© 2019. Published by The Company of Biologists Ltd | Biology Open (2019) 8, bio037762. doi:10.1242/bio.037762

RESEARCH ARTICLE

Integral gene drives for population replacement

Alexander Nash[1], Giulia Mignini Urdaneta[1], Andrea K. Beaghton[1], Astrid Hoermann[1], Philippos Aris Papathanos[2,3], George K. Christophides[1,*] and Nikolai Windbichler[1,*]

ABSTRACT

A first generation of CRISPR-based gene drives has now been tested in the laboratory in a number of organisms, including malaria vector mosquitoes. Challenges for their use in the area-wide genetic control of vector-borne disease have been identified, including the development of target site resistance, their long-term efficacy in the field, their molecular complexity, and practical and legal limitations for field testing of both gene drive and coupled anti-pathogen traits. We have evaluated theoretically the concept of integral gene drive (IGD) as an alternative paradigm for population replacement. IGDs incorporate a minimal set of molecular components, including drive and anti-pathogen effector elements directly embedded within endogenous genes – an arrangement that in theory allows targeting functionally conserved coding sequences without disrupting their function. Autonomous and non-autonomous IGD strains could be generated, optimized, regulated and imported independently. We performed quantitative modeling comparing IGDs with classical replacement drives and show that selection for the function of the hijacked host gene can significantly reduce the establishment of resistant alleles in the population, while drive occurring at multiple genomic loci prolongs the duration of transmission blockage in the face of pre-existing target site variation. IGD thus has potential as a more durable and flexible population replacement strategy.

KEY WORDS: Gene drive, Population replacement, CRISPR, Genome editing, Population modeling

INTRODUCTION

Homing gene drives were first proposed 15 years ago as potential tools for enabling the genetic engineering of natural populations (Burt, 2003), in particular of disease vectors. Gene drives are aimed at either reducing vector population size (population suppression) by imposing a fitness load via the disruption of important vector genes, or by modifying the vector's ability to transmit disease (population replacement). After showing promise in a number of proof-of-principle studies (Chan et al., 2011, 2013; Gantz and Bier, 2015; Simoni et al., 2014; Windbichler et al., 2011), first implementations highlighting their potential use to control disease vectors were demonstrated in two species of malaria vector mosquitoes (Gantz et al., 2015; Hammond et al., 2016). Gene drive research is currently focused on two main areas: (1) studying the nature of target site resistance (Champer et al., 2017; Hammond et al., 2017; KaramiNejadRanjbar et al., 2018) to mitigate its eventual rise; and, conversely, (2) reducing or counteracting the invasive potential of gene drives, in order to minimize the perceived or actual risk associated with the technology. The former strand of research is centered on improving regulatory elements to contain/confine nuclease activity to homing-relevant cell types, identifying target sites that are intolerant to drive-inactivating mutations (Kyrou et al., 2018), and on the addition of further components to the drive constructs, such as multiple guide RNAs (gRNAs) (Champer et al., 2018; Marshall et al., 2017; Noble et al., 2017) or factors that bias repair towards the desired homology-directed pathway (Basu et al., 2015). For limiting gene drive invasiveness, a number of schemes have been proposed; for example, linking multiple driving and non-driving CRISPR/Cas9 transgenes into a chain in which the spread of each construct depends on the prior link in the chain (Esvelt and Gemmell, 2017; Noble et al., 2016).

Although gene drive research is now mainly centered on drives built using the CRISPR genome-editing toolset, proposed strategies adhere to the classic transgene paradigm, namely the use of pre-characterized promoter and terminator elements, each driving tissue-specific transgene expression as required for different functions in the germline (drive) and various somatic tissues (anti-pathogen effect), and with the resulting constructs inserted at arbitrarily chosen genomic sites, e.g. genes that are presumed to be neutral (Fig. 1A). It has been observed that the resulting complex and large constructs can show unexpected behaviors (Gantz et al., 2015; Hammond et al., 2016; Oberhofer et al., 2018), e.g. unexpected fitness effects and reduced homing in hemizygous females as well as incomplete homing events. Such effects could result from maternal deposition of components, the interaction between engineered components, their non-native genomic context, and the fact that isolated regulatory elements may not fully recapitulate expression patterns of the endogenous loci they were derived from (e.g. resulting in leaky expression). Moreover, while a number of modeling studies have mapped out the ideal characteristics and the resulting predicted theoretical behavior of gene drives, the practical implications and limitations for the construction of gene drives based on those schemes are often neglected. For example, each and every one of the molecular components (promoters, gRNAs, fluorescent markers, etc.) of complex constructs, such as those that carry multiple anti-pathogen effectors or those designed around the use of multi-gRNA arrays (Marshall et al., 2017; Noble et al., 2017), all have to satisfy regulatory requirements. Along the same lines, limiting the propagation of gene drives by inserting them into repetitive genomic regions, while attractive in principle, presents a formidable genome engineering challenge (Min et al., 2017).

[1]Department of Life Sciences, Imperial College London, Sir Alexander Fleming Building, South Kensington Campus, London SW7 2AZ, UK. [2]Centre of Functional Genomics, Department of Experimental Medicine, University of Perugia, Perugia 06123, Italy. [3]Department of Entomology, Robert H. Smith Faculty of Agriculture, Food and Environment, Hebrew University of Jerusalem, Rehovot 7610001, Israel.

*Authors for correspondence (g.christophides@imperial.ac.uk; n.windbichler@imperial.ac.uk)

A.N., 0000-0002-7767-8886; P.A.P., 0000-0001-8508-4554; G.K.C., 0000-0002-3323-1687; N.W., 0000-0001-9896-1165

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Received 14 August 2018; Accepted 21 November 2018

Downloaded from http://bio.biologists.org/ by guest on February 13, 2019

Biology Open (2019) 8, bio037762. doi:10.1242/bio.037762

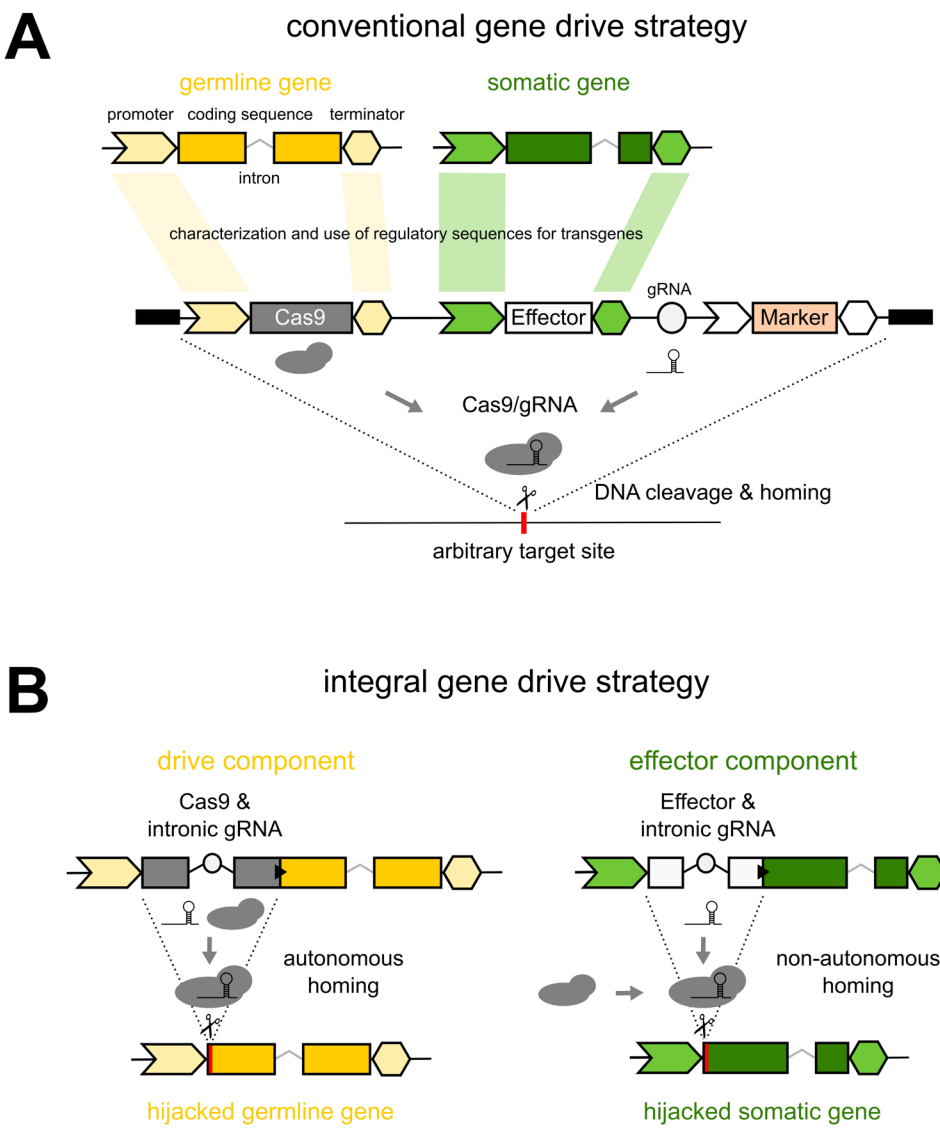

Fig. 1. Schematic overview of the molecular constructs enabling both conventional (A) and integral (B) gene drive approaches for population replacement. The black triangles indicate the 2A ribosomal skipping signal and the circles indicate the gRNA locus. A marker gene is shown for the conventional but not the integral gene drive strategy, but is no requirement for either.

