EVOLUTIONARY GENETICS

Evolutionary dynamics of CRISPR gene drives

Charleston Noble,[1,2,3]* Jason Olejarz,[1]* Kevin M. Esvelt,[2†] George M. Church,[2,3] Martin A. Nowak[1,4,5†]

The alteration of wild populations has been discussed as a solution to a number of humanity's most pressing ecological and public health concerns. Enabled by the recent revolution in genome editing, clustered regularly interspaced short palindromic repeats (CRISPR) gene drives—selfish genetic elements that can spread through populations even if they confer no advantage to their host organism—are rapidly emerging as the most promising approach. However, before real-world applications are considered, it is imperative to develop a clear understanding of the outcomes of drive release in nature. Toward this aim, we mathematically study the evolutionary dynamics of CRISPR gene drives. We demonstrate that the emergence of drive-resistant alleles presents a major challenge to previously reported constructs, and we show that an alternative design that selects against resistant alleles could greatly improve evolutionary stability. We discuss all results in the context of CRISPR technology and provide insights that inform the engineering of practical gene drive systems.

INTRODUCTION

Gene drive systems are selfish genetic elements that bias their own inheritance and spread through populations in a super-Mendelian fashion (Fig. 1A). These elements have been discussed as a means of contributing to the eradication of insect-borne diseases, such as malaria, reversing herbicide and pesticide resistance in agriculture, and controlling destructive invasive species (*1*–*12*). Various examples of gene drive can be found in nature, including transposons (*13*), Medea elements (*14*, *15*), and segregation distorters (*16*–*19*), but for ecological engineering purposes, endonuclease gene drive systems received the most significant attention in the literature (*1*–*10*, *20*–*22*). In general, these elements function by converting drive heterozygotes into drive homozygotes through a two-step process: (i) the drive construct, encoding a sequence-specific endonuclease, induces a double-strand break (DSB) at its own position on a homologous chromosome, and (ii) subsequent DSB repair by homologous recombination (HR) copies the drive into the break site. Any sequence adjacent to the endonuclease will be copied as well; if a gene is present, we refer to it as "cargo," as it is "driven" by the endonuclease through the population.

Although originally proposed over a decade ago (*1*), the chief technical difficulty of this approach—inducing easily programmable cutting at arbitrary target sites—has only recently been overcome by the discovery and development of the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) genome editing system (*23*–*27*). Briefly, Cas9 is an endonuclease whose target site is prescribed by an independently expressed guide RNA (gRNA) via a 20-nucleotide protospacer sequence. Because virtually any position in a genome can be uniquely targeted by Cas9, so-called RNA-guided gene drive elements can be constructed by inserting a suitable sequence encoding both Cas9 and gRNA(s).

Recent studies demonstrated highly functional CRISPR gene drive elements in mosquitoes (*5*, *6*), yeast (*7*), and fruitflies (*8*). In each case, the basic construct consists of a copy of Cas9 with a single corresponding gRNA and cargo sequence (Fig. 1B). Despite drive inheritance of about 95%, on average, in the published studies (compared to 50% expected by Mendelian inheritance), the evolutionary stability of these constructs in large populations has been debated because of the potential emergence of drive resistance within a population (*1*, *2*, *21*). A resistant allele is anticipated to arise whenever the cell repairs the drive-induced DSB using nonhomologous end joining (NHEJ) instead of HR, a process that typically introduces a small insertion or deletion mutation at the target sequence. Because the reported constructs cut only at a single site, a substantial fraction of NHEJ events will create drive-resistant alleles that could prevent the construct from spreading to the entire population (Fig. 1B).

Drive resistance was first mathematically studied in the context of single-cutting homing endonuclease–based drive elements (*21*). There, it was concluded that drive is most effective when the fitness cost of the drive is low and the fitness cost of resistance is high (see section S1 for a description of that work). Unfortunately, in the drive constructs reported thus far, these two requirements are fundamentally at odds: the fitness cost of resistance arises from disruption of the target sequence, but the drive copies itself precisely by disrupting the target sequence.

Here, we study the evolutionary dynamics of an alternative drive architecture that decouples these effects by rescuing the function of the target gene, but only if the drive cassette is successfully copied. This design was first proposed conceptually by Esvelt *et al.* (*2*) but has not yet been modeled or constructed in the laboratory; hence, we refer to it here as the "proposed" construct. It involves targeting multiple sites within the 3′ end of a gene for cutting by the drive and including a completely genetically recoded (*28*–*30*) copy of this 3′ target sequence in the drive construct (Fig. 1C). The 3′ untranslated region of the gene is also replaced with an equivalent sequence to remove all homology between the cut sites and the drive components, which ensures that the drive cassette is copied as a single unit. If repair occurs by HR, then the target gene is restored to functionality as the drive is copied. However, if repair occurs by NHEJ, then the target gene is mutated, potentially resulting in a knockout and a corresponding loss of fitness. Using this design, drive resistance can be selected against by choosing an essential or even haploinsufficient gene as the drive target.

Because the success of this design is contingent on the ability to genetically recode the 3′ end of an essential gene without imposing a large fitness cost, we now briefly discuss the plausibility of this strategy. In a study of CRISPR-based gene drive in yeast, DiCarlo *et al.* (*7*)

[1]Program for Evolutionary Dynamics, Harvard University, Cambridge, MA 02138, USA. [2]Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, MA 02138, USA. [3]Department of Genetics, Harvard Medical School, Cambridge, MA 02138, USA. [4]Department of Mathematics, Harvard University, Cambridge, MA 02138, USA. [5]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: martin_nowak@harvard.edu (M.A.N.); esvelt@mit.edu (K.M.E.)

