Cleave and Rescue, a novel selfish genetic element and general strategy for gene drive

Georg Oberhofer¹,², Tobin Ivy¹,², and Bruce A. Hay¹,²

¹Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125

Edited by James J. Bull, The University of Texas at Austin, Austin, TX, and approved January 7, 2019 (received for review October 2, 2018)

There is great interest in being able to spread beneficial traits throughout wild populations in ways that are self-sustaining. Here, we describe a chromosomal selfish genetic element, CleaveR (Cleave and Rescue (ClvR)), able to achieve this goal. ClvR comprises two linked chromosomal components. One, germline-expressed Cas9 and guide RNAs (gRNAs)—the Cleave—cleaves and thereby disrupts endogenous copies of a gene whose product is essential. The other, a recoded version of the essential gene resistant to cleavage and gene conversion with cleaved copies—the Rescue—provides essential gene function. ClvR enhances its transmission, and that of linked genes, by creating conditions in which progeny lacking ClvR die because they have no functional copies of the essential gene. In contrast, those who inherit ClvR survive, resulting in an increase in ClvR frequency. ClvR is predicted to spread to fixation under diverse conditions. To test these predictions, we generated a ClvR element in Drosophila melanogaster. ClvRtko is located on chromosome 3 and uses Cas9 and four gRNAs to disrupt melanogaster technical knockout (tko), an X-linked essential gene. Rescue activity is provided by tko from Drosophila virilis. ClvRtko results in germline and maternal carryover-dependent inactivation of melanogaster tko (>99% per generation); lethality caused by this loss is rescued by the virilis transgene; ClvRtko activities are robust to genetic diversity in strains from five continents; and uncleavable but functional melanogaster tko alleles were not observed. Finally, ClvRtko spreads to transgene fixation. The simplicity of ClvR suggests it may be useful for altering populations in diverse species.

Significance

There is great interest in spreading beneficial traits throughout wild populations in self-sustaining ways. Here, we describe a synthetic selfish genetic element, CleaveR (Cleave and Rescue (ClvR)), that is simple to build and can spread a linked gene to high frequency in populations. ClvR is composed of two components. The first, germline-expressed Cas9 and guide RNAs (gRNAs), cleave and disrupt versions of an essential gene located elsewhere in the genome. The second, a version of the essential gene resistant to cleavage, provides essential gene function. ClvR spreads by creating conditions in which progeny lacking ClvR die because they have no functional copies of the essential gene. In contrast, those who inherit ClvR survive, resulting in an increase in ClvR frequency.

Author contributions: G.O., T.I., and B.A.H. conceptualized the study; G.O., T.I., and B.A.H. provided the methodology; G.O. and B.A.H. investigated the study; T.I. provided mathematical modeling; G.O. and B.A.H. wrote the manuscript; G.O. and B.A.H. acquired funding; and G.O., T.I., and B.A.H. wrote the paper.

Conflict of interest statement: The authors have filed patent applications on ClvR and related technologies (US Application No. 15/970,728).

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹G.O. and T.I. contributed equally to this work.

¹To 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.1816928116/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1816928116

PNAS Latest Articles | 1 of 10
early embryo development (the consequences of toxin expression). The antidote is a transgene that results in early embryonic expression of a recoded version of this same gene that is resistant to miRNA silencing (the antidote), thereby providing essential gene function in a just-in-time fashion (6, 38, 39). Finally, prokaryotes also contain a number of tightly linked toxin-antidote clusters (including but not limited to type II restriction enzymes and their cognate methyltransferases). While many of these play important roles in cell physiology and defense, there are also multiple lines of evidence showing that some of them act in a selfish manner to increase their representation within populations through postsegregational killing of those that fail to inherit them (and thus the antidote) at cell division as a result of inefficient partitioning, or when in competition with other similar units (plasmids of the same incompatibility group/replicon) that lack them (40, 41). Based on these behaviors bacterial TA systems are sometimes known as addiction modules: the components they encode are fundamentally nonessential (as with the eukaryotic TA systems described above), but once they are acquired, they cannot easily be lost without causing death of the host cell.

The components of naturally occurring TA systems could in principle be adapted to bring about gene drive in other species of interest. While the locus that contains Tribolium Medea has been sequenced, the molecular nature of TA components that account for its behavior remains unknown. Toxins and antidotes associated with C. elegans maternal- and paternal-effect selfish genetic elements are known (35, 34), as are those associated with gamete/spore killing in yeast (36, 37). However, in these latter cases, it is unclear whether the mechanisms of action and any associated gene regulation required for selfish behavior can be transferred across species. Implementation of synthetic Medea was successful in Drosophila, but this relied on detailed knowledge of the molecular genetics that underlie maternal and early zygotic control of embryogenesis. Efforts to translate Medea to species other than the closely related Drosophila suzukii (39) have not yet succeeded. Toxins and antidotes from prokaryotes are well understood at a mechanistic level, and are often likely to be active in eukaryotic systems since many of them target highly conserved processes, such as translation, or promote the degradation of RNA or DNA (40–42). However, the use of these or other gain-of-function toxins and antidotes requires careful titration of the place and time they are transcribed and translated. Achieving such control, as with synthetic Medea elements, is likely to require the expedient development of some specific tools and reagents, including knowledge of details of development, promoters, and regulators of translation and degradation during key stages of developmental events (including maternal–zygotic transition. In sum, while existing TA systems are attractive to consider as a starting point for development of new gene drive systems—since they bring about drive in nature—the available tools do not yet provide a straightforward and general approach to building TA-based chromosomal gene drive methods in diverse species.

