16), and yeast (37). However, these studies also reported the creation of mutant alleles that are resistant to cleavage. In those experiments in which coding regions were targeted for cleavage, resistant alleles included variants in which target gene function was retained (9, 14, 17, 20, 21). The presence of such alleles is predicted to prevent population suppression when genes required for viability or fertility are targeted (3, 10, 11). Resistant alleles can also prevent homing-based population replacement using HEGs that carry a gene of interest into the population for the same reason. Thus, prevention of resistant allele formation is a central problem that must be solved in order for homing to be useful as a drive mechanism.

Multiplexing of gRNAs has been proposed as a way of overcoming this problem (10, 26, 27, 38), but it has only been tested once, very recently (20). In this study, we tested the effects of gRNA multiplexing on resistance allele formation in the context of suppression HEGs developed with components used in earlier gene drive experiments in insects. A key positive finding of our work is that when four gRNAs are used to target a gene, cleavage rates in the germline are very high and resistance allele formation at all target sites is not observed. Similar results were recently reported for a two-gRNA system in *Drosophila* targeting a nonessential gene (20). However, our work also highlights several important problems that remain to be solved. First, homing rates in the male and female germline are overall modest and vary greatly between individuals. Second, HEGs are unstable, with homing resulting in the creation of a high frequency of chromosomes that bear incomplete HEGs lacking one or more essential elements. Third, maternal carryover of a Cas9/gRNA HEG results in extensive cleavage of the paternal allele in zygotes, and this results in the

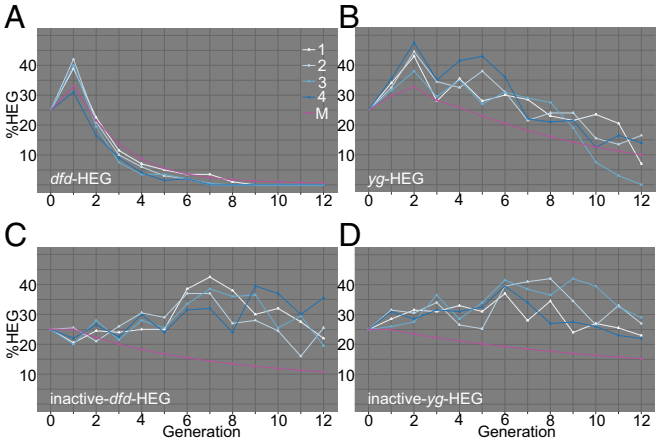

Fig. 4. Drive experiments. (*A*–*D*) Plots of drive experiments. The frequency of HEG-bearing individuals (HEG/+ and HEG/HEG) is indicated on the *y* axis, and the generation number is indicated on the *x* axis. Drive replicates are labeled as 1–4. Predicted drive behavior using fitness costs inferred from data (*A* and *B*) are shown, and the predicted behavior of a strictly Mendelian (M) recessive lethal or sterile allele (*C* and *D*) is shown in pink. Counts are provided in Dataset S2.

GENETICS

creation of HEG heterozygotes (clv/HEG) that experience dramatic fitness costs. Together, these results predict that the two HEGs generated, *Dfd*-HEG and *yg*-HEG, should fail to spread; drive experiments confirmed these predictions, with both elements initially increasing in frequency but then decreasing in frequency and/or being lost from the population. We discuss these and earlier results in more detail below and explore possible paths forward.

Cleavage Rates. For the *yg*-HEG, cleavage rates were 100% in the germline of all but three of 50 individuals (Fig. 2 *B* and *C*). In two of the three fertile females (F8 and F13), cleavage rates were likely zero (homing rates were zero), while in the third fertile female (F4), a cleavage rate of ∼26% was inferred. Low rates of cleavage in these individuals could reflect variation in the levels of Cas9, the timing of its expression, expression of gRNAs, accessibility of the target sites, or reduced activity of Cas9 at temperatures below 37 °C (39). Another possible reason for low (or no) cleavage in these individuals is suggested by our molecular analysis of homing events, which uncovered a high frequency of incomplete homing. In particular, when we examined flies in which homing of the *yg*-HEG had been permitted over 10 generations in a drive experiment (Fig. 4), 25% were found to have lost Cas9 activity but retained the dominant marker used to score for the presence of the HEG (Fig. 2*D*). Thus, some ostensibly HEG-bearing individuals tested for cleavage, such as F8, in which all target sites are wild type in sequence, may lack a functional HEG. Multiple female progeny of F4 carried the same mutation in gRNA4, and multiple female progeny of F13 also carried a distinct common mutation in gRNA4. Both mutations were also observed in our analysis of cleavage events in sterile females (Fig. 3). These results suggest that these sequence variants represent preexisting polymorphisms. Regardless of their origin, the fact that gRNA targets 1–3 were intact in these individuals indicates that cleavage did not occur at all sites. gRNA1–gRNA3 are likely to be functional, given the spectrum of cleavage products observed in sterile clv/Df females (Fig. 3). It is possible that interaction of gRNAs with target sequences is a relatively slow stochastic process, with these simply representing the rare alleles in which gRNA1–gRNA3 did not bind to DNA. It is also possible that gRNA1–gRNA3 are expressed/processed/loaded and/or promote cleavage less efficiently than gRNA4, which may be important if Cas9 levels are limiting. A possibility of particular concern, consistent with our data, is that gRNA4 is preferentially loaded. If there are also polymorphisms in the gRNA4 target site (preexisting or created through NHEJ) that prevent cleavage, the combination of these two events would result in an abundance of Cas9/gRNA complexes that are effectively dead with respect to the target gene, even though the gRNA target sites are intact and the gRNAs are expressed. It could be argued that preferential loading of specific gRNAs is unlikely since naturally occurring CRISPR arrays in prokaryotes often contain multiple protospacer sequences that are ultimately transcribed into the CRISPR-RNA (crRNA) used for Cas9 targeting of DNA (40). However, it is not clear that all crRNAs are loaded equally efficiently. Even if they are loaded equally, an important difference from our work is that the naturally occurring CRISPR arrays are, by definition, the survivors, those containing sequences that allowed them to survive attack by phage, which would require efficient crRNA loading. In contrast, the sequences we incorporate are chosen only with reference to on-target site prediction (41).