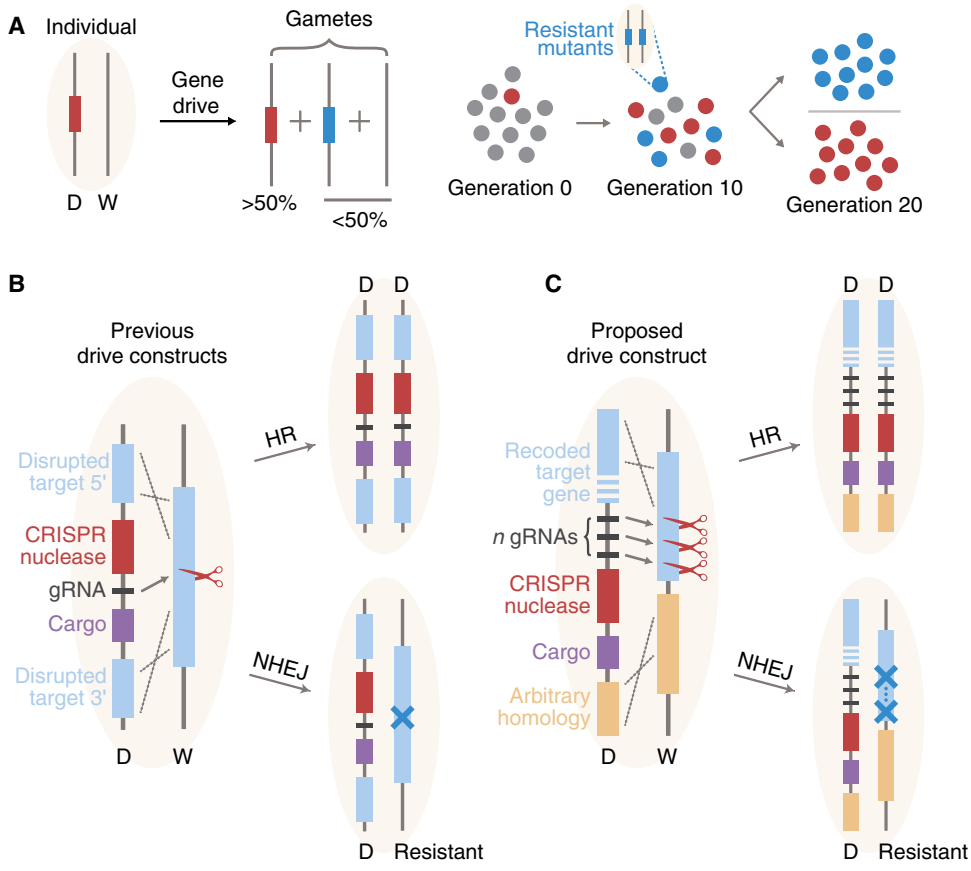

Fig. 1. CRISPR gene drive inheritance and spread in wild populations. (**A**) Inheritance and spread of a gene drive construct, D, in a population of individuals homozygous for the wild type, W. In the late germ line, the drive construct induces a DSB at its own position on the homologous chromosome, which is repaired either by HR, converting the individual to a DD homozygote, or by NHEJ, producing a small insertion/deletion/substitution mutation at the cut site, which results in a drive-resistant allele. There is also the possibility of no modification, in which case the W allele remains unchanged. This mechanism can lead to rapid spread of the gene drive in a population or spread of resistant alleles, depending on their relative fitness effects. (**B**) To achieve this mechanism, previously demonstrated drive constructs are inserted at some target sequence (blue) and carry a CRISPR nuclease (for example, Cas9) with a gRNA, as well as a "cargo gene," which can be chosen arbitrarily for the desired application. Disruption of the target sequence must be nearly neutral for the drive to spread. (**C**) The construct modeled here, which was proposed by Esvelt *et al.* (*2*), reconstitutes the target gene after cutting—so an essential gene can be chosen as the target to select against resistant alleles—and uses multiple (*n*) gRNAs.

showed that a drive construct targeting the essential *ABD1* gene and encoding a recoded copy of *ABD1* functioned with high efficiency without exhibiting "any obvious fitness defects as compared to wild-type strains." In the most comprehensive study of essential gene recoding to date, Ostrov *et al.* (*30*) showed that computationally minimizing disruption of existing RNA-binding motifs and secondary structures while preserving overall codon usage allowed the elimination of seven codons from 91% of essential genes in *Escherichia coli*, with an overall fitness cost of less than 10%. Moreover, many attempted recodings were costless on the first try without requiring optimization. Wang *et al.* (*31*) obtained similar results. Finally, work in *Drosophila* on underdominance-based drive systems (*11*, *32*) has shown that partial recoding of haploinsufficient genes in metazoans is possible, although in both studies this involved RNA interference.

In addition to 3′ target recoding, the construct uses multiple gRNAs. The use of multiple gRNAs offers two important benefits with respect to resistance: (i) all gRNA target sites must be mutated or lost before a single allele becomes drive-resistant, and (ii) if cutting occurs at two or more gRNA target sites simultaneously, then the intervening DNA sequence is lost, resulting in a large deletion and a knockout of the target gene. This is in contrast to single-cutting constructs, where a knockout can be avoided by an in-frame indel or substitution mutation.

RESULTS

To study this construct, we formulate a deterministic model (see Materials and Methods and sections S2 and S7) that considers the evolution of a large population of diploid organisms and focuses on a specific locus with $2n + 2$ alleles (Fig. 2A). First, there are the wild-type (W) allele and the gene drive allele with n gRNAs (D). There are then n distinct "cost-free" resistant alleles that are resistant to drive-induced cutting at 1, 2, …, n target sites but are otherwise identical to the wild type (denoted S_1, S_2, …, S_n). These could arise via, for example, mutations that block cutting by disrupting the gRNA target sequences but do not cause a shift in the reading frame. Finally, there are n distinct "costly" resistant alleles, which have fitness effects that are distinct from those of the wild type (denoted R_1, R_2, …, R_n). Only the alleles S_n and R_n are fully resistant to cutting by the drive. We also refer to the wild-type allele as S_0 for notational convenience. Last, we say that individuals having genotype AB, where A and B are any of the

alleles above, have fitness f_{AB} (alternatively, genotype AB is associated with a cost $1 - f_{AB}$) and produce gametes having haplotype C with probability $p_{AB,C}$. Note that these probabilities $p_{AB,C}$ abstract all individual-level drive dynamics and are agnostic to the mechanism that produces drive. We allow these parameters to be arbitrary for our analytical calculations and derive corresponding results that hold for any underlying drive mechanism—including both the previous drive constructs and the new ones considered here.

For numerical simulations, we further consider a mechanistic model that explicitly describes the mechanism of drive in individuals (Fig. 2B and section S7.3). We assume that, in the germ line of an individual that is heterozygous for a drive construct and a susceptible allele (DS_i, where $0 \le i < n$, or DR_i, where $1 \le i < n$), each susceptible target site undergoes cutting independently with probability q. If there is at least one cut, then HR occurs with probability P, whereas NHEJ occurs with probability $1 - P$. If HR occurs, then the cell is converted to a drive homozygote. However, if mutagenic NHEJ occurs, then there are a few possibilities, depending on the number of cuts.