Here, we report the creation of a TA-based chromosomally located selfish genetic element whose components are simple and interchangeable, and likely to be generally available across species. Our starting point is the fact that site-specific alteration of DNA in the germline, mediated by Cas9 and gRNAs or other site-specific nucleases, followed by error-prone repair or creation of larger deletions, can be used most simply to disrupt the function of a gene, in our case an essential gene. Site-specific base editing enzymes (43) can be employed toward a similar end. Here, we focus on site-specific nucleases. Novel versions of essential genes that share limited or no nucleotide sequence similarity with the endogenous version, and are thus unrecognizable, can rescue the viability and fertility of individuals that otherwise carry only loss-of-function (LOF) versions of the essential gene (44–46). Recombination and gene conversion can occur between a cleaved locus and an uncleaved counterpart located elsewhere in the genome to which it has sequence similarity (47), and this could lead to the creation of functional, cleavage-resistant alleles at the endogenous essential gene locus. Reducing or eliminating sequence similarity between the cleaved version of the essential gene and an unrecognizable rescuing version can prevent such events (48). Finally, in the case of diploids, for many essential genes (haplosufficient recessive lethal or sterile), heterozygotes for a LOF allele are, at least to a first approximation, fit (49–51).

Under the above conditions, a cassette that includes germline-expressed Cas9 and gRNAs, designed to cleave in trans and thereby disrupt any endogenous wild-type copies of an essential gene, and a recoded version of the essential gene resistant to cleavage and recombination or gene conversion with cleaved versions of the wild-type allele, and therefore able to rescue those who carry it in cis, behaves as a selfish genetic element, which we refer to as CleaveR [Cleave and Rescue (ClvR)] (Fig. 1A). The toxin, Cas9 and gRNAs, works in trans by creating a permanent, potentially lethal change to the host genome wherever the targeted locus is located. However, this lethality only manifests itself in those who fail to inherit ClvR and its cis-acting antidote, the Rescue transgene. In contrast, those who inherit ClvR and the Rescue transgene contained within it survive, resulting in an increase in the frequency of individuals with ClvR-bearing chromosomes compared with those carrying non–ClvR-bearing counterparts. (Fig. 1, and other examples in SI Appendix, Figs. S1 and S2). This represents a form of postsegregational killing and leads cells, organisms, and populations to become dependent on (addicted to) the ClvR-encoded Rescue transgene (the antidote) for their survival. An analogy can be drawn with one strategy used to force the maintenance of a costly, nonessential plasmid in the absence of antibiotic selection. This involves locating an unconditionally essential gene (normally chromosomal) on the plasmid in cells that otherwise lack a functional copy of the essential gene (52). A ClvR element simply has the added feature that it provides the mechanism by which the endogenous version of the essential gene is inactivated in addition to the mechanism promoting survival in its absence. In Results and Discussion, we consider the specific case of ClvR behavior in a diploid animal, Drosophila melanogaster, as a model for other species such as mosquitoes, for which there has long been interest in the idea of altering wild populations so that they are unable to transmit diseases such as dengue, yellow fever, chikungunya, or malaria.

Results and Discussion

ClvR and the locus it targets for inactivation can be located on the same chromosome or on different chromosomes. The specific relationship is not important for gene drive since cleavage occurs in trans, wherever the target gene is located, while rescue only occurs in cis, in those who inherit ClvR. ClvR behavior is illustrated in Fig. 1B–D for the case in which ClvR is located on an autosomal and the haplosufficient essential gene targeted for cleavage is located on the X chromosome (see below and Figs. 2–5 for related experiments). Cleavage by Cas9 followed by inaccurate repair creates LOF alleles of the essential gene in the adult female germline (Fig. 1B). Diploid germ cells survive because they carry a copy of ClvR, which includes the recoded Rescue. In animals, haploid gametes lacking ClvR and a functional copy of the essential gene (e.g., some female gametes in Fig. 1C and D) will generally survive and be functional because essential gene products utilized during the haploid stage are expressed during the diploid stage and shared between the products of meiosis (53–55). However, in other organisms in which extensive transcription occurs during the haploid stage (e.g., plants and fungi), gametes lacking ClvR will be lost if transcription of the targeted essential gene is required during the haploid phase for gamete survival or function (SI Appendix, Fig. S1).

Here, we focus on animals. When a heterozygous female mates with a wild-type male, female progeny survive because they inherit a wild-type copy of the essential gene from their father. Some males who inherit the X-linked LOF allele from their mother also survive because they inherit an autosomal copy of ClvR, while others die because they inherit the X-linked LOF allele and the wild-type non–ClvR-bearing autosomal homolog (Fig. 1C). If there is maternal
carrii the essential gene inherited from the father can be converted to LOF allele. Arrows indicate conversion from wild-type to LOF allele. Large red Xs indicate offspring that die because they lack any source of essential gene function. The color of the centromere (large circle) indicates whether the chromosome is X or autosomal.

