Integral Gene Drives: an "operating system" 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[1,4] & Nikolai Windbichler[1,4]

1. Department of Life Sciences, Imperial College London, Sir Alexander Fleming Building, South Kensington Campus, London, United Kingdom
2. Centre of Functional Genomics, Department of Experimental Medicine, University of Perugia, Perugia, Italy
3. Department of Entomology, Robert H. Smith Faculty of Agriculture, Food and Environment, Hebrew University of Jerusalem, Rehovot, Israel
4. to whom correspondence should be addressed (g.christophides@imperial.ac.uk, n.windbichler@imperial.ac.uk)

Abstract

First generation CRISPR-based gene drives have now been tested in the laboratory in a number of organisms including malaria vector mosquitoes. A number of challenges for their use in the area-wide genetic control of vector-borne disease have been identified. These include the development of target site resistance, their long-term efficacy in the field, their molecular complexity, and the practical and legal limitations for field testing of both gene drive and coupled anti-pathogen traits. To address these challenges, we have evaluated the concept of Integral Gene Drive (IGD) as an alternative paradigm for population replacement. IGDs incorporate a minimal set of molecular components, including both the drive and the anti-pathogen effector elements directly embedded within endogenous genes - an arrangement which we refer to as gene "hijacking". This design would allow autonomous and non-autonomous IGD traits and strains to be generated, tested, optimized, regulated and imported independently. We performed quantitative modelling 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 hedging drive over multiple genomic loci prolongs the duration of transmission blockage in the face of pre-existing target-site variation. IGD thus has the potential to yield more durable and flexible population replacement traits.

Introduction

Homing gene drives were first proposed 15 years ago as potential tools for enabling the genetic engineering of natural populations (Burt 2003). After showing promise in a number of proof-of-principle studies (Chan et al. 2011; Windbichler et al. 2011; Chan et al. 2013; Simoni et al. 2014; V. M. Gantz and Bier 2015), first implementations highlighting their potential for the use to control disease vectors by population suppression and replacement were recently demonstrated in two species of malaria vector mosquitoes (A. Hammond et al. 2016a; Valentino M. Gantz et al. 2015a). Gene drive research is currently focused on two main areas: (i) studying the nature of target site resistance (A. M. Hammond et al. 2017a; Champer et al. 2017b; KaramiNejadRanjbar et al. 2018) to mitigate its eventual rise, and, conversely, (ii) 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, to identify target sites that are intolerant to drive-inactivating mutations and on the addition of further components to the drive constructs such as multiple gRNAs (Noble et al. 2017; Marshall et al. 2017; Champer et al. 2018a) or factors that bias repair towards the desired homology-directed pathway. For limiting gene invasiveness, a number of schemes have been proposed, for example linking multiple driving and non-driving CRISPR/Cas9 transgenes into a chain where the spread of each construct depends on the allele frequency of the prior link in the chain (Noble et al. 2016; Esvelt and Gemmell 2017).

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 (Figure 1A). It has

been observed that the resulting complex and large constructs can show unexpected behaviors (Valentino M. Gantz et al. 2015b; A. Hammond et al. 2016b), possibly resulting from the interaction between engineered components with each other, 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. Moreover, while a number of modelling studies have mapped out the ideal characteristics and the resulting predicted theoretical behavior of gene drives, what is often neglected are the practical implications and limitations for the construction of gene drives based on those schemes. 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 have all 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).

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. 2002a). 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 exist (Kim et al. 2004; Ito et al. 2002b; Corby-Harris et al. 2010; Isaacs et al. 2011; Jasinskiene et al. 2007), 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, and hence cannot be tested in the absence of gene drive complicating this crucial step (Figure 2A). 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. This also highlights the fact that, despite ongoing discussions (Z. Adelman et al. 2017), no clear pathway for safely testing gene drives has emerged that would allow research to progress step by step from lab to field deployment. It was not until 2017 that the first transgenic mosquito strain, carrying a dominant male-sterilizing transgene (Windbichler, Papathanos, and Crisanti 2008), was imported by the Target Malaria consortium to an African partner nation, after completing a regulatory pathway that lasted nearly 2 years. Indeed, 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 therefore inform the design of gene drives, including minimizing their molecular complexity and allowing each of their components to be assessed independently. Such an approach could accelerate the pace of gene drive research and testing as it moves closer to the field. Here we evaluate a novel strategy, Integral Gene Drive (IGD), for implementing population replacement that is summarized in Figures 1B and 2B. IGD is specifically conceived with the aforementioned molecular, population dynamic, practical, operational and regulatory challenges in mind.