Blocking parasite development in genetically modified mosquito vectors is an area intensively researched long before efficient gene drive had first been demonstrated (Ito et al., 2002). The modification of vector genes involved in immunity and vector-parasite interactions or the introduction of exogenous effectors, such as antimicrobial peptides and antibodies specifically targeting the parasite, are the two cardinal approaches to interfere with *Plasmodium* transmission. A growing set of anti-pathogen effectors now exists (Corby-Harris et al., 2010; Isaacs et al., 2011; Ito et al., 2002; Jasinskiene et al., 2007; Kim et al., 2004), yet these traits have been assessed exclusively against laboratory strains of *Plasmodium falciparum* or the rodent parasite *Plasmodium berghei*. Thus, the efficacy of these effectors against genetically diverse polymorphic isolates of the parasite is currently unknown. The necessary experiments can realistically only be performed in a disease-endemic setting as they require the recruitment of gametocyte carriers from the human population. However, population replacement strains, as currently envisioned, carry one or multiple anti-pathogen effector traits directly coupled to the endonuclease (Gantz et al., 2015) and hence cannot be tested in the absence of gene drive, complicating this crucial step. Alternatively, standard transgenic effector strains must first be generated and used to perform these pilot experiments, which would require further genetic engineering steps (possibly altering their properties) to enable gene drive later on. Here, we evaluate a novel strategy, integral gene drive (IGD), for implementing population replacement that is summarized in Fig. 1B. IGD is specifically conceived with the aforementioned molecular, population-dynamic, practical, operational and regulatory challenges in mind.

RESULTS

IGD drive components

In contrast to the design of conventional population replacement constructs (Fig. 1A), IGDs integrate the endonuclease coding sequence (e.g. Cas9) directly within an endogenous gene, the function and expression of which is confined to the male and/or female germline where homing occurs (Fig. 1B). The presence of Cas9 should ideally have no significant negative effect on the expression of the hijacked host gene. To guarantee accurate translation of both Cas9 and the endogenous host gene, their open reading frames are linked via the 2A ribosome-skipping signal, resulting in the production of two distinct functional polypeptide chains from a compound transcript. A similar strategy has been successfully used to generate endogenously driven reporter genes

Downloaded from http://bio.biologists.org/ by guest on February 13, 2019

Biology Open (2019) 8, bio037762. doi:10.1242/bio.037762

(Rojas-Fernandez et al., 2015); however, it needs to demonstrated that such an arrangement can indeed be neutral with respect to the host gene and at the same time allow sufficient expression of Cas9. For example, inefficient processing has been observed with multicistronic transgenes (Cruz-Teran et al., 2017; Liu et al., 2017). Both an N-terminal as well as a C-terminal fusion of Cas9 to the host gene is possible. However, one consequence of the former arrangement is that incomplete homing events or frameshift mutations inactivating Cas9 would also lead to the loss of function of the hijacked host gene. Therefore, selection at the population level would be expected to maintain the integrity of the Cas9 open reading frame, although it may still be inactivated by point mutations.

The regulatory elements of a number of germline genes have so far been validated in flies and mosquitoes. These include promoter elements derived from the *vasa* (Papathanos et al., 2009), *nanos* (Calvo et al., 2005) or *beta2-tubulin* (Catteruccia et al., 2005) genes and other loci (Akbari et al., 2014). These genes are thus ideal candidates for serving as hijacked host genes for the drive component. As shown in Fig. 1B, the gRNA expression cassette can be located within an intron located inside the Cas9 coding sequence. Intronic gRNAs have been demonstrated previously (Ding et al., 2018; Kiani et al., 2014) and we recently explored this concept in insects (A.N. and N.W., unpublished). Intronic gRNAs can either be promoterless and thus mirror the expression pattern of the hijacked host gene, or they can have their own RNA polymerase III promoter element for ubiquitous expression (e.g. in cases the host gene is not itself expressed in the germline as in the case of the effector component described below). In all cases, targeted cleavage mediated by the gRNA and Cas9, which associate with each other in the germline, triggers homing. The drive component is thus an autonomously homing allele of an endogenous gene and is designed to spread in a population (at the expense of wild-type alleles of the same gene) with no other intended effect than seeding it with and increasing the allele frequency of the coupled Cas9 trait.

IGD effector components

For simplicity, we consider here the effector to be an exogenous polypeptide that when expressed in the target tissue reduces or abolishes parasite development in the mosquito. Various mosquito tissues – such as the midgut, the hemocytes and the salivary glands – are at the interface of the vector-parasite interactions and are thus ideal candidates for hosting the expression of effectors. The fat body is another good candidate tissue, as its secretions into the hemocoel can directly interact with (Ito et al., 2002) parasite oocyst and sporozoite stages. Analogous to the drive component, each effector is incorporated into an endogenous gene expressed in any of the above target tissues (Fig. 1B), thereby hijacking that gene for the use of its regulatory regions. In mosquitoes, a limited number of regulatory regions driving transgene expression in these target tissues have been characterized, including the carboxypeptidase promoter shown to drive transgene expression in the midgut following a blood meal and the vitellogenin promoter driving blood-meal-induced expression in the fat body. However, genome-wide expression analyses have identified numerous additional genes specifically expressed in these tissues (Giraldo-Calderón et al., 2015). Indeed, IGD sidesteps the laborious process of first experimentally testing the temporal and spatial specificity of isolated promoter sequences and instead directly utilizes any such suitable loci for targeted insertion of effector transgenes, although a limited number of target sites may be available to achieve this. Again, an N-terminal fusion of the effector transgene with respect to the host gene can guard against incomplete homing events or frameshift mutations that lead to the loss of the function of the host gene, maintaining the integrity of the effector. Another approach for achieving hijacking would be the use of intein splicing (Fig. S1). Each IGD effector transgene also carries a ubiquitously expressed U6-driven intronic gRNA for triggering homing in the germline. Unlike the drive component, however, the effector component does not encode an endonuclease and thus is not able to initiate gene drive on its own. Indeed, targeted cleavage mediated by the gRNA and Cas9 complex can only occur when the latter is provided by the drive component in the germline. Cleavage in the germline triggers homing of the effector component, which is a non-autonomously homing allele of an endogenous gene and is designed to spread in a population, thereby increasing the allele frequency of the coupled effector trait. We define non-autonomous gene drive as the ability of a certain allele, encoding a gRNA that renders itself capable of homing, to only show gene drive, provided that a Cas9 endonuclease is supplied by another genetic element.

Developing IGD traits

Drive and effector traits inserted at various suitable loci throughout the genome could be generated and tested independently from each other, including by different research teams. On the one hand, the drive component can be specifically optimized in the laboratory for its propensity to induce homing, the faithfulness of Cas9 expression and to minimize the rate at which cleavage triggers the undesirable non-homologous end joining (NHEJ) or microhomology-mediated repair (MHMR) pathways. In addition, one would seek to reduce the fitness cost incurred by the expression of Cas9 itself or by its integration interfering with the function of the hijacked host gene. On the other hand, the effector component can be optimized in the laboratory for its efficacy in reducing or blocking parasite/pathogen development and transmission, its own intrinsic fitness cost and any negative effect it may have on the expression of the hijacked host gene. In addition, the rates of non-autonomous homing of the effector component, assisted by the presence of the IGD drive component or another source of Cas9, can be measured under laboratory containment. Laboratory crossing of the transgenic strains harboring the IGD drive and effector components would allow assessing the likely performance of these trait combinations in the field, including the spread of each modified allele and their resulting frequency in cage populations. It should be noted that the existence of independent IGD strains requires them to be appropriately isolated to minimize the risk of them combining before it is desired.

Exploring IGD population dynamics

To predict the behavior of one or multiple interacting IGD traits at the population level, we used a discrete-generation (non-overlapping) model, comparing the dynamics of a classic replacement drive (Beaghton et al., 2017b) to the dynamics of IGD, analyzing protection levels and allelic dynamics over time: although mosquito populations in the wild are overlapping, for this analysis we use a computationally more efficient discrete-generation model versus a continuous time model due to the complexity of the IGD strategy (e.g. 1000 genotypes for the three-locus, four-allele model). We have confirmed that the assumption of discrete generations is reasonable here by comparing the model predictions for the classical effector with those of a continuous time model (Beaghton et al., 2017a,b), and we find that results for allele frequency dynamics and duration of protection match closely over the parameter range investigated (data not shown). To facilitate comparison to the conventional

Downloaded from http://bio.biologists.org/ by guest on February 13, 2019

Biology Open (2019) 8, bio037762. doi:10.1242/bio.037762

replacement drive model, we initially constructed a two-locus model with one drive component hijacking a germline gene (nuclease, locus 1) and one effector component hijacking a somatic gene (effector, locus 2). We then extended this to a three-locus two-effector model with a single gene drive component at locus 1 and effector components at two independent loci 2 and 3, assuming – for the sake of simplicity – that the effectors at the two different host genes have the same molecular biology and fitness parameters. As a baseline, we assume that if an individual has at least one effector component at locus 2 or 3, we consider it to be refractory against malaria. We evaluate the effectiveness of these different drive architectures by calculating the duration of protection, which is affected by the probabilities of different molecular processes, such as homing, and the formation or pre-existence of resistant alleles, by the fitness costs of the nuclease and the effector components, and by the efficacy of the effector. Protection is defined as the reduction in vectorial capacity, given by the sum of the genotype frequencies with at least one effector component at either locus times their degree of reduction on vector competence. A baseline set of parameter values (Table 1) was chosen to be consistent with existing published work on mosquitoes (Hammond et al., 2016) and for ease of comparison to the classical replacement drive model (Table 2).