**Population Genetic Behavior of ClvR.** The behavior of such a ClvR element, located on an autosomal target, is illustrated in Fig. 2 for a conservative germline cleavage rate of 90% (actual rates, >99%; Fig. 3) and various release percentages and fitness costs, without (Fig. 2A) and with (Fig. 2B) maternal carryover-dependent cleavage (actual rates, >99%; Fig. 3). ClvR is predicted to behave as a low-threshold gene drive mechanism (no deterministic threshold for an element with no fitness cost), spreading to transgene-bearing genotype fixation for a wide range of release percentages and fitness costs. However, in contrast to a HEG, which can spread quickly from low frequency (56), spread of ClvR is very frequency dependent: slow when introduced at low frequency, and fast when introduced at high frequency (Fig. 2A and B). Maternal carryover-dependent cleavage is not essential for ClvR-dependent drive (Fig. 2A) but can speed the process and allow the drive element to tolerate larger fitness costs (Fig. 2B).

Finally, the behavior of many genes is described as haploinsufficient, this designation often reflects the results of characterization under controlled laboratory conditions. Characterization of the same heterozygotes under other environmentally relevant conditions may uncover varying levels of haploinsufficiency (cf. ref. 57). Given that wild populations carrying gene drive elements will experience a variety of biotic and abiotic environmental conditions, it is important to understand how haploinsufficiency would affect ClvR-dependent drive. To explore this, we examined the behavior of a ClvR located on one autosome, targeting an unlinked locus on a different autosome, with a single functional version of the target gene resulting in some level of haploinsufficiency (Fig. 2C). We modeled a two locus autosomal scenario rather than that of an autosomal ClvR targeting the X since most essential genes are on autosomes, and to be able to capture the effects of haploinsufficiency in both sexes. Interestingly, ClvR is predicted to bring about population alteration under a wide variety of conditions if the essential gene targeted is haploinsufficient (Fig. 2C), or even haplolethal (Fig. 2D).

**Synthesis of ClvRtko in Drosophila melanogaster.** To create ClvR in Drosophila melanogaster, we first generated a construct carrying a recoded version of D. melanogaster’s X-linked tko locus, which encodes the conserved, essential, and haploinsufficient mitochondrial ribosomal protein rps12 (58). To minimize homology of the rescue transgene with D. melanogaster tko, and thereby limit opportunities for recombination or gene conversion between the two (47, 48), we utilized the tko locus from a distantly related species, Drosophila virilis. We also introduced six additional silent coding sequence mutations to further reduce homology with the D. melanogaster gene (SI Appendix, Fig. S3). The tko rescue construct (tkoA) includes a dominant td-tomato marker, and an attP recombination site. It was introduced into the D. melanogaster genome on the third chromosome, at 68E, using Cas9 mediated HR, generating tkoA flies (SI Appendix, Fig. S4A).

---

**Fig. 1.** Basic structure of a ClvR element, and its behavior in a diploid, with and without maternal carryover. (A) Components of a ClvR element. (B) Behavior of ClvR on the right arm of an autosome and an essential gene located on X chromosome. The long thick horizontal black bar represents a chromosome with ClvR on the right arm of an autosome (see experiments below for an experimental implementation), while the shorter horizontal black bar represents an X chromosome carrying an essential gene. The identity of genes, alleles, and protein and RNA products are indicated. Arrows are drawn from a wild-type allele of the essential gene to the cleaved product resulting from Cas9 activity. (C) Results of a cross between a heterozygous ClvR-bearing female and a wild-type male, in the absence of maternal carryover of Cas9/gRNA complexes. Arrows indicate conversion from wild-type to LOF allele. (D) Same cross as in C, but with maternal carryover of Cas9/gRNAs sufficient to convert wild-type alleles of the essential gene inherited from the father into LOF alleles. The dashed boxes highlight the paternal X chromosome before and after cleavage and creation of a LOF allele. Arrows indicate conversion from wild-type to LOF allele. Large red Xs indicate offspring that die because they lack any source of essential gene function. The color of the centromere (large circle) indicates whether the chromosome was inherited from a female (red) or male (blue) parent. The Y chromosome is shown as a short horizontal black bar with a angled segment, and a blue centromere.
Population genetic behavior of ClvR when targeting a haplosufficient (A and B) or haploinsufficient (C and D) essential gene. (A and B) A discrete generation, deterministic population frequency model of ClvR spread in which cleavage occurs in the male and female germline; ClvR located on an autosome and the essential gene is located on the X (see data in Figs. 3 and 5) through a single panmictic population, for varying initial release percentages and fitness costs, without (A) or with (B) maternal carryover-dependent cleavage. The heatmap indicates the number of generations required for the ClvR-bearing genotype to approach fixation (i.e., >99% of the total population). (C) Heatmap showing the number of generations required for the ClvR-bearing genotype to reach fixation (<99% ClvR-bearing) for different initial release percentages and haploinsufficient fitness costs (100% = haplolethal), for an autosomal version of ClvR targeting a second unlinked autosomal locus, with maternal carryover. (D) Individuals traced showing the fate of a ClvR from (C) targeting a haplolethal gene, for different release percentages. The horizontal line represents an approximation of the unstable equilibrium frequency (≈31.5%); genotype frequencies do not change significantly over 20 generations). Genotype frequencies greater than equilibrium, 36%, 41%, and 46%; those below, 26%, 21%, and 16%. Note that the term “Release %” for all heatmaps refers to the percentage of homozygous transgenic males compared with wild-type males and females after a release has occurred (e.g., a 40% release means that 40% of the population is of the adult progeny carried ClvR). We then carried out a panmictic population, for varying initial release percentages and its self is assumed to have no fitness cost. Such costs would further increase the minimum release percentages required for drive to occur, as in A and B.

second step, transgenes expressing Cas9 and four gRNAs designed to recognize and cleave DNA within the D. melanogaster tko coding region, but not that of D. viridis tko, were integrated into the attP site in tkoA rescue construct-bearing flies (SI Appendix, Figs. S4 B and C). The gRNAs were each expressed under the control of a U6 polymerase III promoter (59). Cas9 was expressed under the control of nanos regulatory sequences, which drive expression in the male and female germline (60). Nano-driven Cas9 also results in extensive maternal, but not paternal, carryover of active Cas9/gRNA complexes into the zygote (29, 61). The final construct is designated ChR\textsuperscript{tko} (Fig. 3A), and flies that carry it as ChR\textsuperscript{tko} flies.