IGD drive components

In contrast to the design of conventional population replacement constructs that reflects the classic transgene paradigm (Figure 1A), IGDs integrate the endonuclease coding sequence (e.g. Cas9) directly within an endogenous gene whose function and expression is confined to the male and/or the female germline where homing occurs (Figure 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 (Rojas-Fernandez et al. 2015). 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 the function of the hijacked host gene. Therefore, selection would be expected to maintain the integrity of the Cas9 open reading frame to a certain degree.

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, Benton, and Crisanti 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 Figure 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 (Nash et al, in preparation). 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. 2002c) 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 (Figure 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. 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. Other approaches for achieving hijacking than the use of the 2A signal are also conceivable but are not discussed here in detail (Supplementary Figure 1). Each IGD effector transgene also carries a 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 in trans by the drive component. 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 with the intended effect of increasing the allele frequency of the coupled effector trait.

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 (Figure 2B). 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 NHEJ or MHMR repair 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.

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 conventional fully-driving transgenes harboring a population replacement payload (Figure 2A). 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 (Figure 2B). Strains harboring the IGD drive component are also conceived to be molecularly simple, although they will 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.

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, potentially giving rise to type II errors 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 modelling 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. 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.

Exploring IGD population dynamics

To predict the behavior of one or multiple interacting IGD traits on the population level, we used a discrete-generation (non-overlapping) model comparing the dynamics of a classic replacement drive (Beaghton et al., 2017) to the dynamics of IGD, analyzing protection levels and allelic dynamics over time. To facilitate comparison to the conventional 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 & 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 Figure 3. We find that, using identical baseline values (Tables 1 & 2) to facilitate comparison (e.g., drive transmission, loss-of-function mutations during 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 approximately 4.5 years of protection against malaria transmission, while protection given by a classical gene drive lasts approximately 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 costs of expression due to the nuclease (s_n=0.05) are less than for the effector (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 (Supplementary Figure 2). 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.

We found that the existence of pre-existing resistant alleles at the effector locus among the factors that most significantly reduce the duration of protection (Supplementary Figure 2). 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 modelling 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. With pre-existing resistant alleles (10% allele frequency at all 3 loci), 95% protection lasts for 38 generations (Figure 4), 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. Figure 4 shows the boost in duration of protection for these baseline parameters (h_n=0.05, h_{e2}=h_{e3}=0.1) 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 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%.

Testing and deployment of IGDs

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. However, the modularity and interdependence of IGD components does provide a straightforward pathway for moving testing from self-limited to self-sustaining traits in the field by modulating the propensity to spread in the population (Figure 5). First, an inundative release of an effector strain alone would allow to assay (by recapture) mosquito fitness and performance under field conditions and to detect 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 consists of releasing a limited-drive strain, containing an effector component and a non-driving source of Cas9, which can trigger a limited and local 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 modelling 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 to control 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 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.

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 monolithic 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.

Modelling 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 tends 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 a human population. This approach also resolves 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. The IGD strategy allows targeting 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. 2017a) 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. It has been suggested that resistant alleles are susceptible to early stochastic loss (Beaghton, Beaghton, and Burt 2017) while we consider loss of gene function alleles to arise predominantly as a result of frameshift mutations, suggesting that some of the assumptions underpinning our deterministic model are conservative.

We have determined which parameters are key in influencing the duration of protection: the cost of expressing the nuclease (s_n), the 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 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. As such, significant penalties can be imposed upon the efficacy of the drive component without sizeable reduction in the generation 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 (Z. N. Adelman and Tu 2016), however, the modularity of the IGD strategy may be an advantage in this respect.