Our data on escapers of cleavage by the *Dfd*-HEG, while representing small numbers, may be more suggestive of a situation in which multiple gRNAs are often utilized. Thus, in contrast to the case of *yellow-g*, in which cleavage rates were low in females that gave rise to escapers (F4, F8, and F13), the four *Dfd*-HEG/+ males that gave rise to +/Df viable progeny still had very high cleavage rates (only one or two viable progeny per adult). The five rare escapers' progeny from these four males were wild type in sequence at all target sites, consistent with models in which cleavage simply failed to occur at any of the four sites some of the time. Little is known about how specific gRNAs are chosen from a cellular pool and what regulates the kinetics of subsequent steps. Much remains to be learned about how to ensure contemporaneous, if not simultaneous, loading and cleavage at multiple sites within a gene. The hypothetical scenario outlined above involving preferential loading of a gRNA targeting a sequence that can mutate to resistance demonstrates how easily even multiplexing-based strategies to prevent resistance development could potentially be defeated. Exploration of these and other variables, such as the levels of gRNAs and Cas9, nuclear import of Cas9, stability of particular gRNAs, and accessibility of specific loci in the germline, should provide guidance on how to maximize opportunities for multiple cleavage events.

Homing Rates. In our experiments, each HEG included four gRNAs, recognizing sequences distributed over more than 2 kb and targeting genes required for viability or fertility. Other published work in *Drosophila* targeted different genes, all nonessential. In addition, all previous work used only a single gRNA (14, 17, 20, 21), with the exception of Champer et al. (20), who also tested a HEG that included two gRNAs targeting nearby (within 100 bp) sequences. Given these many differences, elaborated on below, we do not attempt to derive general principles regarding homing rates in *Drosophila*. Instead, we compare our observations and those of others with an eye to understanding how differences in homing rates can be mechanistically explained and how homing rates of HEGs utilizing multiple gRNAs can be improved.

In contrast to the generally very high rates of germline cleavage of target loci, homing rates for individual *yg*-HEG males and females, as well as for *Dfd*-HEG males, showed great variability, ranging from 0 to 83% (mean values: ♂*Dfd*-HEG = 0.33, ♂*yg*-HEG = 0.19, and ♀*yg*-HEG = 0.26). Some variability in homing rate between individuals has also been noted previously (17, 20) in experiments targeting several different loci (*y*, *w*, and *cn*), although the variability we observe is particularly high. The average homing rate we observed is also lower than that observed previously (17, 20). The basis for this variability is unknown, but important to understand, as it seems generally unrelated to the frequency of cleavage, which is required for homing to occur. Very high homing rates were inferred in one earlier *Drosophila* study in which *y* was targeted with a HEG that lacked a dominant marker with which to follow homing independent of mutation of the *y* locus (14). However, the vasa promoter used in these experiments results in extensive maternal carryover and is active in somatic cells (17, 20), leaving open the possibility that the high rates of homing inferred may reflect, to some extent, cleavage in somatic cells of heterozygotes rather than homing.

One possible explanation for the high variability we observe in homing is that having multiple Cas9/gRNA complexes bound to the locus (but before cleavage) interferes with some aspect of homologous recombination-dependent repair initiated following cleavage at one of these sites. For example, it has been suggested that binding of one Cas9/gRNA complex could interfere with binding to a neighboring site as a result of DNA supercoiling (42). Such a mechanism might contribute to the great variability in homing rates, since which combination of target sites is occupied at any one time in a specific germ cell, and what the order of cleavage is, will vary. If such an effect does exist, it may be possible to overcome it by increasing the distance between gRNA targets and/or changing their orientation (although other possible effects, discussed below, may argue for placing cleavage sites near each other). The timing of Cas9 and gRNA expression may also matter. While we observe very high frequencies of cleavage, which indicates that Cas9 and gRNAs are expressed and active, the exact timing of this expression and its variability are unknown. The time at which cleavage occurs is likely to be

important for the choice of repair pathway. It may also be important that target loci for suppression HEGs such as *Dfd* and *yellow-g* are, by definition, not important for germline development since, otherwise, homing (unless it occurred after gene function was required) would result in loss of cells in which homing occurred. *Dfd* and *yellow-g* are not expressed at appreciable levels in the germline. Thus, it is also possible that lack of expression creates a chromatin environment in which homing steps subsequent to cleavage are rendered challenging. Finally, we note that in a multiplex design, the sequences on either side of a cut site often do not correspond to those present in the homology arms flanking the HEG (Fig. 1). They do in single gRNA designs (14–17, 20), and they also do when both outer gRNAs cleave in our multiplex design (Fig. 1); however, only one end (gRNA1 alone or + gRNA2/gRNA3, gRNA4 alone or + gRNA2/gRNA3) or no ends (gRNA2 and/or gRNA3) correspond to the homology arms in other cleavage scenarios. Our gRNAs span a region of 2.2 kb within the *yellow-g* gene and 8.8 kb within the *Dfd* locus. Thus, ends generated by cleavage at any combination of gRNAs other than the combination of gRNA1 and gRNA4 presumably require extensive resection before homing rate-dependent repair can occur. This may reduce the frequency and/or efficiency of homing rate versus other competing repair pathways in copying a complete HEG. Such a model would suggest that locating multiple target sites near each other could prevent resistance allele formation while also maintaining a high homing rate. Evidence consistent with this possibility comes from experiments in which the nonessential *w* locus was targeted by a HEG utilizing one gRNA or two gRNAs designed to bind within 100 bp of each other. Interestingly, the two-gRNA HEG had a reduced frequency of resistant alleles and an increased homing frequency compared with the single gRNA HEG (20). If the location of the break with respect to the location of homology arms is critical, it will be important to understand how to locate sites so as to minimize negative consequences for homing rate-based repair. Modeling suggests that four or more gRNAs will be needed to avoid resistant allele formation in large wild populations (10, 11, 27). It will be interesting to see if close spacing of sites targeted by four gRNAs allows for the creation of a HEG that maintains the benefits observed with two gRNAs for increased homing rate, as well as the prevention of resistant allele formation observed herein and by Champer et al. (20).