**Genetic Behavior of ClvR**. Matings between males that carry a LOF mutation for the X-linked eye pigmentation gene white (w\textsuperscript{1118}), and that are heterozygous for ChR\textsuperscript{tko} on the third chromosome (w\textsuperscript{1118}, ChR\textsuperscript{tko}/+), where + indicates a third chromosome that does not carry ChR\textsuperscript{tko}, and homozygous w\textsuperscript{1118}, ++ females resulted in high levels of progeny viability to adulthood (95.2 ± 2.0%), similar to those for the w\textsuperscript{1118} strain used for transformation (95.9 ± 2.0%). In addition, ~50% (50.1 ± 3.0%) of the adult progeny carried ChR\textsuperscript{tko}, as expected for Mendelian segregation and high ChR\textsuperscript{tko} heterozygote fitness. Matings among homozygous ChR\textsuperscript{tko} flies also resulted in high levels of viability to adulthood (95.1 ± 1.7%), indicating that the presence of ChR\textsuperscript{tko} components (in the likely absence of functional D. melanogaster tko; see below) does not result in obvious fitness costs. In contrast, when heterozygous w\textsuperscript{1118}, ChR\textsuperscript{tko}/+ females were mated with homozygous w\textsuperscript{1118}, ++ males, 53.6 ± 1.3% of progeny did not reach adulthood, and all surviving adults carried ChR\textsuperscript{tko}. On the basis of these results, we infer that the presence of ChR\textsuperscript{tko} in mothers results in a very high frequency (>99%) of mutational inactivation of the D. melanogaster tko locus in the adult female germline and in the zygote through maternal carryover-dependent cleavage of the paternal allele. In consequence, those who fail to inherit ChR\textsuperscript{tko} die, while those who inherit a single copy of ChR\textsuperscript{tko} thrive (SI Appendix, Table S1 A and B).

To obtain estimates of the rate of female adult germline- and maternal carryover-dependent cleavage and subsequent D. melanogaster tko inactivation, we repeated the cross between ChR\textsuperscript{tko}/+ females and wild-type males with larger numbers of individuals (see also SI Appendix, Table S5, for additional experiments of this type with genetically diverse strains). All but one of 3,736 progeny that survived to adulthood (cleavage rate of >99.9%) carried ChR\textsuperscript{tko} (Fig. 3B and SI Appendix, Table S2). To estimate male germline cleavage rates, we carried out a cross between ChR\textsuperscript{tko}/+ males and females that carried a lethal tko LOF allele [tko\textsuperscript{3}, a frameshift mutation at amino acid 108 that introduces a premature stop codon (62)] in trans to the balancer chromosome FM7 (the balancer prevents meiotic recombination between X chromosomes), which is wild type for tko and carries a dominant mutation in the Bar gene, B\textsuperscript{1}. Female progeny that inherit the
maternal tko<sup>3</sup> allele (identified by their failure to carry the dominant B<sup>1</sup> marker), and that lack ClvR<sup>tko</sup> (and therefore lack the td-tomato and GFP markers), should die if D. melanogaster tko was inactivated in the parental male germline and survive if it was not. Eight females carrying the tko<sup>3</sup> allele and lacking ClvR<sup>tko</sup> were recovered compared with 907 that carried tko<sup>3</sup> and ClvR<sup>tko</sup>, for a minimum male germline cleavage rate of >99.9% (Fig. 3C and SI Appendix, Fig. S5 and Table S3). ClvR<sup>tko</sup>-dependent rescue of the tko<sup>3</sup> mutant phenotype is indicated by the large numbers of tko<sup>3</sup>/Y; ClvR<sup>tko</sup>/+ progeny (880), compared with none for tko<sup>3</sup>/Y; ClvR<sup>tko</sup>/+ (Fig. 3C).

**X Chromosomes In Which a tko LOF Allele Was Not Created Following Exposure to ClvR<sup>tko</sup> Remain Sensitive to Cleavage by ClvR<sup>tko</sup>.** We sequenced the D. melanogaster tko locus from each of the nine X chromosomes above, in which a tko LOF allele was not created (escapers) following exposure to maternal or paternal ClvR<sup>tko</sup>. In the single escaper coming from a ClvR<sup>tko</sup>/+ mother, all four gRNA target sites were unaltered. For seven escapers coming from the ClvR<sup>tko</sup>/+ father, there was a common 3 bp in-frame deletion within the gRNA1 target site, and the remaining three target sites were unaltered. For escaper M3, a mixed sequencing signal, which may be indicative of nuclear mosaicism, was obtained. When each of the above escaper chromosomes was isolated in a male and the male crossed to ClvR<sup>tko</sup>/+ females, all surviving progeny inherited the ClvR<sup>tko</sup> td-tomato and GFP markers, showing that the D. melanogaster tko locus remained sensitive to cleavage (SI Appendix, Fig. S5 and Table S4).