The potential theoretical properties of IGDs should be seen in conjunction with the practical advantages of accelerating and simplifying the generation, testing, optimization, regulation and deployment of gene drives for population replacement. The IGD strategy offers an increased degree of versatility in testing and release of components. Modelling of effector-only and limited-drive releases 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 have suitably diminished.

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 IDGs. 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. 2017a; A. M. Hammond et al. 2017b). We limited our analysis to one autonomous drive component and two non-autonomous effectors 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 2-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 behaviour while significantly complicating construct design (Oberhofer, Ivy, and Hay 2018; Champer et al. 2018b). Finally, modelling 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.

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. 2017) 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), that 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 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, 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 expressions 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 modelled using 10 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 m_i mutations) that lead to loss of function at each locus (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 and h_{n1}, h_{e2} 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 modelling 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. (2017), 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 2-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 3-loci model, we assume that the reduction in vector competence 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 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 3-loci 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 2-locus model, this yields:

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

For the 3-locus model:

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

References

Adelman, Zach, Omar Akbari, John Bauer, Ethan Bier, Cinnamon Bloss, Sarah R. Carter, Craig Callender, et al. 2017. “Rules of the Road for Insect Gene Drive Research and Testing.” *Nature Biotechnology* 35 (8): 716–18.

Adelman, Zach N., and Zhijian Tu. 2016. “Control of Mosquito-Borne Infectious Diseases: Sex and Gene Drive.” *Trends in Parasitology* 32 (3): 219–29.

Akbari, Omar S., Philippos A. Papathanos, Jeremy E. Sandler, Katie Kennedy, and Bruce A. Hay. 2014. “Identification of Germline Transcriptional Regulatory Elements in Aedes Aegypti.” *Scientific Reports* 4 (February): 3954.

Anopheles gambiae 1000 Genomes Consortium, Data analysis group, Partner working group, Sample collections—Angola:, Burkina Faso:, Cameroon:, Gabon:, et al. 2017. “Genetic Diversity of the African Malaria Vector Anopheles Gambiae.” *Nature* 552 (7683): 96–100.

Beaghton, Andrea, Pantelis John Beaghton, and Austin Burt. 2017. “Vector Control with Driving Y Chromosomes: Modelling the Evolution of Resistance.” *Malaria Journal* 16 (1): 286.

Beaghton, Andrea, Andrew Hammond, Tony Nolan, Andrea Crisanti, H. Charles J. Godfray, and Austin Burt. 2017a. “Requirements for Driving Antipathogen Effector Genes into Populations of Disease Vectors by Homing.” *Genetics* 205 (4): 1587–96.

———. 2017b. “Requirements for Driving Antipathogen Effector Genes into Populations of Disease Vectors by Homing.” *Genetics* 205 (4): 1587–96.

Burt, Austin. 2003. “Site-Specific Selfish Genes as Tools for the Control and Genetic Engineering of Natural Populations.” *Proceedings. Biological Sciences / The Royal Society* 270 (1518): 921–28.

Calvo, Eric, Marika Walter, Zachary N. Adelman, Alyssa Jimenez, Sedef Onal, Osvaldo Marinotti, and Anthony A. James. 2005. “Nanos (nos) Genes of the Vector Mosquitoes, Anopheles Gambiae, Anopheles Stephensi and Aedes Aegypti.” *Insect Biochemistry and Molecular Biology* 35 (7): 789–98.

Catteruccia, Flaminia, Jason P. Benton, and Andrea Crisanti. 2005. “An Anopheles Transgenic Sexing Strain for Vector Control.” *Nature Biotechnology* 23 (11): 1414–17.

Champer, Jackson, Jingxian Liu, Suh Yeon Oh, Riona Reeves, Anisha Luthra, Nathan Oakes, Andrew G. Clark, and Philipp W. Messer. 2018a. “Reducing Resistance Allele Formation in CRISPR Gene Drive.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (21): 5522–27.

———. 2018b. “Reducing Resistance Allele Formation in CRISPR Gene Drive.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (21): 5522–27.

Champer, Jackson, Riona Reeves, Suh Yeon Oh, Chen Liu, Jingxian Liu, Andrew G. Clark, and Philipp W. Messer. 2017a. “Novel CRISPR/Cas9 Gene Drive Constructs in Drosophila Reveal Insights into Mechanisms of Resistance Allele Formation and Drive Efficiency in Genetically Diverse Populations.” https://doi.org/10.1101/112011.