A comparison of a conventional replacement strategy and the IGD two-loci model is summarized in Fig. 2. We find that, using identical baseline values (Tables 1 and 2) to facilitate comparison [e.g. drive transmission, loss-of-function mutations during homology-directed repair (HDR), costs for nuclease and effector expression], the IGD strategy conveys 95% protection for 81 generations, compared to 39 generations for the classical replacement drive. This translates to ∼4.5 years of protection against malaria transmission, while protection given by a classical gene drive lasts ∼2.1 years (Depinay et al., 2004; Mordecai et al., 2013). As in the conventional strategy, resistant alleles are generated at each locus, eventually replacing the constructs, since they do not carry the cost associated with expressing either the nuclease (locus 1) or the effector (locus 2). If the cost (s_n) of expression due to the nuclease is less than the cost (s_{e2}) of expression for the effector (s_n=0.05 versus s_{e2}=0.1), resistance replaces the transgene faster at the effector locus, while the nuclease persists for longer in the population, whereas in the conventional strategy, the compound cost for expressing both causes the construct to be lost rapidly. We find that at the same rate of formation of resistant alleles, their impact is reduced in the IGD strategy since mutations that result in a loss of function of the hijacked endogenous target gene and are selected against – unlike the conventional strategy, which assumes the target to be neutral.

In order to determine which parameters have the strongest effect on the duration of protection, we varied each while retaining all others at baseline values (Fig. S2). We find similar dependencies as for the classical model; however, for a number of parameters IGD appears more robust with minor or no effects on the duration of protection evident. For example, increasing the proportion of resistance and loss-of-function alleles formed by either NHEJ or aberrant HDR at locus 1 within a biologically sensible range does not affect the duration of protection. This can be explained by the main effect the drive component (locus 1) has on locus 2, which is to convert wild-type alleles to effector alleles in heterozygous individuals, allowing the effector to rapidly propagate in the population and to establish protection before resistance at locus 1 takes over. Protection starts decreasing only when resistant alleles start forming at locus 2 and eventually replace effector alleles in the population. The eventual subsequent loss of the drive allele at locus 1 and its replacement by resistant alleles is no longer of any consequence to the duration of protection because the conversion of wild-type alleles at locus 2 into effector has already taken place.

Table 1. Parameters and baseline values for the IGD model

Symbol	Parameter[+]	Baseline value
cc_0	Initial release frequency of homozygous drive and effector construct	10^{-4}
dd_0	Initial release frequency of homozygous drive construct	0
ee_0	Initial release frequency of homozygous effector construct	0
$rr1_0$	Initial frequency of resistance (locus 1)	0
$rr2_0$	Initial frequency of resistance (locus 2)	0
$rr3_0$	Initial frequency of resistance (locus 3)	0
d_1	Transmission rate of drive (locus 1)	0.99
d_{2a}	Transmission rate of drive (locus 2) if drive component present in heterozygosity	0.99
d_{2b}	Transmission rate of drive (locus 2) if drive component present in homozygosity	0.99
d_{3a}	Transmission rate of drive (locus 3) if drive component present in heterozygosity (l)	0.99
d_{3b}	Transmission rate of drive (locus 3) if drive component present in homozygosity (l)	0.99
l_1	Resistance arising from HDR at locus 1	10^{-4}×(1/3)
l_2	Resistance arising from HDR at locus 2	10^{-4}×(1/3)
l_3	Resistance arising from HDR at locus 3 (l)	10^{-4}×(1/3)
u_1	Resistance arising from NHEJ at locus 1	0.5×(1/3)
u_2	Resistance arising from NHEJ at locus 2	0.5×(1/3)
u_3	Resistance arising from NHEJ at locus 3 (l)	0.5×(1/3)
L_1	Loss of gene function arising from HDR at locus 1	10^{-4}×(2/3)
L_2	Loss of gene function arising from HDR at locus 2	10^{-4}×(2/3)
L_3	Loss of gene function arising from HDR at locus 3 (l)	10^{-4}×(2/3)
U_1	Loss of gene function arising from NHEJ at locus 1	0.5×(2/3)
U_2	Loss of gene function arising from NHEJ at locus 2	0.5×(2/3)
U_3	Loss of gene function arising from NHEJ at locus 3 (l)	0.5×(2/3)
s_m	Cost of loss of gene function	1
s_{d1}	Cost of hijacking target locus 1 (drive)	0
s_{d2}	Cost of hijacking target locus 2 (effector)	0
s_{d3}	Cost of hijacking target locus 3 (effector) (l)	0
s_n	Cost of expressing nuclease at locus 1	0.05
s_{e2}	Cost of expressing effector at locus 2	0.1
s_{e3}	Cost of expressing effector at locus 3 (l)	0.1
h_m	Dominance coefficient for loss of gene function	0.2
h_{d1}	Dominance coefficient for hijacking at locus 1	0.5
h_{d2}	Dominance coefficient for hijacking at locus 2	0.5
h_{d3}	Dominance coefficient for hijacking at locus 3 (l)	0.5
h_n	Dominance coefficient for expressing nuclease	0.5
h_{e2}	Dominance coefficient for expressing effector (locus 2)	0.5
h_{e3}	Dominance coefficient for expressing effector (locus 3) (l)	0.5
h_{rc1}	Dominance coefficient for refractoriness (one effector allele)	1
h_{rc2}	Dominance coefficient for refractoriness (two effector alleles) (l)	1
h_{rc3}	Dominance coefficient for refractoriness (three effector alleles) (l)	1
r_c	Homozygous degree of refractoriness	1

l, three-locus model only; HDR; homology-directed repair; NHEJ, non-homologous end joining.

An exemplar set of parameter values that is consistent with most of the extensive published work on mosquitoes (Hammond et al., 2016) has been chosen for both the classical gene drive model (Beaghton et al., 2017a,b) and the IGD drive model. These parameters define various aspects of molecular biology, fitness and vector competence effects, and initial genotype frequencies. Parameters shared by both models are given the same baseline values to facilitate comparison (e.g. drive transmission, loss-of-function mutations during HDR, costs for nuclease and effector expression).

Downloaded from http://bio.biologists.org/ by guest on February 13, 2019

Biology Open (2019) 8, bio037762. doi:10.1242/bio.037762

Table 2. List of parameters and baseline values for the 'classical' model

Symbol	Parameter	Baseline value
cc_0	Initial release frequency of homozygous construct	10^{-4}
d	Transmission rate of drive	0.99
u	Probability of resistance (functional)	0.5
l_n	Probability of nuclease gene loss during homing	10^{-4}
l_e	Probability of effector loss during homing	10^{-4}
l_{ne}	Probability of nuclease and effector loss during homing	10^{-4}
s_d	Cost of target site disruption	0
s_n	Cost of nuclease expression	0.05
s_e	Cost of effector expression	0.1
h_d	Dominance coefficient for target site disruption	0.5
h_n	Dominance coefficient for nuclease expression	0.5
h_e	Dominance coefficient for effector expression	0.5
h_{rc}	Dominance coefficient for refractoriness	1
r_c	Homozygous degree of refractoriness	1

An exemplar set of parameter values, consistent with published work conducted in relation to classical replacement gene drives (Beaghton et al., 2017a,b).

We found the existence of pre-existing resistant alleles at the effector locus among the factors that most significantly reduce the duration of protection (Fig. S3). By contrast, levels of protection begin to crash only when initial resistance at locus 1 approaches 80% i.e. that target sequence represents a minor allele. Should pre-existing resistance alleles occur with a frequency of 10% at both loci, 95% protection is reduced from 81 generations to 15 generations. Pre-existing resistance must be assumed to be present in significant proportions in many target species including mosquitoes (Anopheles gambiae, 1000 Genomes Consortium et al., 2017), and modeling has already shown that resistant alleles arising from standing genetic variation are generally more likely to contribute to resistance than from new mutations induced by the drive (Unckless et al., 2017). Having investigated the effect of pre-existing resistance alleles in a two-locus IGD model, which showed that only the effector locus is particularly sensitive to pre-existing resistance, we considered next our three-locus two-effector model. Deploying two-effector or multi-effector strains should sustain protection for longer, since for the protection to disappear, resistance will need to develop or pre-exist for the effector at both loci in a significant fraction of the population. We find that releasing a two-effector driving strain into a population without pre-existing resistance yields extended protection of 103 generations (Fig. 3A). With pre-existing resistant alleles (10% allele frequency at all three loci), 95% protection lasts for 38 generations (Fig. 3B), a significant increase in the duration of protection when compared to a single-effector strain release under identical conditions.