**ClvR<sup>tko</sup> Functions in Diverse Genetic Backgrounds.** To alter wild populations, a gene drive mechanism must be able to function in diverse genetic backgrounds. To begin to explore this topic with ClvR, we crossed ClvR<sup>tko</sup>/+ females to males from Global Diversity Lines (GDL) isolated from five different continents (63), and used in previous work investigating Cas9 function in the context of engineered HEGs (27). After each generation, we scored the frequency of ClvR<sup>tko</sup> flies, collected 30 virgins, and backcrossed them again to males from each of the GDL lines. Results are summarized in SI Appendix, Table S5. All offspring were ClvR<sup>tko</sup>-bearing for each of six generations (7,882 progeny scored). While these results do not preclude the existence of unlinked genetic variants and/or gRNA target polymorphisms in wild populations that would result in decreased rates of cleavage and LOF mutation creation, they show that the system is not specific to a common laboratory strain [SI Appendix, Table S6, shows all gRNA target site polymorphisms in strains from the 1000 fly genomes project (64)].

**Molecular Nature of Mutations Created in D. melanogaster tko Created Following Exposure to ClvR<sup>tko</sup>.** To analyze the mutations in D. melanogaster tko created by ClvR<sup>tko</sup> we selected 2 ClvR<sup>tko</sup>-bearing male progeny from each of nine individual single crosses (18 total flies) between heterozygous ClvR<sup>tko</sup> females and w<sup>1118</sup> males.
LOF alleles can be created via cleavage followed by NHEJ, or via locus spanning the gRNA-binding sites are D. virilis. The gRNA1 target site contained indels of varying size in all 18 individuals. The gRNA2 target site contained a likely preexisting polymorphism in four individuals (also observed in roughly half of the 1000 Fly Genome Project strains (64)), and a 2 bp deletion in 3. The gRNA3 target site was unaltered in all individuals, and the gRNA4 target contained indels in nine individuals. Somewhat surprisingly, larger deletions between target sites were not observed. This raises the possibility, suggested by others (65), that close juxtaposition of multiple target sites—in our case, four target sites within a 250-bp region of the tko ORF—limits Cas9’s ability to simultaneously interact with and/or cleave multiple nearby target sites as a consequence of Cas9-dependent DNA supercoiling.

One implication of such a model is that mutations should accumulate at additional target sites over time, as the target sites first cleaved by Cas9 are rendered nonfunctional for further Cas9 binding due to mutation within the gRNA target site. To explore this possibility, and the general question of whether all gRNA target sites can be cleaved, we sequenced the melanogaster tko locus from a homozygous $C{hr}^{Rko}$ stock that had been inbred for three generations (SI Appendix, Fig. S6 C and D, and Table S7B). Among the 12 analyzed males, all 12 had mutations at the gRNA1 target site. The gRNA2 target site was mutated in five, unaltered in one individual, and carried the suspected common polymorphism in the remaining six. The gRNA3 target site was mutated in 1 fly, and the gRNA4 target site was mutated in all 12 flies. Thus, cleavage events accumulate over time, and all sites can be cleaved, although cleavage efficiencies differ (from 100% for gRNA1 in generation 1 to 8% for gRNA3 after three generations).

The mutations we observe presumably arise initially from error-prone repair by nonhomologous end-joining (NHEJ) or microhomology-mediated end-joining pathways (Fig. 4). However, we note that ChR elements may also utilize HR and homing to create new LOF alleles when the ChR-bearing individuals introduced into the wild population carry (as the above results indicate they will) uncleavable LOF indels in the targeted essential gene. For example, if ChR-bearing individuals carrying LOF indels in the essential gene mate with wild-type, ChR-bearing progeny will be heterozygous for chromosomes that carry the LOF indels and the wild-type version of the essential gene. In the germline of these individuals, the LOF indel-bearing chromosome (which is uncleavable) can serve as a template for HR-mediated repair of cleaved wild-type alleles, converting them to the LOF sequence (Fig. 4). Such behavior in cleavage heterozygotes was recently described in yeast (66). Further implications of homing-dependent alteration of the essential gene locus are discussed below.