If there is exactly one cut, then one gRNA target is lost on the susceptible allele. If the susceptible allele was initially functional (S_i), then with probability γ, it retains function and converts to S_{i+1}; otherwise, it loses function and converts to R_{i+1}. We assume that the parameter γ is the probability that the reading frame is unaffected, so $\gamma = 1/3$. If the susceptible allele is initially nonfunctional (R_i), then we assume that it cannot regain function, so it converts to R_{i+1}.

If there are two or more cuts, then all j susceptible gRNA targets between and including the outermost damaged targets in the locus are lost ($2 \le j \le n - i$). The resulting allele is certainly nonfunctional and thus converts to R_{i+j}. The probability distribution for the number of lost targets is described in section S7.3.2. It follows directly from our assumptions that cutting at each target site is independent and that sequential cutting and repair events do not occur.

Regarding initial conditions, our simulations and analytical invasion analysis assume that drive homozygotes (genotype DD) are released into a population consisting initially of fully susceptible wild-type homozygotes (genotype S_0S_0). However, depending on the sequence targeted by the drive, standing genetic variation in real populations could result in preexisting resistance at one or more gRNA targets. For example, in a genome-wide analysis of 192 inbred strains of *Drosophila melanogaster* derived from a single natural population, Mackay *et al.* (*33*) found the genome-wide averaged polymorphism value (*34*) to be $\pi = 0.0056$. If we assume that polymorphism at each base pair is independent, then the number of mismatches at a gRNA target sequence in a particular individual is binomial with 20 trials and success probability π. If each gRNA can tolerate, on average, one mismatch in its target, then single guide–resistant alleles should exist at frequencies roughly on the order of 10^{-3}. Further assuming that resistance at each gRNA is independent, two guide–resistant alleles should exist at frequencies roughly on the order of 10^{-5}, and so on. In this example and with these assumptions, using five gRNAs would reduce the frequency

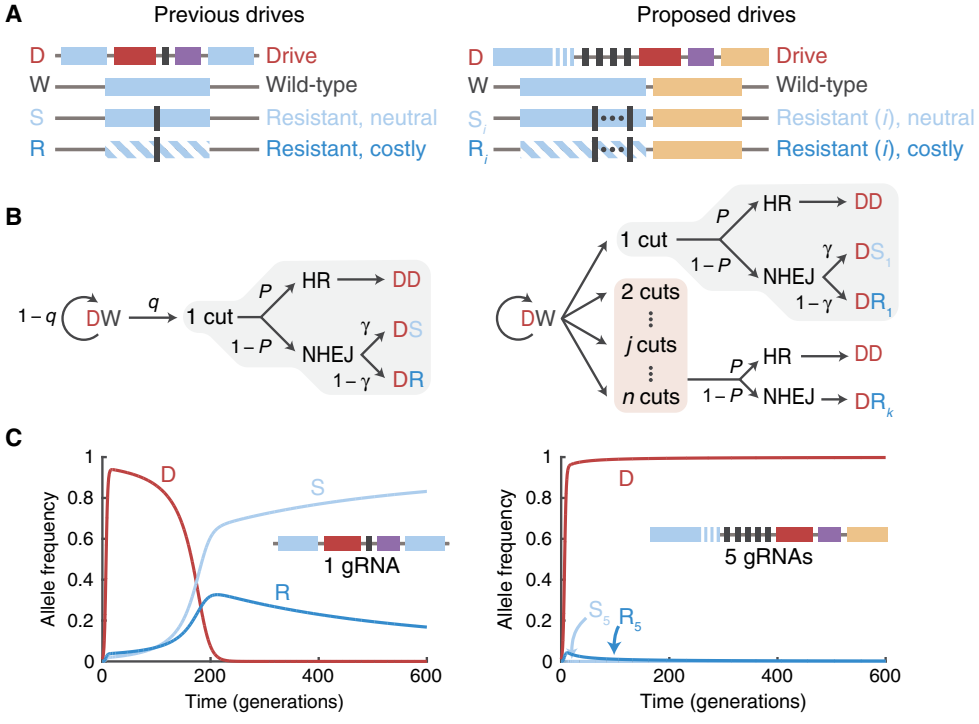

Fig. 2. Modeling framework and representative simulations. (**A**) We consider $2n + 2$ alleles, where n is the number of drive target sites (prescribed by CRISPR gRNAs): the drive construct (D), the wild type (W), n "neutral" resistant alleles (S_i), and n "costly" resistant alleles (R_i). Previous drives (left) used one target site, whereas our proposed drives use multiple target sites (right). (**B**) Conversion dynamics within DW germline cells during early gametogenesis. Cutting occurs at each susceptible target independently with probability q. Then, repair occurs by HR with probability P or by NHEJ with probability $1 - P$. In the case of a single cut (light gray), if there is NHEJ repair, then repair produces a functional target gene with probability γ or a nonfunctional target with probability $1 - \gamma$. Two or more cuts (light red) certainly produce nonfunctional targets after NHEJ repair. (**C**) Representative simulations using high cutting and HR probabilities ($q = P = 0.95$), for an initial drive release of 1% in a wild-type population, with $\gamma = 1/3$. Fitness parameters are (left) $f_{SS} = f_{SR} = 1$, $f_{SD} = 95\%$, $f_{RR} = 99\%$, $f_{DD} = f_{DR} = (99 \times 95\%) = 94.1\%$, where S refers to neutral alleles (either S or W), and (right) $f_{SS} = f_{SR} = 1$, $f_{SD} = f_{DD} = f_{DR} = 95\%$, $f_{RR} = 1\%$, where S and R refer to alleles W, S_1, …, S_5 and R_1, …, R_5, respectively. See section S7.3.2 for details regarding our assignments of the inheritance probabilities.

of preexisting fully resistant alleles to 10^{-12}. Of course, complications could arise, such as nonindependence of polymorphism within or between guides, so we anticipate this to serve as a low estimate of the frequency of preexisting resistance in a natural population. Therefore, before any application is considered, standing variation in the target population should be carefully measured, and the target gene as well as the number of guides should be adjusted accordingly.