A second effector component allele at a separate locus could also be introduced after a given time to extend the duration of protection. The second effector is driven through the population when in the presence of the nuclease allele at locus 1, extending the duration of protection until the time when the second effector is in its turn replaced by resistant alleles. Fig. 3C and D show the boost in duration of protection for these baseline parameters (i.e. cost of expression of the nuclease less than that of the effector) when the second effector is added at the time when the level of protection from the first effector has dropped to 67% and to 95%. When protection drops below a certain level, effector release restores maximal protection. The lower the cost of expression of the nuclease (s_n), the longer it will be present in the population and the longer protection can be extended. The condition for full protection to be achieved after release of the second effector at locus 3 is that the allelic frequency of the nuclease is greater than ≅55% at the time of release. If the second effector is released before protection starts

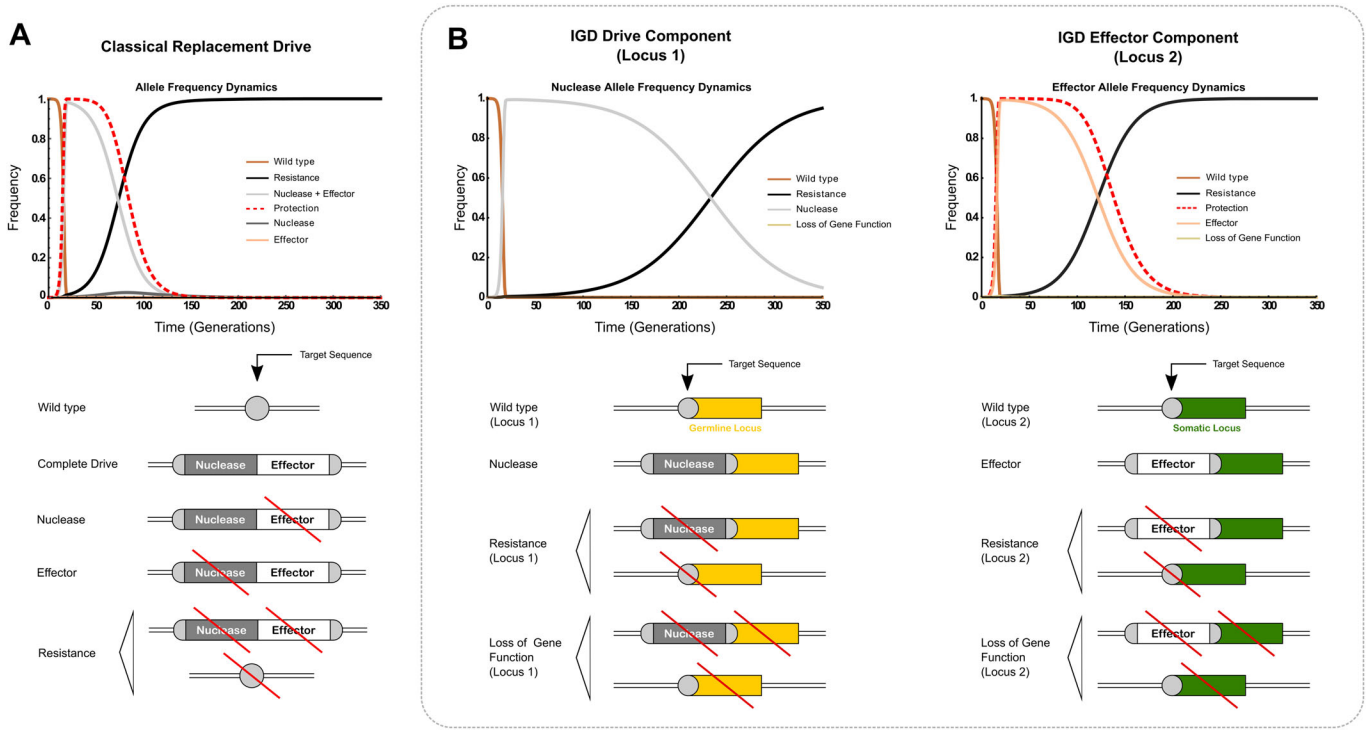

Fig. 2. Comparison of allele frequency dynamics of the conventional (A) and integral (B) gene drive strategies using baseline parameter values. Integral drive dynamics is displayed here in terms of the modeled behavior of the constituent components at the two loci. The dashed red lines in graphs show the proportionate reduction in vectorial capacity of the target population.

Downloaded from http://bio.biologists.org/ by guest on February 13, 2019

Biology Open (2019) 8, bio037762. doi:10.1242/bio.037762

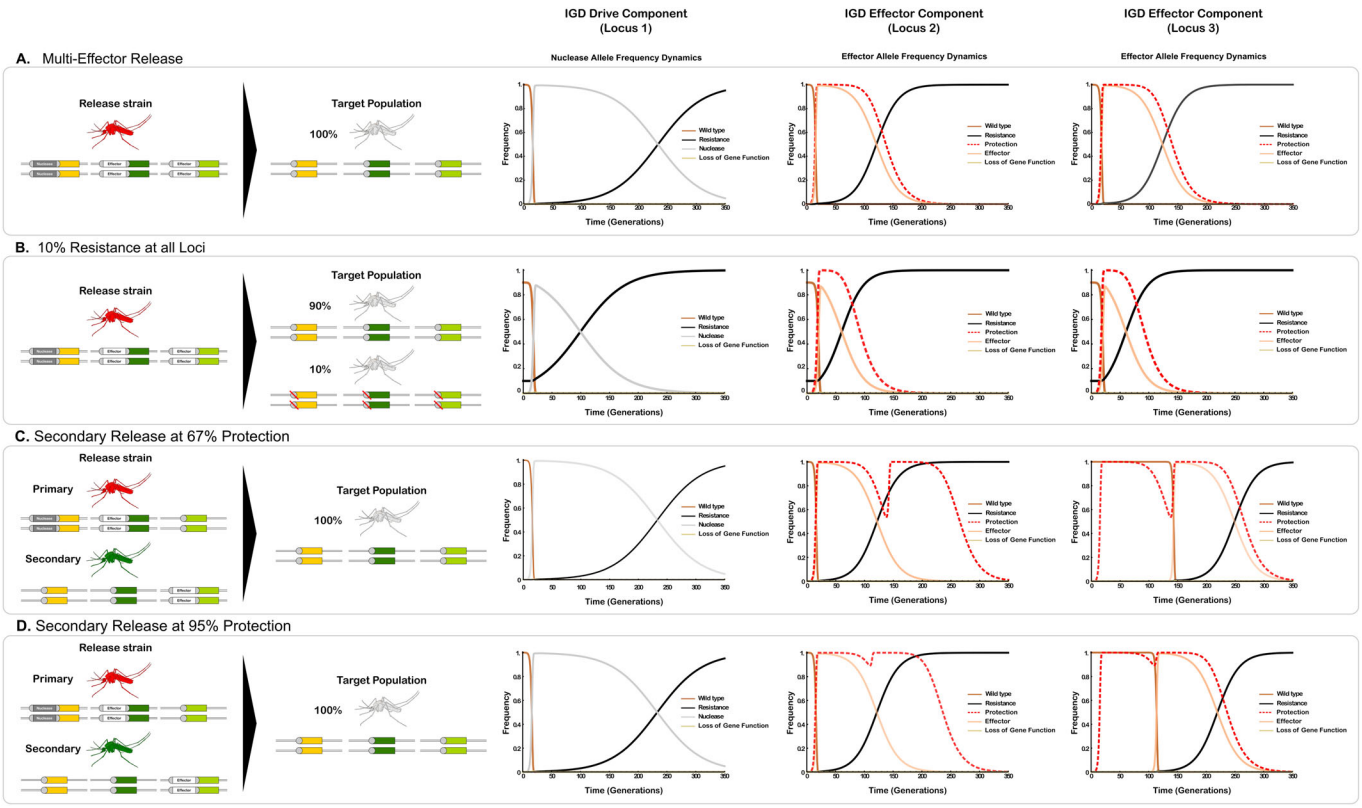

Fig. 3. Dynamics of the three-locus two-effector model. The model assumes the release of transgenics for all loci (locus 1, drive; locus 2 and 3, effectors) into either a population that is wild type for all loci (A) or that carries pre-existing resistance at all loci at 10% allele frequency (B). Staggered release of the second effector strain at 95% (C) and 67% (D) protection, whilst all other parameters are maintained at their respective baseline value. Integral drive dynamics is displayed here in terms of the modeled behavior of the constituent components at the three loci. The dashed red lines in graphs show the proportionate reduction in vectorial capacity of the target population.

significantly decreasing, it can optimally prolong the duration of maximal protection. We find that the ideal time for secondary effector release at locus 3 is when protection drops below 95%.

DISCUSSION

The IGD paradigm reflects our view that a successful gene drive intervention will have to involve the use of multiple interacting traits rather than attempting to make singular constructs evolution safe. It consequently should allow (informed by continuous monitoring) the ability to flexibly react to predictable as well as unexpected developments in a target vector or parasite population by adapting the release strategy in a context-dependent manner. Control would require the constant development and refinement of engineered genetic traits rather than the one-off or continued application of a static product.

Modeling of allele frequency dynamics suggests that for the parameters investigated, the IGD strategy could confer significant advantages over conventional replacement drive designs. By integrating components into endogenous loci, undesired repair outcomes following DNA cleavage are predicted to result in loss of function of the hijacked gene. Selection, in turn, reduces the rate at which such resistant alleles accrue. Whilst IGD does not prevent the onset of resistance, it could significantly extend the duration and level of protection available to the human population. This approach also could resolve the longstanding issue of the arbitrary selection of a target sequence for population replacement where neutral sites show poor conservation, whereas conserved sites are likely to be functionally constrained and thus costly to disrupt. For example, trying to tackle this conundrum, Gantz and colleagues disrupt the kynurenine hydroxylase gene in their replacement approach with unclear consequences for the behavior of these drives in the field (Gantz et al., 2015). The IGD strategy allows the targeting of a coding sequence of an endogenous gene while at the same time the insertion of IGD components is aimed to be neutral with respect to that gene, although it remains to be demonstrated experimentally if this is easily achievable. Notably, naturally occurring homing endonuclease genes, via their association with introns or inteins, propagate in a similar manner, i.e. by targeting highly conserved sites without disrupting their function.