**$C{hr}^{Rko}$ Spreads to Genotype Fixation in D. melanogaster.** Our combined results show that $C{hr}^{Rko}$ results in a very high frequency of germine and maternal carrierover-dependent mutational inactivation of the D. melanogaster tko locus (>99% per generation); the lethality caused by this loss can be efficiently rescued using the D. virilis transgene; the high frequency of $C{hr}^{Rko}$-dependent mutational inactivation of D. melanogaster tko and rescue by D. virilis tko is robust to genetic diversity; and cleaved but functional D. melanogaster tko alleles resistant to further cleavage, which could limit drive, were not observed. These observations predict that $C{hr}^{Rko}$ will spread to genotype fixation. To test this prediction, we initiated two drive experiments. In one experiment, $w^{1118}$; $C{hr}^{Rko}$ heterozygous males were mated with $w^{1118}$; +/+ females, creating a progeny population used to seed the first generation in which $C{hr}^{Rko}$ was present in one-half of the individuals, at a total population allele frequency of 25%. In a second experiment, homozygous $w^{1118}$; $C{hr}^{Rko}$ males and $w^{1118}$; +/+ males were premated with equal numbers of $w^{1118}$, +/+ females, which were then combined and used to seed the first generation (25% $C{hr}$-bearing individuals), also resulting in an initial $C{hr}^{Rko}$ allele frequency of 25%. This level of introduction, although substantial, is not unreasonable as it is substantially lower than that used in earlier nontransgenic insect population suppression programs (67). As a control, we carried out similar drive experiments utilizing flies that carry the Rescue-only tko construct, tkoA, and that are wild type at the endogenous tko locus ($w^{1118}$, tkoA). tkoA carries the td-tomato marker and the Rescue transgene, but lacks gRNAs and Cas9, and is thus expected to show Mendelian transmission. $w^{1118}$; tkoA/+ males were mated with $w^{1118}$; +/+ females (also wild type for tko), creating a progeny population used to seed the first generation in which tkoA was present in one-half of the individuals, at a total population allele frequency of 25%. For the first drive experiment, five replicate population cages were followed for 18 generations (drive 1, Fig. 5A). For the second drive experiment, four replicate populations were followed for 16 generations (drive 2, Fig. 5B). For the control, four tkoA populations were followed for 10 generations. In both $C{hr}^{Rko}$ drive experiments $C{hr}^{Rko}$ spread to genotype fixation between six and nine generations for all replicates. In contrast, the control transgene, tkoA, remained near its introduction frequency in all populations. As expected based on modeling, wild-type (+) alleles at the third chromosome locus into which $C{hr}^{Rko}$ was inserted were still present in the five drive 1 populations (Fig. 5D and SI Appendix, Table S8), but since
wild-type alleles of *D. melanogaster* *tko* (SI Appendix, Fig. S6 and Table S7), these chromosomes are trapped in *ChR<sup>tko</sup>*/+ heterozygotes.

**Strategies for Maintaining ClvR Functionality over Time.** In any gene drive-based strategy for altering the makeup of a population, the cargo and drive mechanisms are subject to separation, mutational inactivation, and loss of efficacy. Resilience, an ability to respond to these forces in ways that maintain and/or restore the ability to alter populations over time, is essential. Mutation of cargo genes or loss of effectiveness as a result of evolution of the host, or of other species such as pathogens on which they are meant to act, requires that strategies be available for removing an old element from the population and replacing it with a new one. This can be achieved using an approach analogous to that proposed for synthetic *Medea* selfish genetic elements (6, 68), in which a second-generation *ClvR*, *ChR<sup>ns+1</sup>*, is located at the same site as the first-generation element, *ChR<sup>n</sup>* with *ChR<sup>ns+1</sup>* targeting essential gene<sup>ns+1</sup>, while also carrying the original rescuing copy of essential gene<sup>n</sup>. Because progeny carrying *ChR<sup>n</sup>* are sensitive to loss of essential gene<sup>ns+1</sup>, only those carrying *ChR<sup>ns+1</sup>* survive, regardless of their status with respect to *ChR<sup>n</sup>* (SI Appendix, Fig. S7). Opportunities for physical separation of Cargo from a functional Rescue can also be minimized, as with *Medea* (6), by interleaving Cargo and Rescue transgenes in various ways (SI Appendix, Figs. S8–S10).

Clavage is required for *ChR* selfish behavior, and can fail as a result of mutation within target sites or Cas9/gRNAs. Mutations within the target sites that create uncleavable, but functional alleles of the target locus (resistant alleles), can lead to loss of *ChR* from the population if its presence is associated with a fitness cost. Resistant alleles can arise from de novo mutations, from preexisting natural variation in the population, and as a result of error-prone NHEJ or microhomology-mediated end-joining pathways. Error-prone repair is likely to be the most important because the mutation rate per nucleotide/per generation is low, ~10<sup>−8</sup> to 10<sup>−9</sup> (69), and high-frequency preexisting mutations that produce target site resistance to cleavage can be avoided through sequencing of the target population. In contrast, NHEJ-mediated creation of resistance alleles following cleavage can occur frequently [>10<sup>−3</sup> per generation (27, 70)], although use of targets sites that cannot easily mutate to resistance and high fitness may be able to reduce this frequency dramatically (20). Modeling suggests that the probability of completely resistant alleles emerging with a multiplex of gRNAs is approximately equal to that of the probability of resistant alleles emerging at all gRNA target sites simultaneously, that is, p<sup>n</sup>, where p is the probability of a single site mutating to resistance and n is the number of gRNAs/target sites (71). Thus, even for a high rate of single target site mutation to resistance of 10<sup>−2</sup> to 10<sup>−5</sup>, resistant alleles at all target sites might be predicted to arise only infrequently (~10<sup>−6</sup> to 10<sup>−10</sup>) with a four-gRNA *ChR*. However, this calculation assumes no standing variation in the population at any of these sites, that all gRNAs work equally well, and that ectopic gene conversion between the Rescue transgene and the cleaved allele can be completely prevented by recoding.