Now, we address two fundamental questions: whether a CRISPR gene drive will invade a resident wild-type population and, if so, whether it will be evolutionarily stable (*35*). We begin with the former. We find that a CRISPR gene drive will invade a wild population if

$$2p_{\mathrm{WD,D}}f_{\mathrm{WD}} > f_{\mathrm{WW}} \quad (1)$$

A derivation of this result can be found in sections S3 and S7.1. For the drive to spread when initially rare, the advantage from inheritance biasing ($p_{\mathrm{WD,D}}$)—typically about 95% in published studies—must overcome the lower fitness of the drive/wild-type heterozygote (f_{WD}) compared with the wild type (f_{WW}). Note that this condition holds in the context of drive resistance, is agnostic to individual-level drive dynamics, and thus applies both to previous drive architectures and to our proposed architecture. Equation 1 explains the apparent success of CRISPR drive constructs reported in the literature (*5*–*8*), which easily invade wild-type laboratory populations, or would be predicted to do so after optimization of drive expression: Over short time scales, drive resistance is rare and thus does not affect the dynamics.

However, over longer time scales, NHEJ-mediated resistance will markedly affect the dynamics. We find that a resident drive population is stable against invasion by resistant alleles if and only if

$$\max_{\mathrm{A}\in\mathrm{S}\cup\mathrm{R}}(2p_{\mathrm{DA,A}}f_{\mathrm{DA}}) < f_{\mathrm{DD}} \quad (2)$$

Here, the maximization is over all nondrive alleles S_0, …, S_n and R_1, …, R_n. Intuitively, the drive is stable if and only if no other allele can invade, and each of these has an invasion condition identical in form to Eq. 1 (sections S4 and S7.2).

Disconcertingly, Eq. 2 suggests that drive constructs are necessarily unstable in sufficiently large populations. An individual who is heterozygous for the drive and the fully resistant cost-free allele S_n has probability $p_{\mathrm{DSn,Sn}} = {}^1/_2$ of producing an S_n gamete, and this individual has fitness equivalent to (or potentially greater than) the drive/wild-type heterozygote. Thus, if the drive construct has lower fitness than the wild type, and if the fully resistant cost-free allele has a nonzero rate of production in the population, then the latter will certainly invade a resident drive population. This is especially problematic for highly deleterious population suppression drives, as in the study by Hammond *et al.* (*6*), which have low fitness relative to the wild-type and less costly resistant alleles.

However, population alteration drives (sometimes referred to as replacement drives) might not require long-term persistence in a population to produce their desired effect. Some applications might still be successful as long as the drive construct attains and persists at a sufficiently high frequency in the population over some length of time.

To quantify the relative effectiveness of the previous and proposed drive architectures, we consider three quantities: (i) the maximum frequency achieved by a drive construct released in a wild population, (ii) the time required for a drive construct to attain 90% of its maximum frequency, and (iii) the frequency of the drive construct after 200 generations, roughly the longest relevant time scale for a typical application. We compute these quantities numerically for drives featuring cutting and HR probabilities, consistent with average drive inheritance rates observed in previous fruitfly (*8*) and mosquito (*5*, *6*) experiments ($q = P = 0.95$, modeling a reported drive inheritance rate of roughly 95% from DW individuals).

Our results suggest that, as anticipated from Eq. 1, both the previous and proposed drive constructs should spread similarly in the short term, immediately following release (Fig. 3, A, B, and D). However, over longer time scales, the two constructs undergo markedly different dynamics. The proposed drive constructs, released at an initial frequency of 1% in a wild population, using five gRNAs and targeting an essential gene, can attain >99% frequency in a population (Fig. 3, B and C) in 10 to 20 generations (Fig. 3, B and D) and remain above 99% for at least 200 generations (Fig. 3, B and E). Furthermore, this is seen over a large range of drive fitness costs, up to approximately 30% (Fig. 3, C to E). In contrast, the previously demonstrated constructs attain maximum frequencies between 90 and 95% over a narrower range of fitness values (Fig. 3, A and C) and demonstrate significantly reduced stability (Fig. 3E). In particular, previous constructs exceeding 8% fitness cost invariably fall below their initial release frequency in fewer than 200 generations.

DISCUSSION

In summary, we constructed and analyzed a mathematical model of CRISPR gene drive that includes multiplex cutting via multiple gRNAs and allows for multiple costly and cost-free resistant alleles. Our results suggest that previously demonstrated CRISPR gene drives constructed as proofs of principle should effectively invade wild populations—consistent with experimental observations—but could have limited utility due to their inherent instability, brought about by their production of resistant alleles and vulnerability to preexisting ones. We studied an alternative drive architecture, first proposed by Esvelt *et al.* (*2*), which contains (i) multiple CRISPR gRNAs that target the 3′ end of a gene and (ii) a recoded copy of the target gene that is functional but resistant to cutting. We discussed the plausibility of building such a construct in light of recent experimental reports, and we concluded that this architecture could substantially improve the stability of CRISPR gene drives by minimizing the effects of NHEJ-mediated resistance.

Another alternative strategy that we have not modeled here would involve multiple independent single-guide drive constructs targeting the same locus. This is conceptually symmetric to the strategy considered here: Rather than a single drive with multiple (n) gRNAs ("multiple guides"), one might consider multiple (n) drives with one gRNA each ("multiple drives"). In this strategy, each independent drive would behave similarly to the previously demonstrated constructs. The multiple-drive strategy would likely outperform the previous strategy, but we anticipate that it would not outperform the multiple-guide strategy. This is because, in the multiple-drive strategy, each gRNA target can independently undergo NHEJ-mediated mutation, providing stepping stones to fully resistant alleles. Furthermore, the multiple-drive strategy lacks the benefit of large NHEJ knockouts from multiple simultaneous cuts, which help combat cost-free resistance (Fig. 2B, red box), although it would be capable of editing regions unimportant to fitness. Also, regardless, each single-guide drive construct could itself be built in the way we have described here, by using multiple gRNAs.