There are several noteworthy assumptions made in the model. It is assumed following previous work (Beaghton et al., 2017b) that at baseline values, the cost of expressing the nuclease from the drive component (s_n=0.05) is lower than the cost of expressing an effector gene from an effector component (s_{e2}=s_{e3}=0.1). This is justifiable when considering that germline genes tend to have lower expression levels than those present in the soma and that expression of antimalarial effectors needs to be sufficiently strong to ensure effective concentrations. Probabilistic models of finite populations (versus our deterministic model for an effectively infinite population) predict that resistant alleles generated at low rates suffer early stochastic loss (Beaghton et al., 2017a), which would be expected to delay the establishment of such an allele and therefore the onset of resistance. Our deterministic model would in that case be conservative, as it would underestimate duration of protection.

We have determined which parameters are key in influencing the duration of protection: the cost of expressing the nuclease (s_n), the

Downloaded from http://bio.biologists.org/ by guest on February 13, 2019

Biology Open (2019) 8, bio037762. doi:10.1242/bio.037762

cost of disrupting the drive component locus (s_{d1}), the cost of expressing the effector (s_e) and the cost of disrupting its corresponding locus (s_{d2}). Modeling shows that changes in the cost of expressing the nuclease and the cost of disrupting the drive component locus are robustly tolerated. By contrast, the IGD strategy is sensitive to changes in the cost for disruption of the effector component locus, and the cost of expression of the effector. This is due to the efficiency with which the drive component, even if not present at high allele frequencies, is able to convert wild-type alleles into effector alleles. Thus, even if the drive component carries significant costs of expression and disruption, there will not be a sizable reduction in the duration of protection. These findings may be useful in the design of IGD traits, particularly with respect to the effector components. Minimizing the cost of effector expression and avoiding disruption of the hijacked gene are critical to the successful implementation of an IGD-based release. Previous work has highlighted the difficulty with which these parameters can be evaluated in a field setting (Adelman and Tu, 2016); however, the modularity of the IGD strategy may be an advantage in this respect.

Our initial models included here must be extended to evaluate the potential of IGD in more realistic settings with full seasonality and spatial heterogeneity. Currently, spatial effects are not considered, and the simulated population is considered to be distributed homogeneously within a contained landscape. Previous work has been conducted to investigate the effects of spatial interactions upon the propagation of replacement drives (Eckhoff et al., 2017) and these models should be extended to IGDs. The inclusion of sex-specific parameters would allow the consideration of sex-specific effectors and associated fitness effects (Beaghton et al., 2017b). This is particularly relevant given that, in order to improve the efficiency and duration of the IGD strategy, it may prove useful to restrict effector activity to females and homing activity to males, reducing the net fitness cost on the population as a whole. Moreover, this approach may help to overcome issues associated with maternal deposition of Cas9 mRNA into the embryo (Champer et al., 2017; Hammond et al., 2017). We limited our analysis to one autonomous drive component and two non-autonomous effector components. More powerful models could explore the dynamics of multiple drive and effector alleles and their interaction over more complex geographical scales and metapopulation structures. A set of separate hijacked host genes could be used in sequence to guarantee that the level of Cas9 in the population remains high even when, at particular loci, resistant alleles predominate in some areas or subpopulations. Equally, multiple effector strains could be generated expressing several effector molecules from different host genes and used to ensure that most individuals in each population carry transmission-blocking effectors acting in various parasite-relevant tissues, even if resistant alleles are in circulation. Our simple two-effector model already indicates that hedging homing over multiple loci in this manner could be a viable strategy to tackle the issues of resistance and standing variation in target populations, particularly as the use of multiple gRNAs at one site may only create a marginal difference to drive behavior while significantly complicating construct design (Champer et al., 2018; Oberhofer et al., 2018). Finally, modeling could explore whether effector components could also be used in conjunction with Cas9-expressing suppressive strains to ensure a reduction in the vectorial capacity of mosquito populations that have been reduced in size but not eliminated. IGD population replacement could thus operate alongside a suppression program and would be a safeguard in the case of a population and transmission bounce following the intervention.

Testing and deployment of IGDs

Recently, the first transgenic mosquito strain, carrying a dominant male-sterilizing transgene (Windbichler et al., 2008), was imported by the Target Malaria consortium to an African partner nation, after completing a prolonged regulatory pathway (https://targetmalaria.org/, accessed 12 October 2018), including an independent ecological risk assessment. It is expected that gene drive strains will face a significantly tougher and prolonged regulatory pathway compared to strains harboring such a conservative genetic modification. These foreseeable regulatory and operational challenges must inform the design of gene drives, and the potential theoretical properties of IGDs should thus be considered in conjunction with the practical aspects of generation, testing, optimization, regulation and deployment of gene drives for population replacement.

Field testing of the effector component would be aided by the simplicity of the genetic modification, i.e. insertion of a single effector coding sequence containing the intronic gRNA, with all other functions provided by the hijacked host gene. Given the inability of the effector component to spread autonomously, the regulatory threshold for such strains is expected to be lower than that of conventional fully driving transgenes harboring a population replacement payload (Fig. 4A). This would facilitate import of these strains to disease-endemic countries and the swift testing of effector traits against polymorphic isolates of the parasite, as well as within varying mosquito genetic backgrounds (Fig. 4B). Strains harboring the IGD drive component are also conceived to be molecularly simple, although they would be more difficult to import, regulate and deploy as they carry autonomous gene drive elements. However, the drive component on its own is not designed to have any phenotypic effect on fitness or vectorial capacity of the mosquito. Thus, an inadvertent release and spread of such a strain would not be expected to have any relevant effect on mosquito biology and carries a relatively lower risk compared to suppressive gene drives or replacement drives that also disrupt mosquito genes (Gantz et al., 2015; Kyrou et al., 2018), although a source of Cas9 (the drive component) would then be present in the population and hence no containment would be afforded to even non-autonomous drive elements carrying compatible guides. The lack of a fitness cost may also mean that such alleles could persist in a population for longer, even in the case of an unintended release.

The scientific working group on a pathway to deployment of gene drive mosquitoes for elimination of Malaria in Sub-Saharan Africa (James et al., 2018) recommends the use of fluorescent markers to track gene drive constructs pre- and post-deployment. However, gene drives can decouple from genetically linked fluorescent marker genes within a single generation of homing (e.g. via incomplete homing events (Hammond et al., 2016; Oberhofer et al., 2018), potentially giving rise to type II errors (false negatives) during monitoring, arguably the most crucial error as it would suggest the absence of constructs in populations or regions in which active gene drives are in fact propagating. Fluorescent markers, other than those used for transgenesis and that can be subsequently removed, are therefore to be avoided in our design and modeling of IGD population replacement, as they also increase molecular and regulatory complexity. We assume that molecular genotyping will be the only viable approach for gene drive monitoring in general. Replacement drives including IGDs, unlike suppressive drives, can be constituted as true-breeding strains, which should facilitate the exclusive use of molecular markers during implementation.

The IGD strategy offers an increased degree of versatility by allowing independent testing and release of overall less complex

Downloaded from http://bio.biologists.org/ by guest on February 13, 2019

Biology Open (2019) 8, bio037762. doi:10.1242/bio.037762

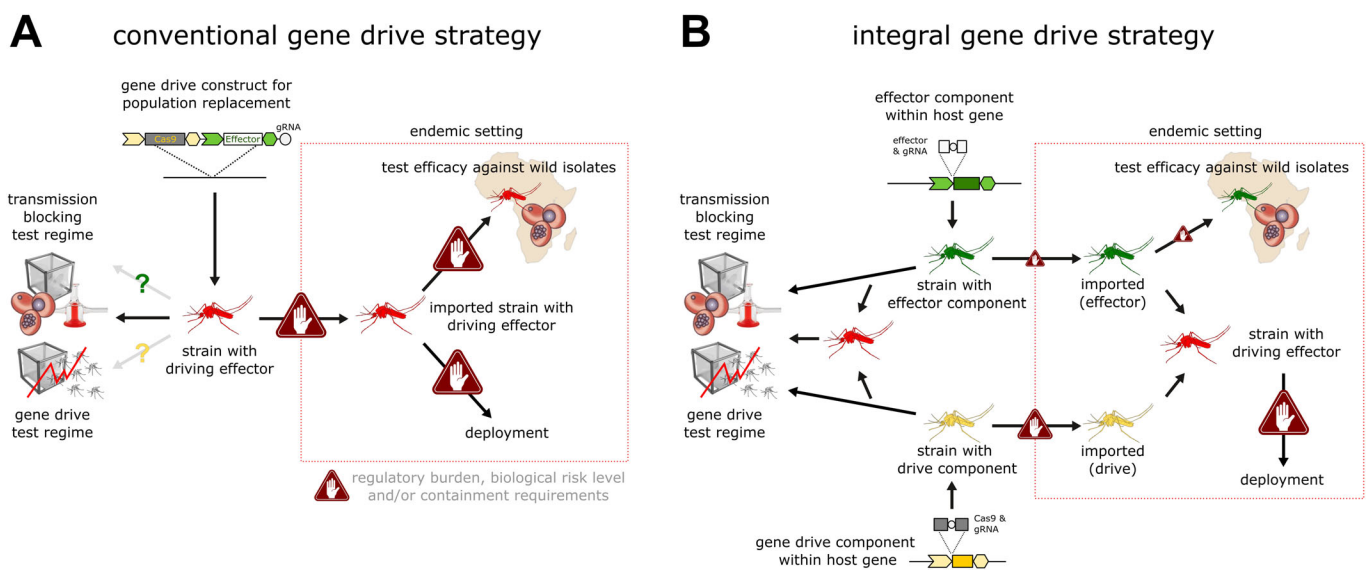

Fig. 4. Comparison of the possible future steps and requirements for laboratory and field testing of both conventional (A) and integral (B) gene drive strategies for population replacement of malaria vectors.

components. The scientific working group on gene drives recommends a stepwise pathway for the deployment of gene drive mosquitoes i.e. to progress testing from laboratory studies, possibly involving large indoor and outdoor cage trials, to small-scale isolated and open releases to full-scale open releases (James et al., 2018). However, it is not always clear how limiting drive propagation can be guaranteed with conventional gene drive designs, as the level of ecological or geographical isolation achievable at different release sites is yet to be fully understood. No obvious pathway for safely testing gene drives has emerged that would allow research to progress step by step from laboratory to field deployment.