The results reported herein, using laboratory and global diversity strains (0 resistant alleles out of more than 11,000 progeny scored; Fig. 3 and SI Appendix, Table S5), along with other recent work on HEGs (29, 61), provide experimental support for the idea that multiplexing of gRNAs can prevent the creation of cleavage-resistant, but functional alleles. Use of target sites that cannot easily mutate to a cleavage-resistant but high-fitness genotype have also been used toward a similar end (20). Targeting highly conserved housekeeping genes such as *tko* supports both strategies. Nonetheless, given that drive in very large populations has not yet been attempted, we briefly consider a “worst-case” scenario involving resistant alleles, *ChR*, and a panmictic population, to gain some feeling for the consequences of resistant alleles on *ChR* lifetime. We suppose that alleles that are completely resistant to four gRNAs, and with high fitness, arise at a high frequency of 10<sup>−5</sup> per generation, that the presence of *ChR* results in a significant fitness cost of 20% when homozygous (10% when heterozygous), and that *ChR* is introduced at a low (10%) or a high (50%) frequency. Under these conditions, *ChR*-bearing individuals constitute ≥99% of the population for 456 generations when introduced at a frequency of 10%, and 713 generations when introduced at a frequency of 50%. If homing of resistant alleles into cleaved wild-type alleles in heterozygotes carrying *ChR* is now included (Fig. 6A), *ChR* lifetime at high frequency (≥99% transgene-bearing) is modestly reduced to 409 generations for a 10% introduction frequency (Fig. 6B) and 707 generations for a 50% introduction frequency (Fig. 6C). The effect of homing is limited because it requires the presence

---

Fig. 5. *ClvR* spreads to genotype fixation in Drosophila. The frequency of *ClvR*-bearing individuals (*ClvR*<sup>+</sup> and *ClvR/ClvR*) is indicated on the y axis and the generation number on the x axis. Drive replicates in red; predicted drive behavior in dotted black lines. (A) Drive 1: *ClvR<sup>+/+</sup>* *XX 1<sup>st</sup>*/*<sup>118</sup> as generation 0. (B) Drive 2: *ClvR<sup>ns0</sup>/Clv<sup>ns0</sup> *XX 1<sup>st</sup>*/*<sup>118</sup> and *tko<sup>118</sup>* *XX 1<sup>st</sup>*/*<sup>118</sup> at a 1:1 ratio as generation 0. (C) Control drive: *tkoA<sup>+</sup>* *XX 1<sup>st</sup>*/*<sup>118</sup> as generation 0. For the control drive, we used flies carrying construct *tkoA* (Methods) that had only the rescue and the *td-tomato* marker, but no *Cas9* and gRNAs. (D) Allele frequency of *ClvR<sup>ns0</sup>* in drive 1. Replicates coming from drive 1 in red. Model (black) is the predicted allele frequency inferred from modeling of the drive using parameters estimated from the data in Fig. 3, and assuming no fitness cost to carrying *ClvR* (see SI Appendix, Table S8, for counts).
of a specific genotype, a ClvR-bearing mother carrying both a wild-type and a resistant allele at the essential gene locus (Fig. 6A), and the speed at which ClvR elements transform wild-type alleles into LOF alleles works to limit the frequency of such individuals.

With respect to mutational inactivation of Cas9 and gRNAs, ClvR-dependent drive of Cargo into a population is predicted to be remarkably insensitive to loss of these components when ClvR is introduced area-wide, even when inactive versions of Cas9/gRNAs are present at significant frequencies (5% of the ClvR-bearing individuals) in the initial population, and carriers of these mutant versions are more fit than those carrying intact Cas9 (Fig. 6D–F). Drive is robust because so long as active ClvR elements are present, the population is rapidly driven toward Rescue- and thus Cargo-bearing genotype fixation by the ongoing loss of endogenous wild-type copies of the essential gene.

Once all endogenous alleles of the essential gene are rendered nonfunctional, the population is locked into a Rescue—and thus Cargo-bearing—state regardless of whether Cas9 and gRNAs are still active. These points notwithstanding, we note that ClvR dynamics in the presence of resistant alleles at the target site or inactive Cas9 are likely to be more complicated in spatially structured populations that also include migration of wild types, a topic that remains to be explored. Strategies for further constraining the ability of Cas9/gRNAs to mutate to inactivity that involve forcing Cas9 and gRNAs to bring about transcription of the rescue as well as cleavage of the essential gene can also be envisioned (Fig. 6G). In one such strategy, a Cas9-VPR fusion protein is utilized. Cas9-VPR mediates cleavage at full length target sites. Cas9-VPR can also bind truncated gRNA target sites and drive transcription of a nearby gene, but is unable to cleave these sites (72). In this way, the same gRNAs and Cas9 are used

---

**Fig. 6.** The consequences of target site/Cas9/gRNA inactivation for the spread of cargo, and ways of selecting against inactivation. (A) Illustration of how repair using HR and the resistant allele as a template can result in an increase in frequency of a resistant allele. (B) Genotype frequencies of ClvR (red line) and a resistant allele (blue) for an element that is introduced at a 10% genotype frequency (release of homozygous ClvR males), and that carries a 20% fitness cost, with 100% homing. (C) Same as in B, but for a 50% introduction frequency. (D) Behavior of ClvR for an element that is introduced at a 10% genotype frequency (release of homozygous ClvR males), and that carries a 10% fitness cost. No inactive versions of ClvR (dead Cas9) are present (0% null). (E) Same as in D, but with versions of ClvR that lack active Cas9 introduced, so as to make up 5% of the initial ClvR-bearing population. The fitness cost of dead Cas9-bearing ClvR elements is assumed to be half that (5%) of the fully functional element. (F) Heatmap showing number of generations needed for Cargo to reach transgene-bearing genotype fixation (>99%) for different release percentages and fitness costs, in which versions of ClvR that lack active Cas9 are introduced so as to constitute 5% of the ClvR-bearing individuals, for each release percentage. Fifty percent of the fitness cost associated with ClvR is assumed to be due to Cas9 activity, with the rest being due to Cargo. Thus, the wild-type, non-ClvR chromosome always has the highest fitness. Compare with Fig. 2B. (G) A hypothetical circuit that selects against mutation of Cas9/gRNAs to inactivity in which Cas9 activity is made essential for Rescue function.
for cleavage of the endogenous version of the essential gene and transcription of the Rescue transgene.