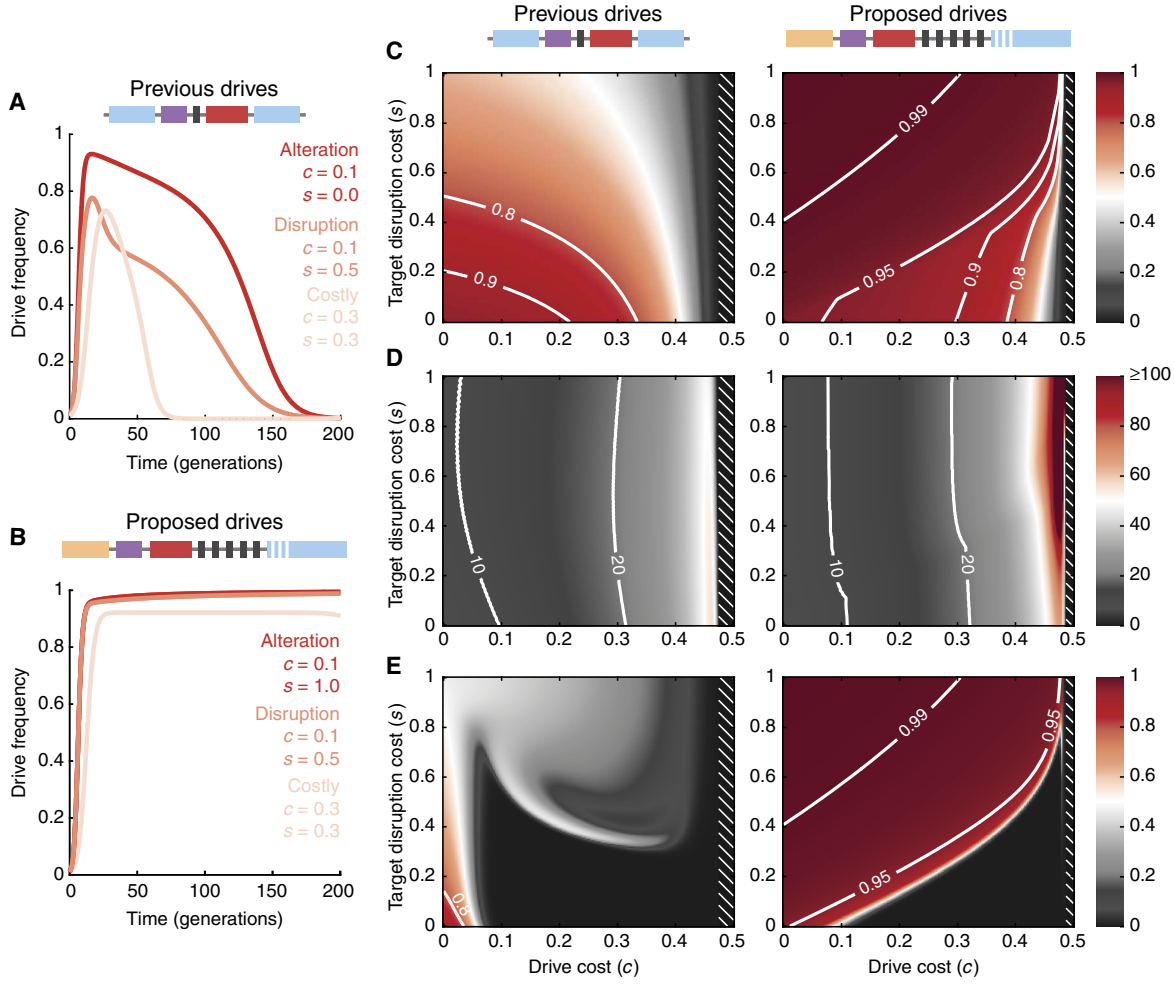

Fig. 3. Quantitative comparison of previously demonstrated and recently proposed drive constructs. (**A** and **B**) Drive frequency over time for three particular scenarios: a low-cost alteration drive carrying a cargo gene and targeting a neutral site (previous drives) or an essential gene (proposed drives) (red), a low-cost drive whose aim is to disrupt an important target gene (orange), and a high-cost drive (tan). (**C**) Maximum drive allele frequency (heat) observed in simulations across 200 generations, following an initial release of drive-homozygous organisms comprising 1% of the total population. In white hatched regions, Eq. 1 is not satisfied, so no invasion occurs. (**D**) Generations to 90% of the maximum frequency. (**E**) Frequency of the drive constructs after 200 generations, a measure of stability in the population. Parameters used are as follows: (throughout) $q = P = 0.95$, $\gamma = 1/3$; (previous drives) $n = 1$, $f_{SS} = f_{SR} = 1$, $f_{SD} = 1 - c$, $f_{DD} = f_{DR} = (1 - c)(1 - s)$, $f_{RR} = 1 - s$; (proposed drives) $n = 5$, $f_{SS} = f_{SR} = 1$, $f_{SD} = f_{DD} = f_{DR} = 1 - c$, $f_{RR} = 1 - s$, where S and R refer to any alleles S_0, …, S_n and R_1, …, R_n, respectively. Inheritance probabilities are assigned as illustrated in Fig. 2B and described in section S7.3.2.

An important caveat of our work is that we specifically studied resistance that is genetically encoded at the drive locus and is generated by the action of the drive. Many other mechanisms of resistance are certainly possible. For example, standing genetic variation and de novo mutation might be important considerations, particularly if the target locus is not highly conserved. However, in recent work (*36*), Unckless *et al.* showed that NHEJ-mediated resistance should be more impactful for realistic NHEJ rates (specifically, greater than the inverse of the population size). Aside from these mechanisms of within-locus resistance, resistance could also arise in trans, for example, as heightened ribonuclease activity or as the evolution of small RNAs that would lead to knockdowns via RNA interference. In addition, even beyond direct molecular effects, resistance could arise via higher-level effects, for example, as selection for inbreeding behavior in hermaphrodites in response to extremely costly population suppression drives, as recently studied by Bull (*37*). The large variety of potential resistance mechanisms underscores the need for further theoretical and experimental work on this topic.

Although our work has focused on how to maximize the invasibility and stability of gene drive systems, "global" CRISPR gene drives, such as those considered here, should only be actively developed for severe problems that (i) cause a great deal of suffering and (ii) have few other potentially viable solutions. Examples include malaria and schistosomiasis. Other applications—such as precision alterations to local populations—will require robust methods to ensure limited spatial and/or temporal spread. Toward this aim, there are several existing approaches, including nondrive strategies such as multi-locus assortment (*38*) and threshold-dependent drives (like toxin-based underdominance systems) (*11*, *14*). Moreover, we, among others, recently proposed an alternative theoretical approach termed "daisy drive" (*39*).

In conclusion, our results suggest three concrete design principles for future CRISPR gene drive systems. Constructs will minimize the impact of misrepair and thus maximize evolutionary stability if (i) multiple gRNAs with minimal off-target effects are used, (ii) disruption of the target locus is highly deleterious, and (iii) any cargo genes are as close to neutral as possible.