HEG Instability for a Suppression HEG. In our crosses to the Df lines, we observed the homing rate to be highly variable between replicates and low on average (19–33%). When we assayed flies carrying the *td-tomato* marker, we found that five of 20 tested had lost Cas9 activity after 10 generations in a drive experiment (Fig. 2*D*). Evidence for partial homing of Cas9 HEGs in *Drosophila* has also been observed recently by others (20, 21). One important implication of our observation is that scoring for the presence of a dominant marker located within the HEG provides only an upper estimate on the rate of complete homing events. To mitigate this problem, it may be possible to link Cas9 and the dominant marker into a single dicistronic transcript by making use of 2A-like sites or internal ribosome entry sites (43–46), although expression of the dominant marker would then be limited to the germline. A second implication is that cargo genes incorporated into HEGs may be lost at appreciable frequencies. Tests of a replacement HEG carrying a cargo gene in *Anopheles stephensi* did not report such events (15). However, it is possible (species difference notwithstanding) that this again reflects the fact that these authors used a single gRNA in which cleavage results in two ends that are identical to the homology arms flanking the HEG. Perhaps this allows for more efficient complete homing, compared with our multiplex gRNA scenario in which the ends generated by cleavage may need to be extensively chewed back to initiate homing rate with the HEG-bearing chromosome.

In our sequencing analysis of cleavage events, we also found that five of 20 analyzed flies had lost the dominant marker but retained some fraction of Cas9 (functionality not tested) and another five flies carried only remnants of the gRNA cassettes but no Cas9 nor *td-tomato* (Fig. 3). Thus, not only are the observed homing rates too low for a HEG to bring about population suppression (3, 5, 11) but the homology-directed repair (HDR) process, at least in our HEG configuration in *Drosophila*, is prone to error. Incomplete homing is not fatal in the case of a suppression HEG since it still creates a nonfunctional copy of the gene. However, it can prevent effective population replacement if the HEG is meant to carry a cargo gene into the populations (15). Finally, we also observed recombination events involving the gRNA cassettes in five of 20 analyzed events, showing that the repetitive sequences in our multipromoter gRNA design are also prone to undergo recombination, with sequences between them getting lost in the process. A similar partial loss of HEG elements was observed previously with transcription activator-like effector nuclease (TALEN)-based HEGs, which also consist of highly repetitive sequences (13). The repeats can be reduced by using polycistronic gRNAs expressed from a single promoter (47), although this may result in different expression (processing) levels for the different gRNAs (48). In addition, the gRNA scaffolds would still have some level of repetitiveness. Thus, it remains to be shown if this approach will result in more stable copying of the HEG during the HDR process.

How can the frequency of complete homing events be increased? First, the NHEJ pathway can be suppressed (49). Second, Cas9 activity can be limited to stages of the cell cycle during which HDR is the dominant repair pathway by fusing cell cycle-specific degrons to Cas9 (50, 51). It is possible that low rates of homing in *Drosophila* are (for unknown reasons) a species-specific problem and that rates in important target species, such as mosquitoes, are higher (7, 12, 13, 15, 16). However, if this latter possibility is correct, it must have a genetic basis, which will be important to understand since polymorphisms in the relevant genes would provide a basis for selection of drive suppressors. In addition, as discussed above, the relationship between the sequences at the cleaved ends and those present in the homology arms of the HEG needs to be explored so as to maximize cleavage while still allowing for a high frequency of complete homing. It will be particularly interesting to explore the consequences of locating multiple target sites within a small region (20). Finally, it may be possible to adjust (reduce) gRNA number and still prevent resistance allele formation through the careful choice of target sites that encompass highly conserved and essential sequences less able to undergo cleavage-dependent mutation to resistance and high fitness.

Spreading of the HEG Through the Female Germline Is Severely Limited. An important problem encountered in our work is maternal deposition and activity of HEG components (i.e., Cas9, gRNAs). Similar problems were observed and discussed previously, but with a focus on resistant allele generation (9, 11, 15–17, 20). However, in the case of a suppression HEG, germline carryover is devastating since most offspring coming from HEG-bearing females, including HEG-bearing heterozygotes, show loss-of-function phenotypes in somatic tissues. Thus, when the *Dfd*-HEG came from a female in the cross *Dfd*-HEG/+ × Df/TM3,Sb, the number of progeny carrying what should be a wild-type balancer chromosome (clv/TM3,*Sb* and HEG/TM3,*Sb*) was only 8% of that obtained when the sexes were switched. For the *yg*-HEG, only 17% of flies that inherited the HEG from the mother were fertile, again showing the negative effect on fitness due to maternal carryover. *Nanos* plays roles in somatic tissues in the nervous system of *Drosophila* (52). However, it is unlikely