The modularity and interdependence of IGD components does provide a straightforward path for moving testing from self-limited to self-sustaining traits in the field by modulating the propensity to spread in the population (Fig. 5). First, an inundative release of an effector strain alone would allow assaying (by rates of recapture) mosquito fitness and performance under field conditions and detection of any unintended effects prior to deployment. When tested in the absence of a drive component, effector strains will not convert the field population, permitting safe testing of individual effector components. A second scenario shown here (Fig. 5) consists of releasing a limited-drive strain, containing an effector component and a non-driving source of Cas9, which can trigger a limited and localized spread of the effector trait and allow evaluation of its drive performance and perhaps its epidemiological effect in reducing disease transmission. This Cas9 source permits super-Mendelian inheritance of effector components within the field population but is itself inherited at a Mendelian rate, and modeling suggests that both would be lost. This strategy therefore facilitates the testing of effector component homing in the field, without the perceived risk posed by using a driving source of Cas9. Modulating the allele frequency of the non-driving Cas9 trait via inundative releases of varying magnitudes would allow control of the expected level of spread and the resulting allele frequency of the effector. The effect of these two first strategies is self-limiting, perhaps allowing a test site to be re-used following the dissipation of the released alleles. When individual traits have been sufficiently tested separately and in conjunction in the laboratory, as well as in self-limiting pilot experiments in the field, one can consider the release of the fully driving strain carrying both drive and effector

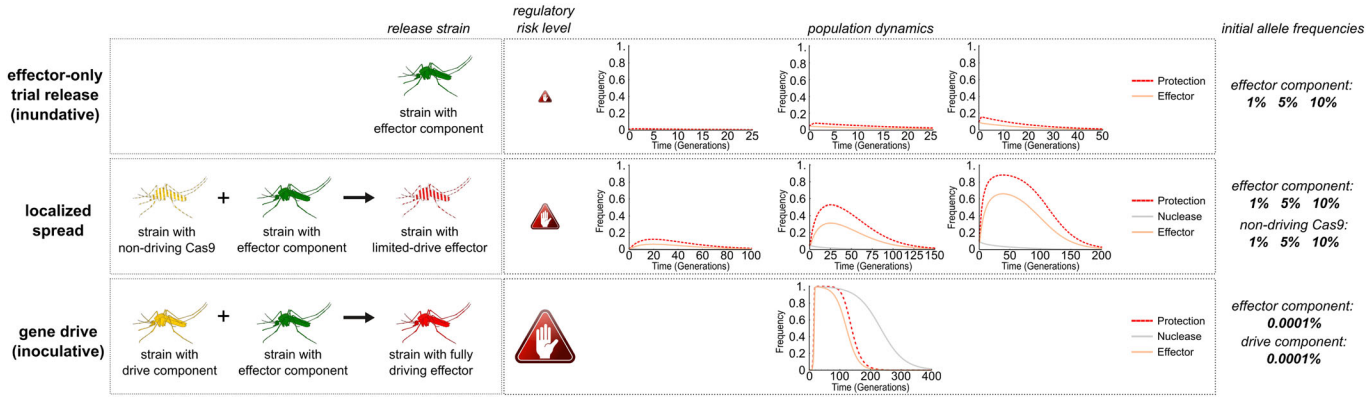

Fig. 5. Evaluation of the allele frequency dynamics of limited release strategies. Release of strains carrying an IGD effector component only (upper row) or of a limited-drive strain combining a non-driving Cas9 locus with an effector (middle row) permit field testing of components in the absence of gene drive with the resulting changes in the target population dependent on the release size practically achievable. Release of the fully driving strain results in replacement of the wild-type population, even when released at low frequencies (lower row).

Downloaded from http://bio.biologists.org/ by guest on February 13, 2019

Biology Open (2019) 8, bio037762. doi:10.1242/bio.037762

component alleles. The release of such a transgenic strain would then trigger full population-wide gene drive in the field and propagation, according to the previously described dynamics. It is important to highlight that, unlike conventional designs, here the performance and behavior of the IGD effector trait is likely to be unchanged by the addition of the drive component. A regular population replacement strategy would require, for various stages of testing, different driving and non-driving constructs to be made that could display significant differences. Modeling of effector-only and limited-drive releases thus suggests how IGD permits safe evaluation of effector components in the field, without need for a driving Cas9 source. This may alleviate concerns relating to the invasive nature of gene drives. From a logistical standpoint, the limited effect that such releases have on a localized test area facilitate the testing of multiple effectors in the same locale, once the presence of previously released transgenes has suitably diminished.

MATERIALS AND METHODS

Discrete-generation population genetics model

The code for all models used in this study is available at https://github.com/genome-traffic/igd. The classical model for a drive+effector construct [see Model I (Beaghton et al., 2017b)] considers five different alleles: the wild-type allele (w), the complete drive construct, which has both a functional nuclease and a functional effector (c=n+e); a nuclease-only construct, which has a functional nuclease but a defective effector (n); an effector-only construct, which has a functional effector but a defective nuclease (e); and functional-resistant alleles (r), which are not recognized by the nuclease and have no functional nuclease or effector. Resistant alleles can either be pre-existing in the population or arise via NHEJ and MHMR repair pathways, as well as incomplete HDR. We define d as the transmission rate of drive, u as the parameter for resistance arising from end-joining repair, and l_n, l_e, and l_{ne} as the probability of loss of function of effector, nuclease or both during HDR. Therefore, allele contributions from germline cleavage and homing from w/c are according to $w{:}c{:}n{:}e{:}r$ in proportions $(1-d)$ $(1-u)$: d $(1-l_n-l_e-l_{ne})$: d (l_e): d (l_n): $(1-d)$ $u+l_{ne}$ d, and in w/n, according to $w{:}n{:}r$ in proportions $(1-d)$ $(1-u)$: d $(1-l_n)$: $(1-d)$ $u+l_n$ d.

While the IGD model may involve multiple nuclease genes and effectors on many different loci, here we consider a simplified version with transgenes on either two or three independent loci i (with i=1,2,3). Locus 1 corresponds to the drive component, and loci 2 and 3 correspond to effector components. At each locus i, there are four possible alleles: a wild-type allele (w_i), a transgene t_i (corresponding to either the nuclease gene as the transgene at the first locus, t_1=n_1, or to an effector component at the second or third locus, t_2=e_2 and t_3=e_3), and two types of alleles at each locus that are resistant to the drive and do not have an intact nuclease or effector, r_i and m_i. The first type of resistant allele, r_i, arises from incomplete homing or mutations that are in-frame and do not cause loss of the function of the host gene, and therefore are not considered to carry any fitness cost (similarly to the wild type). The second type of resistant allele, m_i, corresponds to a mutation that results in frameshift of the host gene, disrupting the endogenous locus. If the host gene is an essential gene, m_i alleles are considered to convey lethality when homozygous. Resistance can be pre-existing or occur either by NHEJ or by incomplete HDR at either locus. We neglect resistant alleles created from spontaneous mutations, as rates are likely to be low, compared to generation of resistance during homing.

An individual is considered to have IGD drive if it carries at least one functional drive component at locus 1 and at least one functional effector component at locus 2 (and at locus 3 if an additional effector is included as part of the strategy). This is a necessary condition for the effector component to home and propagate in the population at a super-Mendelian rate. For both models, we assume that the initial field population consists entirely of the wild-type allele, and there may be pre-existing resistance due to standing genetic variation at each locus (although the baseline pre-existing resistance is set to zero). Individuals homozygous for different IGD drive components are subsequently released as a relative proportion of the field population.