MATERIALS AND METHODS

Throughout this work, we study a genetics-based evolutionary dynamics model. We consider the evolution of diploid individuals, x_{IJ}, where I,J = W, D, R_1, R_2, …, R_n, S_1, S_2, …, S_n. Here, D corresponds to the drive with n gRNAs; R_1, R_2, …, R_n correspond to alleles that are resistant to cutting at 1, 2, …, n target sites, respectively, and S_1, …, S_n are resistant alleles with no fitness cost, and W corresponds to the wild type (which we also denote by S_0 for notational convenience). In the Supplementary Materials (section S2, extended to neutral resistance in section S7), we present a continuous-time model for the evolutionary dynamics of this population, as well as derivations for the invasion and stability conditions discussed above. Here, we briefly describe this model. First, it makes the following assumptions: (i) an infinitely large population; (ii) random mating; (iii) standard segregation of allele pairs at meiosis, unless an individual has genotype DA (where A is one of S_0, …, S_{n-1} or R_1, …, R_{n-1}), in which case gametes receive a D allele with probability $p_{\mathrm{DA,D}}$ or an A allele with probability $p_{\mathrm{DA,A}}$; and (iv) viability selection where each genotype IJ has fitness f_{IJ}.

Using these rules, we can formally express the rates at which each of the $2n + 2$ types of gametes is produced in terms of the frequencies of individuals in the population. We denote by $F_{\mathrm{D}}(t)$ the rate (at time t) at which drive gametes (D) are produced by individuals in the population. We denote by $F_{\mathrm{S}_i}(t)$ the rate (at time t) at which wild-type gametes ($i = 0$) or gametes with varying levels of cost-free resistance ($1 \le i \le n$) are produced by individuals in the population. Last, we denote by $F_{\mathrm{R}_i}(t)$ the rate (at time t) at which gametes with varying levels of costly resistance ($1 \le i \le n$) are produced by individuals in the population. We have

$$F_{\mathrm{D}}(t) = f_{\mathrm{DD}}x_{\mathrm{DD}}(t) + \sum_{k=1}^{n} p_{\mathrm{R}_k\mathrm{D,D}} f_{\mathrm{R}_k\mathrm{D}} x_{\mathrm{R}_k\mathrm{D}}(t) + \sum_{k=0}^{n} p_{\mathrm{S}_k\mathrm{D,D}} f_{\mathrm{S}_k\mathrm{D}} x_{\mathrm{S}_k\mathrm{D}}(t)$$

$$F_{\mathrm{S}_i}(t) = \sum_{k=0}^{n} \frac{1+\delta_{ki}}{2} f_{\mathrm{S}_k\mathrm{S}_i} x_{\mathrm{S}_k\mathrm{S}_i}(t) + \frac{1}{2}\sum_{k=1}^{n} f_{\mathrm{R}_k\mathrm{S}_i} x_{\mathrm{R}_k\mathrm{S}_i}(t) + \sum_{k=0}^{i} p_{\mathrm{S}_k\mathrm{D,S}_i} f_{\mathrm{S}_k\mathrm{D}} x_{\mathrm{S}_k\mathrm{D}}(t)$$

$$F_{\mathrm{R}_i}(t) = \sum_{k=1}^{n} \frac{1+\delta_{ki}}{2} f_{\mathrm{R}_k\mathrm{R}_i} x_{\mathrm{R}_k\mathrm{R}_i}(t) + \frac{1}{2}\sum_{k=0}^{n} f_{\mathrm{R}_i\mathrm{S}_k} x_{\mathrm{R}_i\mathrm{S}_k}(t) + \sum_{k=1}^{i} p_{\mathrm{R}_k\mathrm{D,R}_i} f_{\mathrm{R}_k\mathrm{D}} x_{\mathrm{R}_k\mathrm{D}}(t) + \sum_{k=0}^{i-1} p_{\mathrm{S}_k\mathrm{D,R}_i} f_{\mathrm{S}_k\mathrm{D}} x_{\mathrm{S}_k\mathrm{D}}(t)$$

where δ_{ki} is the Kronecker δ. $x_{IJ}(t)$ denotes the frequency of individuals (at time t) with genotype IJ, where I, J = D, S_0, S_1, …, S_n, R_1, …, R_n. Similarly, f_{IJ} is the fitness of IJ individuals, and $p_{IJ,\mathrm{K}}$ denotes the probability of an individual with genotype IJ producing a K gamete. From conservation of probability, we have the following identities

$$p_{\mathrm{R}_k\mathrm{D,D}} + \sum_{i=k}^{n} p_{\mathrm{R}_k\mathrm{D,R}_i} = 1$$

$$p_{\mathrm{S}_k\mathrm{D,D}} + \sum_{i=k}^{n} p_{\mathrm{S}_k\mathrm{D,S}_i} + \sum_{i=k+1}^{n} p_{\mathrm{S}_k\mathrm{D,R}_i} = 1$$

Notice that type R_nD and type S_nD individuals are fully resistant to being manipulated by the drive construct; such a fully resistant individual shows standard Mendelian segregation in its production of gametes. Thus, we have $p_{\mathrm{R}_n\mathrm{D,R}_n} = p_{\mathrm{S}_n\mathrm{D,S}_n} = {}^1/_2$.

The selection dynamics are modeled by the following system of equations

$$\begin{aligned}
\dot{x}_{\mathrm{DD}}(t) &= F_{\mathrm{D}}^2(t) - \psi^2(t)x_{\mathrm{DD}}(t)\\
\dot{x}_{\mathrm{R}_i\mathrm{D}}(t) &= 2F_{\mathrm{R}_i}(t)F_{\mathrm{D}}(t) - \psi^2(t)x_{\mathrm{R}_i\mathrm{D}}(t)\\
\dot{x}_{\mathrm{S}_i\mathrm{D}}(t) &= 2F_{\mathrm{S}_i}(t)F_{\mathrm{D}}(t) - \psi^2(t)x_{\mathrm{S}_i\mathrm{D}}(t)\\
\dot{x}_{\mathrm{R}_i\mathrm{S}_j}(t) &= 2F_{\mathrm{R}_i}(t)F_{\mathrm{S}_j}(t) - \psi^2(t)x_{\mathrm{R}_i\mathrm{S}_j}(t)\\
\dot{x}_{\mathrm{R}_i\mathrm{R}_j}(t) &= (2-\delta_{ij})F_{\mathrm{R}_i}(t)F_{\mathrm{R}_j}(t) - \psi^2(t)x_{\mathrm{R}_i\mathrm{R}_j}(t)\\
\dot{x}_{\mathrm{S}_i\mathrm{S}_j}(t) &= (2-\delta_{ij})F_{\mathrm{S}_i}(t)F_{\mathrm{S}_j}(t) - \psi^2(t)x_{\mathrm{S}_i\mathrm{S}_j}(t)
\end{aligned}$$