that somatic expression in the zygote contributes to the *yellow-g* sterile phenotype since no fitness cost was observed in HEG/TM3,*Sb* progeny when the HEG came from a father. In addition, while *nanos* is expressed in the nervous system, there are no reports of its expression in somatic follicle cells. The high maternal carryover fitness costs were reflected in the results of our drive experiments, in which both HEGs decreased in frequency or were lost entirely despite significant levels of homing. To mitigate the problem of maternal carryover, different promoters are needed. Expression of the HEG could be limited to the male germline since, based on our observations and those of others (20, 36), carryover of Cas9 from sperm into the zygote is minimal. However, modeling suggests that for a female fertility or viability locus, the genetic load introduced into the population as a result of homing in males only is lower than with homing in the female or in both sexes (3). Thus, conditions that allow for adult germline-specific Cas9 activity in females or both sexes are likely to be more useful for suppression HEGs. Recent evidence in support of this hypothesis comes from the observation that homing rates are increased and carryover-mediated fitness costs are reduced when Cas9 expression and/or activity is more tightly limited through the use of different germline promoters in the mosquito (22). Degrons could also be incorporated into the Cas9 protein to give it a shorter half-life or modulate the phase of the cell cycle in which it is expressed. However, Cas9 levels must not be reduced to such an extent that cleavage is no longer efficient. Cas9 inhibitor proteins could potentially also be used to temporally limit Cas9 activity, if they were expressed under the control of a late oogenesis-specific promoter, so as to inhibit Cas9 activity during early embryogenesis, after cleavage and homing had already occurred during oogenesis (53, 54).

To summarize, our results show that while gRNA multiplexing can prevent the appearance of alleles resistant to homing, a number of other issues remain to be addressed in order for HEG-based population suppression to become a reality. These include which germline promoters and other strategies to use to limit carryover and, at the same time, promote homing rate; how to increase the frequency of homing rate-dependent repair versus NHEJ; how to guarantee equal loading of members of a multiplex set of gRNAs; and how to locate target sites and gRNAs so as to maximize the probability of multiple cleavage events while also maximizing the probability of homing rate that results in complete homing events. In approaching these questions, particularly in the context of suppression HEGs, loci whose activity is required in adults for female fertility, such as *yellow-g*, are particularly useful. First, females lacking *yellow-g* are viable but sterile, with an easily scored collapsed egg phenotype. Second, because chromosomal deficiencies that lack *yellow-g* are available in *Drosophila*, females can be isolated that only carry sequences from the cleaved/homed allele, making it straightforward to characterize sequence changes at the locus. Third, it is conserved in other Diptera, including *Drosophila* pests such as *D. suzukii* and disease vectors such as mosquitoes. In addition, earlier work in the mosquito showed that HEG-based targeting of *yellow-g* leads to female sterility, although also with resistance allele formation (16). Thus, understanding how to effectively home at high frequency into *yellow-g* in *Drosophila* may provide insights that can be translated to other species.

Methods

Target Gene Selection and gRNA Design. We chose two target genes for the suppression HEG drives, the HOX gene *Dfd* (31) and *yellow-g* (30). The mosquito homolog of *yellow-g* has been tested for homing and suppression drive using a single gRNA previously (16). Disruption of *Dfd* results in embryonic lethality, while disruption of *yellow-g* results in defects in egg shell formation, resulting in female sterility. Potential gRNAs were ranked by on-target activity (41) in the Benchling software suite (https://benchling.com) and selected to be spaced out over target gene exons in conserved regions (Fig. 1*A*).

Cloning of HEG Constructs. A list of all primers used in this study can be found in Dataset S3. The starting construct pCFD3-dU6:3gRNA was a gift from Simon Bullock, Division of Cell Biology, MRC Laboratory of Molecular Biology, Cambridge, United Kingdom (plasmid 49410; Addgene) (33). The BbsI cut sites to insert a gRNA into the scaffold were replaced with BsmBI sites by digesting with BbsI and ligating annealed oligos as described on flyCRISPR (flycrispr.molbio.wisc.edu) [forward (FWD) oligo: GTCGGGAGACGGACGTCTCT, reverse (REV) oligo: AAACAGAGACGTCCGTCTCC)] The *vermilion* marker gene was replaced by digesting with HindIII and ligating in a *white* marker gene.

We started by replacing the single gRNA scaffold with an optimized one published previously (34), in which the T base at position 4 was mutated to a G and the duplex was extended by 5 bp. The plasmid described above was digested with XbaI and BglII. Two fragments were amplified with the same plasmid as the template using primers 1-opti-g-FWD1 + 2-opti-g-REV1 and 3-opti-g-FWD2 + 4-opti-g-REV2. The nucleotide changes mentioned above were introduced in the primer overhangs of 2-opti-g-REV1 and 3-opti-g-FWD2. The construct was assembled in a two-fragment Gibson assembly (55) to yield plasmid pCFD3-dU6:3gRNA-BsmBI-white-optimized-sg-scaffold.

Next, we added a second single gRNA promoter similar to pCFD4-U6:1_U6:3 tandem gRNAs (33). The pCFD3-dU6:3gRNA-BsmBI-white-optimized-sg-scaffold was digested with XbaI. The U6:1 promoter was amplified from genomic DNA extracted from a *w-* stock using the Qiagen DNeasy Blood & Tissue Kit with primers 5-U6:1-FWD1 + 6-U6:1-REV1. The single gRNA scaffold was amplified from the uncut plasmid using primers 7-U6:1-FWD2 + 8-U6:1-REV2. The final plasmid was assembled in a two-fragment Gibson assembly yielding pU6:3-U6:1-tandem.

pU6:3-U6:1-tandem was used to subclone all gRNAs into the single-guide scaffold using the cloning strategy described by Port et al. (33). The pU6:3-U6:1-tandem was digested with BsmBI, and the two gRNAs were inserted in a one-fragment Gibson reaction with the gRNA sequences encoded in the primer overhangs.