Cleavage and homing can occur only in the germline of genotypes with a wild-type and nuclease allele at locus 1 (w_1/n_1 at the first locus). Cleavage and homing at effector locus 2 (and locus 3 if an additional effector is included at another host gene) can occur only in those genotypes if there is at least one nuclease allele at locus 1 and a wild-type and effector allele at locus 2 (w_2/e_2 at locus 2). Transmission of the nuclease at locus 1 occurs with probability d_1. Transmission of the effector transgene at locus 2 occurs with probability d_{2a} when the drive component is heterozygous, and probability d_{2b} when homozygous, and similarly for an effector at locus 3 if included (d_{3a} d_{3b}). Resistance to the drive, sometimes accompanied by loss of gene function, is considered to occur during homing. We conservatively consider mutations to produce resistance (r_i) in 1/3 of cases, and loss of gene function (m_i) in 2/3, here predominantly caused by frameshift mutations. Resistant alleles (r_i) arise at loci 1, 2 and 3 from incomplete HDR with probabilities l_1, l_2 and l_3, and by NHEJ with probabilities u_1, u_2 and u_3, respectively. Resistant alleles (m_i) that cause loss of the endogenous gene function occur via incomplete HDR at loci 1, 2 and 3 with probabilities L_1, L_2 and L_3, and by NHEJ with probabilities U_1, U_2 and U_3, respectively. Due to germline cleavage, homing, incomplete HDR and repair events, individuals that are heterozygous for the nuclease at locus 1, i.e. w_1/n_1, contribute alleles $w_1{:}n_1{:}r_1{:}m_1$ in proportions $(1-d_1)$ $(1-u_1-U_1)$:d_1 $(1-l_1-L_1)$:$(1-d_1)$ u_1+l_1 d_1:$(1-d_1)$ U_1+L_1 d_1. Allele contributions from individuals with the wild-type and effector allele (w_2/e_2) at locus 2, if there are one or two nuclease alleles at locus 1, are according to $w_2{:}e_2{:}r_2{:}m_2$ in proportions $(1-d_{2k})(1-u_2-U_2)$:d_{2k} $(1-l_2-L_2)$:$(1-d_{2k})$ u_2+l_2 d_{2k}:$(1-d_{2k})$ U_2+L_2 d_2, where k=a, and b corresponds to locus 1 heterozygous or homozygous for the nuclease. If no nuclease allele is present at locus 1, gene transmission at locus 2 is Mendelian, as is inheritance in all other individuals. Similar expression can be written for an additional effector at locus 3.

With four possible alleles at two independent loci, there are 16 gamete types and 100 diploid genotypes; for three independent loci, there are 64 gamete types and 1000 genotypes. The fitness of each genotype is relative to the wild-type homozygote (w_1/w_1; w_2/w_2; w_3/w_3), which has a fitness of one. Fitnesses are modeled using ten parameters for transgenes at two loci and 14 parameters for three loci. We consider the following homozygous fitness costs: the cost of in-frame disruption (as caused by an intact transgene or resistant allele r_i at the host gene) at locus 1, 2 and 3 (s_{d1}, s_{d2}, s_{d3}), the cost of disruption (by mutations of type m_i) that lead to loss of function at each locus (cost assumed to be the same at all loci, s_m), the cost of expressing the nuclease from the drive component (s_n) and the cost of expressing the effector from the effector components (s_{e2}, s_{e3}). The corresponding dominance coefficients are h_{d1}, h_{d2}, h_{d3}, h_m, h_n, h_{e2} and h_{e3}. The range of fitness costs is from 0 (no cost) to 1 (lethal) and dominance coefficients range from 0 (completely recessive) to 1 (completely dominant). The fitness of each genotype is derived as the product of costs at each locus associated with site disruption, number of nuclease components and number of effector components. For example, for a two-loci model (drive at locus 1 and effector at locus 2), the fitness of a genotype that is heterozygous for the transgene at both loci (w_1/n_1; w_2/e_2) is given by $(1-h_{d1}\,s_{d1})$ $(1-h_{d2}\,s_{d2})$ $(1-h_n\,s_n)$ $(1-h_{e2}\,s_{e2})$ to reflect costs of host gene disruption by transgenes (for baseline parameters, this cost is set to zero) at both loci as well as the cost of expressing the nuclease (at locus 1) and the effector (locus 2).

Allele frequencies and genotype abundances are modeled using deterministic discrete-generation recursion equations. We assume a one-life-stage model (adults) with a field population composed of equal numbers of male and females with the same genetic and fitness parameters, such that allelic and genotypic frequencies are equal between them. Mating is random, with unsuccessful mating events not considered. We assume the population to be sufficiently large to ignore stochastic effects. The system of equations is solved numerically using Wolfram Mathematica [Wolfram Research, Inc., Mathematica, Version 11.3, Champaign, IL (2018)].

As in Beaghton et al. (2017a,b), the effect of the IGD strategy on transmission of disease is dependent on the frequency of each genotype in the population, and the reduction in vector competence when one or more effector components is present. For the two-loci model, the reduction is denoted by $h_{rc1}r_c$ if the effector component is present at locus 2 in one copy (heterozygous) and by r_c if two copies (homozygous) are present. For the three-locus model, we assume that the reduction in vector competence

Downloaded from http://bio.biologists.org/ by guest on February 13, 2019

Biology Open (2019) 8, bio037762. doi:10.1242/bio.037762

depends on the total number of effector alleles, giving $h_{rci}\, r_c$ for i=1, 2 or 3 effector alleles in total over loci 2 and 3, and r_c for i=4 alleles in total (i.e. homozygous for the effector element in both loci). Values of r_c range from 0 (no effect) to 1 (total transmission blockage), and the dominance coefficient for refractoriness h_{rc1} (and h_{rc2} and h_{rc3} for the three-locus model) ranges from 0 (completely recessive) to 1 (completely dominant). We quantify the effect in terms of the reduction in vectorial capacity at time t as $1-V_C$[t], where V_C[t] is the vectorial capacity. V_C[t] is calculated as the sum over the genotype frequencies multiplied by their individual vector competence. For the two-locus model, this yields:

$$\begin{aligned} V_C[t] = &\sum (\text{Freq. of genotypes with no effector}) \\ &+ (1 - h_{rc1} r_c) \sum (\text{Freq. of genotypes heterozygous for effector}) \\ &+ (1 - r_c) \sum (\text{Freq. of genotypes homozygous for effector}) \end{aligned}$$

For the three-locus model:

$$\begin{aligned} V_C[t] = &\sum (\text{Freq. of genotypes with no effector}) \\ &+ (1 - h_{rc1} r_c) \sum (\text{Freq. of genotypes with one effector element}) \\ &+ (1 - h_{rc2} r_c) \sum (\text{Freq. of genotypes with two effector elements}) \\ &+ (1 - h_{rc3} r_c) \sum (\text{Freq. of genotypes with three effector elements}) \\ &+ (1 - r_c) \sum \begin{pmatrix} \text{Freq. of genotypes homozygous} \\ \text{for effector at loci 2 and 3} \end{pmatrix} \end{aligned}$$

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: A.K.B., A.H., P.A.P., G.K.C., N.W.; Methodology: A.N., G.M.U., A.K.B.; Software: A.N., G.M.U.; Validation: A.K.B., N.W.; Formal analysis: A.N., G.M.U., A.K.B.; Investigation: A.K.B., N.W.; Resources: A.N.; Writing - original draft: A.N., G.M.U., A.K.B., P.A.P., N.W.; Writing - review & editing: A.K.B., N.W.; Supervision: A.K.B., N.W.; Project administration: G.K.C., N.W.; Funding acquisition: G.K.C., N.W.

Funding

This work was supported by the Bill and Melinda Gates Foundation (OPP1158151 to N.W. and G.K.C.), the European Research Council (335724 to N.W.) and the Biotechnology and Biological Sciences Research Council (1655064 to A.N.).

Data availability

The code for all models used in this study is available at https://github.com/genome-traffic/igd.

Supplementary information

Supplementary information available online at
http://bio.biologists.org/lookup/doi/10.1242/bio.037762.supplemental

References

Adelman, Z. N. and Tu, Z. (2016). Control of mosquito-borne infectious diseases: sex and gene drive. *Trends Parasitol.* **32**, 219-229.

Akbari, O. S., Papathanos, P. A., Sandler, J. E., Kennedy, K. and Hay, B. A. (2014). Identification of germline transcriptional regulatory elements in Aedes aegypti. *Sci. Rep.* **4**, 35-39.

Anopheles gambiae 1000 Genomes Consortium, Data analysis group, Partner working group, Sample collections—Angola:, Burkina Faso:, Cameroon:, Gabon:, Guinea:, Guinea-Bissau:, Kenya et al. (2017). Genetic diversity of the African malaria vector Anopheles gambiae. *Nature* **552**, 96-100.

Basu, S., Aryan, A., Overcash, J. M., Samuel, G. H., Anderson, M. A. E., Dahlem, T. J., Myles, K. M. and Adelman, Z. N. (2015). Silencing of end-joining repair for efficient site-specific gene insertion after TALEN/CRISPR mutagenesis in *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA* **112**, 4038-4043.

Beaghton, A., Beaghton, P. J. and Burt, A. (2017a). Vector control with driving y chromosomes: modelling the evolution of resistance. *Malar. J.* **16**, 1-14.

Beaghton, A., Hammond, A., Nolan, T., Crisanti, A., Godfray, H. C. J. and Burt, A. (2017b). Requirements for driving antipathogen effector genes into populations of disease vectors by homing. *Genetics* **205**, 1587-1596.

Burt, A. (2003). Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc. R. Soc. B Biol. Sci.* **270**, 921-928.

Calvo, E., Walter, M., Adelman, Z. N., Jimenez, A., Onal, S., Marinotti, O. and James, A. A. (2005). Nanos (nos) genes of the vector mosquitoes, Anopheles gambiae, Anopheles stephensi and Aedes aegypti. *Insect Biochem. Mol. Biol.* **35**, 789-798.

Catteruccia, F., Benton, J. P. and Crisanti, A. (2005). An Anopheles transgenic sexing strain for vector control. *Nat. Biotechnol.* **23**, 1414-1417.