The quantity $\psi^2(t)$ represents a density-dependent death rate for the individuals in the population. At any given time, t, we require that the total number of individuals sums to 1

$$x_{\mathrm{DD}}(t) + \sum_{i=1}^{n} x_{\mathrm{R}_i\mathrm{D}}(t) + \sum_{i=0}^{n} x_{\mathrm{S}_i\mathrm{D}}(t) + \sum_{i=1}^{n}\sum_{j=0}^{n} x_{\mathrm{R}_i\mathrm{S}_j}(t) + \sum_{i=1}^{n}\sum_{j=1}^{i} x_{\mathrm{R}_i\mathrm{R}_j}(t) + \sum_{i=0}^{n}\sum_{j=0}^{i} x_{\mathrm{S}_i\mathrm{S}_j}(t) = 1$$

To enforce this density constraint, we set

$$\psi(t) = F_{\mathrm{D}}(t) + \sum_{i=1}^{n} F_{\mathrm{R}_i}(t) + \sum_{i=0}^{n} F_{\mathrm{S}_i}(t)$$

For further details about the model, as well as derivations of our invasion and stability conditions, please see sections S2 and S7.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/3/4/e1601964/DC1

section S1. Previous work on homing endonuclease gene drives
section S2. Model for the evolutionary dynamics of a CRISPR gene drive with n gRNAs
section S3. Invasion of the drive construct
section S4. Stability of the drive construct
section S5. Interior equilibria
section S6. Numerical examples
section S7. Neutral resistance
fig. S1. Numerical simulations of the evolutionary dynamics.

REFERENCES AND NOTES

1. A. Burt, Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc. Biol. Sci.* **270**, 921–928 (2003).
2. K. M. Esvelt, A. L. Smidler, F. Catteruccia, G. M. Church, Concerning RNA-guided gene drives for the alteration of wild populations. *Elife* **3**, e03401 (2014).
3. O. S. Akbari, H. J. Bellen, E. Bier, S. L. Bullock, A. Burt, G. M. Church, K. R. Cook, P. Duchek, O. R. Edwards, K. M. Esvelt, V. M. Gantz, K. G. Golic, S. J. Gratz, M. M. Harrison, K. R. Hayes, A. A. James, T. C. Kaufman, J. Knoblich, H. S. Malik, K. A. Matthews, K. M. O'Connor-Giles, A. L. Parks, N. Perrimon, F. Port, S. Russell, R. Ueda, J. Wildonger, Safeguarding gene drive experiments in the laboratory. *Science* **349**, 927–929 (2015).
4. K. A. Oye, K. Esvelt, E. Appleton, F. Catteruccia, G. Church, T. Kuiken, S. B.-Y. Lightfoot, J. McNamara, A. Smidler, J. P. Collins, Regulating gene drives. *Science* **345**, 626–628 (2014).
5. V. M. Gantz, N. Jasinskiene, O. Tatarenkova, A. Fazekas, V. M. Macias, E. Bier, A. A. James, Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6736–E6743 (2015).
6. A. Hammond, R. Galizi, K. Kyrou, A. Simoni, C. Siniscalchi, D. Katsanos, M. Gribble, D. Baker, E. Marois, S. Russell, A. Burt, N. Windbichler, A. Crisanti, T. Nolan, A CRISPR-Cas9 gene

drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nat. Biotechnol.* **34**, 78–83 (2015).
7. J. E. DiCarlo, A. Chavez, S. L. Dietz, K. M. Esvelt, G. M. Church, Safeguarding CRISPR-Cas9 gene drives in yeast. *Nat. Biotechnol.* **33**, 1250–1255 (2015).
8. V. M. Gantz, E. Bier, The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations. *Science* **348**, 442–444 (2015).
9. S. P. Sinkins, F. Gould, Gene drive systems for insect disease vectors. *Nat. Rev. Genet.* **7**, 427–435 (2006).
10. N. Windbichler, M. Menichelli, P. A. Papathanos, S. B. Thyme, H. Li, U. Y. Ulge, B. T. Hovde, D. Baker, R. J. Monnat, A. Burt, A. Crisanti, A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature* **473**, 212–215 (2011).
11. O. S. Akbari, K. D. Matzen, J. M. Marshall, H. Huang, C. M. Ward, B. A. Hay, A synthetic gene drive system for local, reversible modification and suppression of insect populations. *Curr. Biol.* **23**, 671–677 (2013).
12. L. Alphey, Genetic control of mosquitoes. *Annu. Rev. Entomol.* **59**, 205–224 (2014).
13. B. Charlesworth, C. H. Langley, The population genetics of Drosophila transposable elements. *Annu. Rev. Genet.* **23**, 251–287 (1989).
14. C.-H. Chen, H. Huang, C. M. Ward, J. T. Su, L. V. Schaeffer, M. Guo, B. A. Hay, A synthetic maternal-effect selfish genetic element drives population replacement in *Drosophila*. *Science* **316**, 597–600 (2007).
15. C. M. Ward, J. T. Su, Y. Huang, A. L. Lloyd, F. Gould, B. A. Hay, *Medea* selfish genetic elements as tools for altering traits of wild populations: A theoretical analysis. *Evolution* **65**, 1149–1162 (2011).
16. T. W. Lyttle, Segregation distorters. *Annu. Rev. Genet.* **25**, 511–557 (1991).
17. B. Charlesworth, D. L. Hartl, Population dynamics of the segregation distorter polymorphism of *Drosophila melanogaster*. *Genetics* **89**, 171–192 (1978).
18. Y. Tao, D. L. Hartl, C. C. Laurie, Sex-ratio segregation distortion associated with reproductive isolation in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13183–13188 (2001).
19. S. Henikoff, K. Ahmad, H. S. Malik, The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science* **293**, 1098–1102 (2001).
20. A. Deredec, H. C. J. Godfray, A. Burt, Requirements for effective malaria control with homing endonuclease genes. *Proc. Natl. Acad. Sci. U.S.A.* **108**, E874–E880 (2011).
21. A. Deredec, A. Burt, H. C. J. Godfray, The population genetics of using homing endonuclease genes in vector and pest management. *Genetics* **179**, 2013–2026 (2008).
22. N. Windbichler, P. A. Papathanos, F. Catteruccia, H. Ranson, A. Burt, A. Crisanti, Homing endonuclease mediated gene targeting in *Anopheles gambiae* cells and embryos. *Nucleic Acids Res.* **35**, 5922–5933 (2007).
23. M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
24. P. Mali, L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville, G. M. Church, RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
25. L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, F. Zhang, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
26. P. Mali, K. M. Esvelt, G. M. Church, Cas9 as a versatile tool for engineering biology. *Nat. Methods* **10**, 957–963 (2013).
27. J. A. Doudna, E. Charpentier, The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
28. M. J. Lajoie, S. Kosuri, J. A. Mosberg, C. J. Gregg, D. Zhang, G. M. Church, Probing the limits of genetic recoding in essential genes. *Science* **342**, 361–363 (2013).
29. M. J. Lajoie, A. J. Rovner, D. B. Goodman, H.-R. Aerni, A. D. Haimovich, G. Kuznetsov, J. A. Mercer, H. H. Wang, P. A. Carr, J. A. Mosberg, N. Rohland, P. G. Schultz, J. M. Jacobson, J. Rinehart, G. M. Church, F. J. Isaacs, Genomically recoded organisms expand biological functions. *Science* **342**, 357–360 (2013).
30. N. Ostrov, M. Landon, M. Guell, G. Kuznetsov, J. Teramoto, N. Cervantes, M. Zhou, K. Singh, M. G. Napolitano, M. Moosburner, E. Shrock, B. W. Pruitt, N. Conway, D. B. Goodman, C. L. Gardner, G. Tyree, A. Gonzales, B. L. Wanner, J. E. Norville, M. J. Lajoie, G. M. Church, Design, synthesis, and testing toward a 57-codon genome. *Science* **353**, 819–822 (2016).
31. K. Wang, J. Fredens, S. F. Brunner, S. H. Kim, T. Chia, J. W. Chin, Defining synonymous codon compression schemes by genome recoding. *Nature* **539**, 59–64 (2016).
32. O. S. Akbari, C.-H. Chen, J. M. Marshall, H. Huang, I. Antoshechkin, B. A. Hay, Novel synthetic *Medea* selfish genetic elements drive population replacement in *Drosophila*; a theoretical exploration of *Medea*-dependent population suppression. *ACS Synth. Biol.* **3**, 915–928 (2014).
33. T. F. C. Mackay, S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, D. Zhu, S. Casillas, Y. Han, M. M. Magwire, J. M. Cridland, M. F. Richardson, R. R. H. Anholt, M. Barrón, C. Bess, K. P. Blankenburg, M. A. Carbone, D. Castellano, L. Chaboub, L. Duncan, Z. Harris, M. Javaid, J. C. Jayaseelan, S. N. Jhangiani, K. W. Jordan, F. Lara, F. Lawrence, S. L. Lee, P. Librado, R. S. Linheiro, R. F. Lyman, A. J. Mackey, M. Munidasa, D. M. Muzny, L. Nazareth, I. Newsham, L. Perales, L.-L. Pu, C. Qu, M. Ràmia, J. G. Reid, S. M. Rollmann, J. Rozas, N. Saada, L. Turlapati, K. C. Worley, Y.-Q. Wu, A. Yamamoto, Y. Zhu, C. M. Bergman, K. R. Thornton, D. Mittelman, R. A. Gibbs, The *Drosophila melanogaster* genetic reference panel. *Nature* **482**, 173–178 (2012).
34. M. Nei, W. H. Li, Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5269–5273 (1979).
35. M. A. Nowak, *Evolutionary Dynamics* (Harvard Univ. Press, 2006).
36. R. L. Unckless, A. G. Clark, P. W. Messer, Evolution of resistance against CRISPR/Cas9 gene drive. *Genetics* **205**, 827–841 (2017).
37. J. J. Bull, Lethal gene drive selects inbreeding. *Evol. Med. Public Health* **2017**, 1–16 (2017).
38. J. L. Rasgon, Multi-locus assortment (MLA) for transgene dispersal and elimination in mosquito populations. *PLOS ONE* **4**, e5833 (2009).
39. C. Noble, J. Min, J. Olejarz, J. Buchthal, A. Chavez, A. L. Smidler, E. A. DeBenedictis, G. M. Church, M. A. Nowak, K. M. Esvelt, Daisy-chain gene drives for the alteration of local populations. *bioRxiv* (2016).

Acknowledgments

Funding: C.N. received support from the NSF Graduate Research Fellowship Program under grant no. DGE1144152 and from the Lynch Foundation. K.M.E. received support from the Massachusetts Institute of Technology Media Lab. **Author contributions:** M.A.N., G.M.C., and K.M.E. conceived the study. C.N., J.O., and M.A.N. created the evolutionary dynamics model, and all authors analyzed the results. C.N. and J.O. wrote the manuscript with contributions from all authors. **Competing interests:** K.M.E. and G.M.C. are authors of provisional and nonprovisional patents on RNA-guided gene drive systems and CRISPR multiplexing. K.M.E. has pending patents related to the described work: WO2015105928A1 ("RNA-guided gene drives") and WO2015006294A2 ("Orthogonal Cas9 proteins for RNA-guided gene regulation and editing"). G.M.C. has the following patents related to this work: WO2015006290A1 ("Multiplex RNA-guided genome engineering") and WO2016089866A1 ("RNA-guided systems for in vivo gene editing"). **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 19 August 2016
Accepted 10 February 2017
Published 5 April 2017
10.1126/sciadv.1601964

Citation: C. Noble, J. Olejarz, K. M. Esvelt, G. M. Church, M. A. Nowak, Evolutionary dynamics of CRISPR gene drives. *Sci. Adv.* **3**, e1601964 (2017).

Science Advances

Evolutionary dynamics of CRISPR gene drives

Charleston Noble, Jason Olejarz, Kevin M. Esvelt, George M. Church and Martin A. Nowak

Sci Adv **3** (4), e1601964.
DOI: 10.1126/sciadv.1601964