For each of the HEG constructs, two of these guide cassettes were assembled for a total of four guides per construct: p-yelG-left with primers 9-yelG-guide1-left FWD + 10-yelG-guide2-left REV, p-yelG-right with primers 11-yelG-guide4-right FWD + 12-yelG-guide3-right REV, p-dfd-left with primers 13-dfd-guide-left-fwd + 14-dfd-guide-left-rev, and p-dfd-right with primers 15-dfd-guide-right-fwd + 16-dfd-guide-right-rev.

The construct for the fly stock targeting the 5srRNA was generated by inserting a gRNA in the original pCFD3-dU6:3gRNA vector (gRNA target sequence: gagaaccgatgtattcagcg).

HEG Construct. The HEG constructs utilized the *nanos* promoter and 3′ UTR to drive expression of Cas9. The pnos-Cas9-nos was a gift from Simon Bullock (plasmid 62208; Addgene) (33). The miniwhite marker in pnos-Cas9-nos was excised by digesting with ApaI and replaced with a *3xP3-td-tomato* marker cassette (56, 57). The nos-Cas9-nos cassette was then flanked by gypsy insulators: first, by cutting with NheI and inserting a gypsy insulator on the left and, second, by cutting with SacII and inserting the insulator on the right. Finally, the whole insulated nos-Cas9-nos cassette was excised with SapI and NsiI and inverted in a one-fragment Gibson reaction. The inversion was carried out to minimize the possibility that transcriptional readthrough from the *3xP3* promoter, present in the same orientation as Cas9 in the original construct, would lead to somatic expression of Cas9. The resulting plasmid was digested with SnaBI, and the left homology arm, along with the left U6-tandem cassette, was inserted in a two-fragment Gibson reaction. Following that, it was digested with ApaI, and the right homology arm and the right U6-tandem cassette were inserted as above (a detailed cloning scheme and primers used are shown *SI Appendix*, Fig. S1).

HEG Constructs Without Cas9. To generate HEG fly strains that could be easily modified, we created versions of our HEG constructs in which the Cas9 cassette was replaced by an attP landing site. Flies carrying these constructs were also used as control strains in the drive experiments.

The two HEG constructs, *Dfd*-HEG and *yg*-HEG, were digested with HindIII to remove nos-Cas9-nos and the *3xP3-td-tomato* marker. The new construct was assembled in a three-fragment Gibson reaction in which *td-tomato-sv40* expression was driven by the flightin promoter and an attP landing site was included, resulting in the inactive HEG constructs inactive-*Dfd*-HEG and inactive-*yg*-HEG.

Fly Germline Transformation. The two HEG constructs, *Dfd*-HEG and *yg*-HEG, as well as the two inactive HEG constructs, split-*Dfd*-HEG and split-*yg*-HEG, were injected into *w-* flies along with pnos-cas9-nos (33) as an additional source of Cas9. All injections were carried out by Rainbow Transgenic Flies.

The injected generation 0 flies were outcrossed to *w-* flies, and the resulting progeny were screened for the fluorescent *3xP3-td-tomato* eye marker. Due to the high frequency of progeny lethality or female sterility when the HEG was passed through females, transformed flies were kept as heterozygous stocks by outcrossing *td-tomato*–positive male flies to *w-* flies each generation.

Fly Crosses. In the crosses described below, we used the progeny resulting from a cross between heterozygous HEG-bearing males and *w-* virgins. Progeny of HEG-bearing females could not effectively be used because most died (*Dfd*) or gave rise to sterile adult females (*yellow-g*). To analyze homing and cleavage rates of the generated HEG strains, we set up the following crosses with Df lines of the targeted gene. Twenty-five heterozygous HEG-bearing females were crossed individually to Df/Balancer males, and vice versa. The *Dfd*-HEG strain was crossed to Df(3R)*Dfd*13/TM3,*Sb* (catalog no. 1980; Bloomington Drosophila Stock Center) (31). The chromosomal breakpoints of the Df are ∼460 kb upstream and 100 kb downstream of the *Dfd* locus:

$$\begin{array}{lcl} ♀\,Dfd\text{-HEG}/+ & \times & ♂\,\mathrm{Df(3R)Dfd13}/\mathrm{TM3},Sb \\ ♂\,Dfd\text{-HEG}/+ & \times & ♀\,\mathrm{Df(3R)Dfd13}/\mathrm{TM3},Sb \end{array}$$

The resulting progeny of the single fly crosses were scored for possible genotypes: *Dfd*-HEG/TM3,*Sb*; +(or clv)/TM3,*Sb*; +(or clv)/Df(3R)*Dfd*13; and Df(3R)*Dfd*13/*Dfd*-HEG.

The *yg*-HEG strain was crossed to Df(3L)BSC384/TM3,*Sb* (catalog no. 24408; Bloomington Drosophila Stock Center) (58). The chromosomal break-points of the Df are ∼50 kb upstream and 10 kb downstream of the *yellow-g* locus:

$$\begin{array}{lcl} ♀\,yg\text{-HEG}/+ & \times & ♂\,\mathrm{Df(3L)BSC384}/\mathrm{TM3},Sb \\ ♂\,yg\text{-HEG}/+ & \times & ♀\,\mathrm{Df(3L)BSC384}/\mathrm{TM3},Sb \end{array}$$

The resulting progeny were scored for possible genotypes: *yg*-HEG/Df(3L)BSC384; *yg*-HEG/TM3,*Sb*; +(or clv)/Df(3L)BSC384; and +(or clv)/TM3,*Sb*. The flies were also sexed in this case because only females elicit the *yellow-g* phenotype. All these females were outcrossed to *w-* males subsequently to check for sterility.