**Conclusions.** Our findings demonstrate that the genetic composition of a population can be rapidly altered using the relatively simple toolkit of components that make up a ChR gene drive/selfish genetic element: a site-specific DNA-modifying enzyme such as Cas9 and the gRNAs that guide it to specific targets, sequences sufficient to drive gene expression in the germline (which need not be germline-specific), an essential gene to act as target, and a recoded version of the essential gene resistant to sequence modification and able to rescue the LOF condition. Highly conserved housekeeping genes such as tko that participate in universal cellular processes required for cell survival or maintenance of basic cellular functions are good candidates for use in implementation ChR in diverse species since they are essential in most if not all species (44–46). Importantly, modeling shows that drive and the alteration of populations to transgene-bearing genotype fixation can be achieved regardless of whether the essential gene being targeted is haplosufficient or haploinsufficient. This is likely to be important since haploinsufficiency may be more common than appreciated, and the fitness of individuals heterozygous for a LOF allele, under conditions present in the wild, is rarely known in advance. Finally, in the case where LOF alleles in the essential gene are created as a result of cleavage (as opposed to cleavage-independent base editing), ChR does not require utilization of a specific repair pathway.

An important feature of ChR is that the rate at which it spreads is frequency dependent (Fig. 2), very slow when introduced at low frequency, and fast when introduced at high frequency. In consequence, ChR is likely to be most useful when it can be introduced area-wide, rather than from a point source within a larger area of interest. More detailed modeling that takes into account features such as density dependence, migration, and spatial structure is required to fully understand ChR behavior. There are several other important unknowns. First, it is unclear what the costs and consequences are of long-term expression of DNA sequence-modifying enzymes such as Cas9, and if selection for alleles at other loci that result in decreased expression and/or activity may occur. A related unknown is the extent to which diversity in genome sequence in wild populations at the target site or elsewhere will thwart cleavage at the target locus. Our failure to identify cleavage-resistant, but functional tko alleles among >11,000 progeny from crosses of heterozygous ChR-bearing females to wild-type males from a laboratory strain and GDL strains from five continents are promising in this regard, but the level of diversity tested likely pales beside that present in wild populations of some species of interest (73, 74). The problem of sequence diversity is also faced by other drive mechanisms designed to alter populations, such as synthetic Medea (6), some versions of underdominance (15, 16), and HEG-based homing (17, 19), which rely on the recognition of specific nucleotide sequences for their mechanism of action. Only further work in genetically diverse populations of species of interest, in facsimiles of wild environments, will suffice to determine whether synthetic selfish genetic elements able to thrive in the wild can be created.

**Methods**

**Target Gene Selection and gRNA Design.** We selected the tko locus on the X chromosome as the target for the ClvR system. It encodes an essential mitochondrial ribosome protein and is recessive lethal and haploinsufficient (58). We used the benchling software suite to design gRNAs targeting the exonic regions of the gene at four positions, selected based on on-target activity ranking (75). An additional criteria was that the gRNAs have a mutated PAM in the rescue construct to avoid any potential off-target cleavage therein (see below).

**Cloning of ClvR Constructs and Fly Germline Transformation.** All plasmids used in this work were assembled with standard molecular cloning techniques and Gibson assembly (76). All restriction enzymes, enzymes for Gibson Assembly mastermix, and Q5 polymerase used in PCRs were from NEB; gel extraction kits and JM109 cells for cloning were from Zymo Research. The DNA extraction kit was from Qiagen (DNeasy). The gRNA cassette and Cas9 were based on pCDFD4-DUK-u6gRNA, which were a gift from Simon Bullock, Division of Cell Biology, MRC Laboratory of Molecular Biology, Cambridge, United Kingdom (77) (Addgene; #49410 and #62208) and modified as described previously (61). Construct A (SI Appendix, Fig. S4A) was inserted into the fly germline via Cas9-mediated HR. Construct B (SI Appendix, Fig. S4B) was integrated into an attP landing site within flies carrying construct A using the phiC31 site-specific integration system. Detailed procedures can be found in SI Appendix, Supplementary Methods. Construct sequence fasta files can be found in Dataset S1.

**Fly Crosses and Husbandry of ClvR**flies. Fly husbandry and crosses were performed under standard conditions at 26 °C. Rainbow Transgenics carried out all of the fly injections. Containment and handling procedures for ClvR flies were as described previously (61), with G.O. and B.A.H. performing all fly handling. Details are in SI Appendix, Supplementary Methods.

**Data Availability.** All data are available in the main text or SI Appendix. ClvR flies are available on request to labs that will meet or exceed containment guidelines outlined in ref. 61.

**ACKNOWLEDGMENTS.** We thank Marlene Biller and Alexander Sampson for technical assistance, and Jackson Campher and Andrew G. Clark for providing the GDL Drosophila RNA and stocks obtained from the Bloomington Drosophila Stock Center (NIH Grant P40OD018537) were used in this study. This work was carried out with support to B.A.H. from the US Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) Specialty Crop Initiative, under USDA NIFA Award 2012-51811-20086. G.O. was supported by a research fellowship from the Deutsche Forschungsgemeinschaft (OB428/1-1). T.I. was supported by NIH Training Grant ST32GM07616-39.