Champer, J., Reeves, R., Oh, S. Y., Liu, C., Liu, J., Clark, A. G. and Messer, P. W. (2017). Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLOS Genet.* **13**, e1006796.

Champer, J., Liu, J., Oh, S. Y., Reeves, R., Luthra, A., Oakes, N., Clark, A. G. and Messer, P. W. (2018). Reducing resistance allele formation in CRISPR gene drive. *Proc. Natl. Acad. Sci. USA* **115**, 5522-5527.

Chan, Y.-S., Naujoks, D. A., Huen, D. S. and Russell, S. (2011). Insect population control by homing endonuclease-based gene drive: an evaluation in Drosophila melanogaster. *Genetics* **188**, 33-44.

Chan, Y. S., Huen, D. S., Glauert, R., Whiteway, E. and Russell, S. (2013). Optimising homing endonuclease gene drive performance in a semi-refractory species: the Drosophila melanogaster experience. *PLoS ONE* **8**, e54130.

Corby-Harris, V., Drexler, A., de Jong, L. W., Antonova, Y., Pakpour, N., Ziegler, R., Ramberg, F., Lewis, E. E., Brown, J. M., Luckhart, S. et al. (2010). Activation of akt signaling reduces the prevalence and intensity of malaria parasite infection and lifespan in anopheles stephensi mosquitoes. *PLoS Pathog.* **6**, 1-10.

Cruz-Teran, C. A., Tiruthani, K., Mischler, A. and Rao, B. M. (2017). Inefficient ribosomal skipping enables simultaneous secretion and display of proteins in Saccharomyces cerevisiae. *ACS Synth. Biol.* **6**, 2096-2107.

Depinay, J.-M. O., Mbogo, C. M., Killeen, G., Knols, B., Beier, J., Carlson, J., Dushoff, J., Billingsley, P., Mwambi, H., Githure, J. et al. (2004). A simulation model of African Anopheles ecology and population dynamics for the analysis of malaria transmission. *Malar. J.* **3**, 29.

Ding, D., Chen, K., Chen, Y., Li, H. and Xie, K. (2018). Engineering introns to express RNA guides for Cas9- and Cpf1-mediated multiplex genome editing. *Mol. Plant* **11**, 542-552.

Eckhoff, P. A., Wenger, E. A., Godfray, H. C. J. and Burt, A. (2017). Impact of mosquito gene drive on malaria elimination in a computational model with explicit spatial and temporal dynamics. *Proc. Natl. Acad. Sci. USA* **114**, E255-E264.

Esvelt, K. M. and Gemmell, N. J. (2017). Conservation demands safe gene drive. *PLoS Biol.* **15**, e2003850.

Gantz, V. M. and Bier, E. (2015). The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science* **348**, 442-444.

Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E. and James, A. A. (2015). Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proc. Natl. Acad. Sci. USA* **112**, E6736-E6743.

Giraldo-Calderón, G. I., Emrich, S. J., MacCallum, R. M., Maslen, G., Emrich, S., Collins, F., Dialynas, E., Topalis, P., Ho, N., Gesing, S. et al. (2015). VectorBase: an updated Bioinformatics Resource for invertebrate vectors and other organisms related with human diseases. *Nucleic Acids Res.* **43**, D707-D713.

Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., Gribble, M., Baker, D., Marois, E., Russell, S. et al. (2016). A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nat. Biotechnol.* **34**, 78-83.

Hammond, A. M., Kyrou, K., Bruttini, M., North, A., Galizi, R., Karlsson, X., Kranjc, N., Carpi, F. M., D'Aurizio, R., Crisanti, A. et al. (2017). The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLoS Genet.* **13**, e1007039.

Isaacs, A. T., Li, F., Jasinskiene, N., Chen, X., Nirmala, X., Marinotti, O., Vinetz, J. M. and James, A. A. (2011). Engineered resistance to plasmodium falciparum development in transgenic Anopheles stephensi. *PLoS Pathog.* **7**, e1002017.

Ito, J., Ghosh, A., Moreira, L. A., Wimmer, E. A. and Jacobs-Lorena, M. (2002). Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. *Nature* **417**, 452-455.

James, S., Collins, F. H., Welkhoff, P. A., Emerson, C., Godfray, H. C. J., Gottlieb, M., Greenwood, B., Lindsay, S. W., Mbogo, C. M., Okumu, F. O. et al. (2018). Pathway to deployment of gene drive mosquitoes as a potential biocontrol tool for elimination of malaria in sub-Saharan Africa: recommendations of a scientific working group. *Am. J. Trop. Med. Hyg.* **98**, 1-49.

Jasinskiene, N., Coleman, J., Ashikyan, A., Salampessy, M., Marinotti, O. and James, A. A. (2007). Genetic control of malaria parasite transmission: threshold levels for infection in an avian model system. *Am. J. Trop. Med. Hyg.* **76**, 1072-1078.

KaramiNejadRanjbar, M., Eckermann, K. N., Ahmed, H. M. M., Sánchez, C. H. M., Dippel, S., Marshall, J. M. and Wimmer, E. A. (2018). Consequences of resistance evolution in a Cas9-based sex conversion-suppression gene drive for insect pest management. *Proc. Natl. Acad. Sci. USA* **115**, 6189-6194.

Kiani, S., Beal, J., Ebrahimkhani, M. R., Huh, J., Hall, R. N., Xie, Z., Li, Y. and Weiss, R. (2014). CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nat. Methods* **11**, 723-726.

Downloaded from http://bio.biologists.org/ by guest on February 13, 2019

Biology Open (2019) 8, bio037762. doi:10.1242/bio.037762

Kim, W., Koo, H., Richman, A. M., Seeley, D., Vizioli, J., Klocko, A. D. and O'Brochta, D. A. (2004). Ectopic expression of a cecropin transgene in the human malaria vector mosquito Anopheles gambiae (Diptera: Culicidae): effects on susceptibility to Plasmodium. *J. Med. Entomol.* **41**, 447-455.

Kyrou, K., Hammond, A. M., Galizi, R., Kranjc, N., Burt, A., Beaghton, A. K., Nolan, T. and Crisanti, A. (2018). A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. *Nat. Biotechnol.* **36**, 1062-1066.

Liu, Z., Chen, O., Wall, J. B. J., Zheng, M., Zhou, Y., Wang, L., Ruth Vaseghi, H., Qian, L. and Liu, J. (2017). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Sci. Rep.* **7**, 2193.

Marshall, J. M., Buchman, A., Sánchez, H. M. and Akbari, O. S. (2017). Overcoming evolved resistance to population-suppressing homing-based gene drives. *Sci. Rep.* **7**, 3776.

Min, J., Noble, C., Najjar, D. and Esvelt, K. (2017). Daisyfield gene drive systems harness repeated genomic elements as a generational clock to limit spread. *bioRxiv*, 104877.

Mordecai, E. A., Paaijmans, K. P., Johnson, L. R., Balzer, C., Ben-Horin, T., de Moor, E., Mcnally, A., Pawar, S., Ryan, S. J., Smith, T. C. et al. (2013). Optimal temperature for malaria transmission is dramatically lower than previously predicted. *Ecol. Lett.* **16**, 22-30.

Noble, C., Min, J., Olejarz, J., Buchthal, J., Chavez, A., Smidler, A. L., DeBenedictis, E. A., Church, G. M., Nowak, M. A. and Esvelt, K. M. (2016). Daisy-chain gene drives for the alteration of local populations. *bioRxiv*, 057307.

Noble, C., Olejarz, J., Esvelt, K. M., Church, G. M. and Nowak, M. A. (2017). Evolutionary dynamics of CRISPR gene drives. *Sci. Adv.* **3**, 3-10.

Oberhofer, G., Ivy, T. and Hay, B. A. (2018). Behavior of homing endonuclease gene drives targeting genes required for viability or female fertility with multiplexed guide RNAs. *Proc. Natl. Acad. Sci. USA* **115**, E9343-E9352.

Papathanos, P. A., Windbichler, N., Menichelli, M., Burt, A. and Crisanti, A. (2009). The vasa regulatory region mediates germline expression and maternal transmission of proteins in the malaria mosquito Anopheles gambiae: a versatile tool for genetic control strategies. *BMC Mol. Biol.* **10**, 65.

Rojas-Fernandez, A., Herhaus, L., Macartney, T., Lachaud, C., Hay, R. T. and Sapkota, G. P. (2015). Rapid generation of endogenously driven transcriptional reporters in cells through CRISPR/Cas9. *Sci. Rep.* **5**, 9811.

Simoni, A., Siniscalchi, C., Chan, Y.-S., Huen, D. S., Russell, S., Windbichler, N. and Crisanti, A. (2014). Development of synthetic selfish elements based on modular nucleases in Drosophila melanogaster. *Nucleic Acids Res.* **42**, 7461-7472.

Unckless, R. L., Clark, A. G. and Messer, P. W. (2017). Evolution of resistance against CRISPR/Cas9 gene drive. *Genetics* **205**, 827-841.

Windbichler, N., Papathanos, P. A. and Crisanti, A. (2008). Targeting the X chromosome during spermatogenesis induces Y chromosome transmission ratio distortion and early dominant embryo lethality in Anopheles gambiae. *PLoS Genet.* **4**, e1000291.

Windbichler, N., Menichelli, M., Papathanos, P. A., Thyme, S. B., Li, H., Ulge, U. Y., Hovde, B. T., Baker, D., Monnat, R. J., Burt, A. et al. (2011). A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature* **473**, 212-215.

Downloaded from http://bio.biologists.org/ by guest on February 13, 2019