Females coming from a HEG-bearing father with the genotypes +(or clv)/TM3,*Sb* and yg-HEG/TM3,*Sb* were also outcrossed to check for paternal carryover of the HEG components. Only five vials were scored, showing no obvious effects on fertility.

Homing Rate and Cleavage Rate Calculations. For both drives, the homing rate can be estimated by looking at the proportion of HEG/TM3,*Sb* individuals out of all TM3,*Sb* individuals, since no TM3,*Sb* individuals should (in the absence of carryover) have died as a result of HEG activity. As such, the average or expected number of individuals per genotype can be estimated as half of the sum of the number of HEG/TM3,*Sb* and +/TM3,*Sb* individuals, a quantity we shall denote as *G*. If we assume an equal number of offspring for each genotype, we get our estimated homing rate by taking the HEG/TM3,*Sb* individuals, subtracting *G*, and then dividing the difference by *G*. The way to intuit this is as follows: HEG/TM3,*Sb* is a sum of the individuals that inherited the HEG through Mendelian transmission and those that received it from a homing event. Thus, by subtracting *G* and dividing the difference by *G*, we estimate the percentage of the HEG/TM3,*Sb* individuals that received the HEG from a homing event rather than through normal inheritance (the homing rate equation is shown below).

Similarly, since the only way to obtain +/Df-viable individuals (or fertile females in the case of the yG-HEG) is if cleavage does not occur, we can use the frequency of these events to estimate the cleavage rate for both drives by subtracting +/Df from *G* and dividing the difference by *G* (the cleavage rate equation is shown below).

Note that for the yG-HEG, the +/TM3,*Sb* pool represents both +/TM3,*Sb* and clv/TM3,*Sb* since we cannot distinguish between the two phenotypically.

Wild-type alleles are denoted with a "+" symbol. To avoid confusion with the mathematical symbol for addition, we used "&" in the formulas below:

$$\begin{aligned} G &= ((\mathrm{HEG}/\mathrm{TM3},Sb)\&(+/\mathrm{TM3},Sb))/2 \\ \text{homing rate} &= (\mathrm{HEG}/\mathrm{TM3},Sb\text{-}G)/G \\ \text{cleavage rate} &= (G\text{-}(+/\mathrm{Df}))/G \end{aligned}$$

Sequence Analysis of Escapers. We sequenced the targeted region from progeny flies that survived (*Dfd*) or were fertile (*yellow-g*) and that carried the Df chromosome from parental crosses that were HEG/+ × Df/TM3,*Sb*. These were identified as being Df-bearing by virtue of the fact that they lacked the balancer chromosome TM3,*Sb*. Genomic DNA was extracted from single flies using the Qiagen BloodNEasy Kit. The genomic regions containing gRNA target sites were amplified using LongAmp PCR Master Mix (New England Biolabs). Because of the long intron, the *Dfd* genomic region was amplified in two fragments. Fragment 1, containing gRNA1 and gRNA2, was amplified with primer 47-dfd-left-fwd and 48-dfd-left-rev. Fragment 2, containing gRNA3 and gRNA4, was amplified with primer 49-dfd-right-fwd and 50-dfd-right-rev. The same primers were used for sequencing. The *yellow-g* genomic region was amplified with primers 51-yG-fwd and 52-yG-rev. These primers were also used for sequencing, along with an additional primer 53-yG-mid.

Sequencing of Cleavage Events. We selected 10 sterile female progeny of the genotype clv/Df from crosses of female HEG/+ and male HEG/+ flies to individuals that were Df/TM3,*Sb*. Genomic DNA from single flies was extracted using the Qiagen BloodNEasy Kit. The targeted genomic region was amplified with LongAmp PCR Master Mix using primers located outside of the homology arms used for the homing construct (54-yG-outside FWD + 55-yG-outside REV). The resulting amplicons were purified on a gel and sequenced with primers 56-yG-seq1, 57-yG-seq2, 58-yG-seq3, and 59-yG-seq4.

Drive Experiments. Drive experiments were set up in fly food bottles for the two active homing stocks, *Dfd*-HEG and *yg*-HEG, as well as for the two inactive versions, inactive-*Dfd*-HEG and inactive-*yg*-HEG, as the controls. To seed the bottles, we crossed 30 female *w-* virgins to 30 heterozygous males of the corresponding drive or control stocks. After allowing the flies to mate for 2 d, they were transferred to a fresh fly food bottle together with 30 *w-* females that had been similarly mated with *w-* males. All drive experiments were set up in four biological replicates.

The flies were allowed to lay eggs for 2 d in the fresh bottle before being removed. After the next generation of adult flies hatched and mated in this bottle, all flies were dumped on a CO_2 pad and scored for the marker of the homing construct under a fluorescent microscope. To ensure that the bottles did not become overcrowded, we set an artificial population size limit of 200 flies. These were transferred to a new bottle as the next generation, and the cycle was repeated for 12 generations.

Modeling. We use a deterministic, discrete-generation, population frequency framework to model the spread of each HEG through a population, assuming random mating. This model is a variant of one used previously (59). It consists of a series of difference equations used to calculate the expected frequency of each genotype based on the frequencies of all genotypes from the previous generation, augmented by fitness effects, cleavage and homing events, and maternal carryover, and it is finally adjusted by a normalization factor. These equations, while straightforward, are rather lengthy due to the number of possible crosses, so we have provided them directly within our model in a supplementary MATLAB (MathWorks) file (Dataset S4).

Containment of HEG Flies. *Drosophila* research involving transgenesis, including transgenesis experiments designed to bring about gene drive, is governed by NIH guidelines detailed at the NIH website (https://osp.od.nih.gov/wp-content/uploads/NIH_Guidelines.html) in the document entitled "NIH guidelines for research involving recombinant or synthetic nucleic acid molecules (NIH guidelines) April 2016." As excerpted from the above document, Section III-D-4-a, this involves the following: "Recombinant or synthetic nucleic acid molecules, or DNA or RNA molecules derived therefrom, from any source except for greater than two-thirds of eukaryotic viral genome may be transferred to any non-human vertebrate or any invertebrate organism and propagated under conditions of physical containment comparable to BL1 or BL1-N and appropriate to the organism under study (see Section V-B, *Footnotes and References of Sections I-IV*)."

At the California Institute of Technology (Caltech), the Institutional Biosafety Committee (IBC) oversees all research utilizing recombinant and synthetic nucleic acid molecules, as well as genetically modified organisms. The IBC reviews and approves recombinant and synthetic nucleic acid molecule research for compliance with NIH guidelines, Caltech policies, and best laboratory practices. The Caltech IBC has reviewed this work and determined that the use of several physical containment measures provides an appropriate biocontainment strategy. The physical containment utilized in this study included the following: All HEG-bearing flies were kept in nested containers and kept in a dedicated room behind three doors. The outer door was placed on a restricted access key. Doors were additionally fitted with nets, and fly traps were placed in each room compartment. One investigator (G.O.) performed all experiments and fly handling. All flies were autoclaved before disposal.

Figures. Figs. 2 and 4 were plotted in R using the "ggplot2" and "RColorBrewer" packages (60–62). All figures were assembled and labeled in Inkscape.

ACKNOWLEDGMENTS. We thank Marlene Biller for technical assistance and three anonymous reviewers who helped improve and clarify this manuscript. This work was supported by Grant 18-CCB5400-05 from the California Cherry Board (to B.A.H. and G.O.) and by Research Fellowship 428/1-1 from the German Research Foundation (to G.O.). T.I. was supported by NIH Training Grant 5T32GM007616-39.

1. Burt A (2003) Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc Biol Sci* 270:921–928.
2. Stoddard BL (2005) Homing endonuclease structure and function. *Q Rev Biophys* 38: 49–95.
3. Deredec A, Burt A, Godfray HCJ (2008) The population genetics of using homing endonuclease genes in vector and pest management. *Genetics* 179:2013–2026.
4. Eckhoff PA, Wenger EA, Godfray HCJ, Burt A (2017) Impact of mosquito gene drive on malaria elimination in a computational model with explicit spatial and temporal dynamics. *Proc Natl Acad Sci USA* 114:E255–E264.
5. Deredec A, Godfray HCJ, Burt A (2011) Requirements for effective malaria control with homing endonuclease genes. *Proc Natl Acad Sci USA* 108:E874–E880.
6. Preston CR, Flores CC, Engels WR (2006) Differential usage of alternative pathways of double-strand break repair in Drosophila. *Genetics* 172:1055–1068.
7. Windbichler N, et al. (2011) A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature* 473:212–215.
8. Alphey N, Bonsall MB (2014) Interplay of population genetics and dynamics in the genetic control of mosquitoes. *J R Soc Interface* 11:20131071.
9. Hammond AM, et al. (2017) The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLoS Genet* 13: e1007039.
10. Marshall JM, Buchman A, Sánchez C HM, Akbari OS (2017) Overcoming evolved resistance to population-suppressing homing-based gene drives. *Sci Rep* 7:3776.
11. Unckless RL, Clark AG, Messer PW (2017) Evolution of resistance against CRISPR/Cas9 gene drive. *Genetics* 205:827–841.
12. Windbichler N, et al. (2007) Homing endonuclease mediated gene targeting in Anopheles gambiae cells and embryos. *Nucleic Acids Res* 35:5922–5933.
13. Simoni A, et al. (2014) Development of synthetic selfish elements based on modular nucleases in Drosophila melanogaster. *Nucleic Acids Res* 42:7461–7472.
14. Gantz VM, Bier E (2015) Genome editing. The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations. *Science* 348:442–444.
15. Gantz VM, et al. (2015) Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proc Natl Acad Sci USA* 112:E6736–E6743.
16. Hammond A, et al. (2016) A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nat Biotechnol* 34: 78–83.
17. Champer J, et al. (2017) Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLoS Genet* 13:e1006796.
18. Chan Y-S, Naujoks DA, Huen DS, Russell S (2011) Insect population control by homing endonuclease-based gene drive: An evaluation in Drosophila melanogaster. *Genetics* 188:33–44.
19. Chan Y-S, Huen DS, Glauert R, Whiteway E, Russell S (2013) Optimising homing endonuclease gene drive performance in a semi-refractory species: The Drosophila melanogaster experience. *PLoS One* 8:e54130.
20. Champer J, et al. (2018) Reducing resistance allele formation in CRISPR gene drive. *Proc Natl Acad Sci USA* 115:5522–5527.
21. KaramiNejadRanjbar M, et al. (2018) Consequences of resistance evolution in a Cas9-based sex conversion-suppression gene drive for insect pest management. *Proc Natl Acad Sci USA* 115:6189–6194.
22. Hammond AM, et al. (2018) Improved CRISPR-based suppression gene drives mitigate resistance and impose a large reproductive load on laboratory-contained mosquito populations. bioRxiv:10.1101/360339.
23. Jinek M, et al. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–821.
24. Cong L, et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339:819–823.
25. Mali P, et al. (2013) RNA-guided human genome engineering via Cas9. *Science* 339: 823–826.
26. Esvelt KM, Smidler AL, Catteruccia F, Church GM (2014) Concerning RNA-guided gene drives for the alteration of wild populations. *eLife* 3:03401.
27. Noble C, Olejarz J, Esvelt KM, Church GM, Nowak MA (2017) Evolutionary dynamics of CRISPR gene drives. *Sci Adv* 3:e1601964.
28. Dickson LB, et al. (2017) Exon-enriched libraries reveal large genic differences between *Aedes aegypti* from Senegal, West Africa, and populations outside Africa. *G3 (Bethesda)* 7:571–582.
29. Anopheles gambiae 1000 Genomes Consortium, et al. (2017) Genetic diversity of the African malaria vector Anopheles gambiae. *Nature* 552:96–100.
30. Claycomb JM, Benasutti M, Bosco G, Fenger DD, Orr-Weaver TL (2004) Gene amplification as a developmental strategy: Isolation of two developmental amplicons in Drosophila. *Dev Cell* 6:145–155.
31. Hazelrigg T, Kaufman TC (1983) Revertants of dominant mutations associated with the Antennapedia gene complex of DROSOPHILA MELANOGASTER: Cytology and genetics. *Genetics* 105:581–600.
32. Van Doren M, Williamson AL, Lehmann R (1998) Regulation of zygotic gene expression in Drosophila primordial germ cells. *Curr Biol* 8:243–246.
33. Port F, Chen H-M, Lee T, Bullock SL (2014) Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. *Proc Natl Acad Sci USA* 111: E2967–E2976.
34. Dang Y, et al. (2015) Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biol* 16:280.
35. Lin C-C, Potter CJ (2016) Non-Mendelian dominant maternal effects caused by CRISPR/Cas9 transgenic components in Drosophila melanogaster. *G3 (Bethesda)* 6:3685–3691.
36. Galizi R, et al. (2016) A CRISPR-Cas9 sex-ratio distortion system for genetic control. *Sci Rep* 6:31139.
37. DiCarlo JE, Chavez A, Dietz SL, Esvelt KM, Church GM (2015) Safeguarding CRISPR-Cas9 gene drives in yeast. *Nat Biotechnol* 33:1250–1255.
38. Champer J, Buchman A, Akbari OS (2016) Cheating evolution: Engineering gene drives to manipulate the fate of wild populations. *Nat Rev Genet* 17:146–159.
39. Xiang G, Zhang X, An C, Cheng C, Wang H (2017) Temperature effect on CRISPR-Cas9 mediated genome editing. *J Genet Genomics* 44:199–205.
40. Barrangou R, Marraffini LA (2014) CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Mol Cell* 54:234–244.
41. Doench JG, et al. (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* 34:184–191.
42. Farasat I, Salis HM (2016) A biophysical model of CRISPR/Cas9 activity for rational design of genome editing and gene regulation. *PLoS Comput Biol* 12:e1004724.
43. Pelletier J, Sonenberg N (1988) Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* 334:320–325.
44. Jang SK, et al. (1988) A segment of the 5′ nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J Virol* 62:2636–2643.
45. Ryan MD, King AM, Thomas GP (1991) Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *J Gen Virol* 72:2727–2732.
46. Ryan MD, Drew J (1994) Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *EMBO J* 13:928–933.
47. Xie K, Minkenberg B, Yang Y (2015) Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc Natl Acad Sci USA* 112:3570–3575.
48. Port F, Bullock SL (2016) Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs. *Nat Methods* 13:852–854.
49. Chu VT, et al. (2015) Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol* 33:543–548.
50. Lin S, Staahl BT, Alla RK, Doudna JA (2014) Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* 3:e04766.
51. Zielke N, et al. (2014) Fly-FUCCI: A versatile tool for studying cell proliferation in complex tissues. *Cell Rep* 7:588–598.
52. Bhogal B, Plaza-Jennings A, Gavis ER (2016) Nanos-mediated repression of hid protects larval sensory neurons after a global switch in sensitivity to apoptotic signals. *Development* 143:2147–2159.
53. Rauch BJ, et al. (2017) Inhibition of CRISPR-Cas9 with bacteriophage proteins. *Cell* 168:150–158.e10.
54. Harrington LB, et al. (2017) A broad-spectrum inhibitor of CRISPR-Cas9. *Cell* 170: 1224–1233.e15.
55. Gibson DG, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6:343–345.
56. Horn C, Jaunich B, Wimmer EA (2000) Highly sensitive, fluorescent transformation marker for Drosophila transgenesis. *Dev Genes Evol* 210:623–629.
57. Shaner NC, et al. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol* 22: 1567–1572.
58. Cook RK, et al. (2012) The generation of chromosomal deletions to provide extensive coverage and subdivision of the Drosophila melanogaster genome. *Genome Biol* 13:R21.
59. Marshall JM, Hay BA (2014) Medusa: A novel gene drive system for confined suppression of insect populations. *PLoS One* 9:e102694.
60. R Core Team (2013) R: A Language and Environment for Statistical Computing. R version 3.4.3. Available at www.R-project.org/. Accessed September 6, 2018.
61. Wickham H (2009) ggplot2: Elegant Graphics for Data Analysis. Version ggplot2 2.2.1. Available at https://ggplot2.tidyverse.org/. Accessed September 6, 2018.
62. Neuwirth E (2014) RColorBrewer: ColorBrewer Palettes. Version RColorBrewer 1.1.2. Available at https://cran.r-project.org/web/packages/RColorBrewer/index.html. Accessed September 6, 2018.