———. 2017b. “Novel CRISPR/Cas9 Gene Drive Constructs Reveal Insights into Mechanisms of Resistance Allele Formation and Drive Efficiency in Genetically Diverse Populations.” *PLoS Genetics* 13 (7): e1006796.

Chan, Yuk-Sang, David S. Huen, Ruth Glauert, Eleanor Whiteway, and Steven Russell. 2013. “Optimising Homing Endonuclease Gene Drive Performance in a Semi-Refractory Species: The Drosophila Melanogaster Experience.” *PloS One* 8 (1): e54130.

Chan, Yuk-Sang, Daniel A. Naujoks, David S. Huen, and Steven Russell. 2011. “Insect Population Control by Homing Endonuclease-Based Gene Drive: An Evaluation in Drosophila Melanogaster.” *Genetics* 188 (1): 33–44.

Corby-Harris, Vanessa, Anna Drexler, Laurel Watkins de Jong, Yevgeniya Antonova, Nazzy Pakpour, Rolf Ziegler, Frank Ramberg, et al. 2010. “Activation of Akt Signaling Reduces the Prevalence and Intensity of Malaria Parasite Infection and Lifespan in Anopheles Stephensi Mosquitoes.” *PLoS Pathogens* 6 (7): e1001003.

Depinay, Jean-Marc O., Charles M. Mbogo, Gerry Killeen, Bart Knols, John Beier, John Carlson, Jonathan Dushoff, et al. 2004. “A Simulation Model of African Anopheles Ecology and Population Dynamics for the Analysis of Malaria Transmission.” *Malaria Journal* 3 (1): 29.

Ding, Dan, Kaiyuan Chen, Yuedan Chen, Hong Li, and Kabin Xie. 2018. “Engineering Introns to Express RNA Guides for Cas9- and Cpf1-Mediated Multiplex Genome Editing.” *Molecular Plant* 11 (4): 542–52.

Eckhoff, Philip A., Edward A. Wenger, H. Charles J. Godfray, and Austin Burt. 2017. “Impact of Mosquito Gene Drive on Malaria Elimination in a Computational Model with Explicit Spatial and Temporal Dynamics.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (2): E255–64.

Esvelt, Kevin M., and Neil J. Gemmell. 2017. “Conservation Demands Safe Gene Drive.” *PLoS Biology* 15 (11): e2003850.

Gantz, Valentino M., Nijole Jasinskiene, Olga Tatarenkova, Aniko Fazekas, Vanessa M. Macias, Ethan Bier, and Anthony A. James. 2015a. “Highly Efficient Cas9-Mediated Gene Drive for Population Modification of the Malaria Vector Mosquito Anopheles Stephensi.” *Proceedings of the National Academy of Sciences of the United States of America* 112 (49): E6736–43.

———. 2015b. “Highly Efficient Cas9-Mediated Gene Drive for Population Modification of the Malaria Vector Mosquito Anopheles Stephensi.” *Proceedings of the National Academy of Sciences of the United States of America* 112 (49): E6736–43.

Gantz, V. M., and E. Bier. 2015. “The Mutagenic Chain Reaction: A Method for Converting Heterozygous to Homozygous Mutations.” *Science* 348 (6233): 442–44.

Giraldo-Calderón, Gloria I., Scott J. Emrich, Robert M. MacCallum, Gareth Maslen, Emmanuel Dialynas, Pantelis Topalis, Nicholas Ho, et al. 2015. “VectorBase: An Updated Bioinformatics Resource for Invertebrate Vectors and Other Organisms Related with Human Diseases.” *Nucleic Acids Research* 43 (Database issue): D707–13.

Hammond, Andrew, Roberto Galizi, Kyros Kyrou, Alekos Simoni, Carla Siniscalchi, Dimitris Katsanos, Matthew Gribble, et al. 2016a. “A CRISPR-Cas9 Gene Drive System Targeting Female Reproduction in the Malaria Mosquito Vector Anopheles Gambiae.” *Nature Biotechnology* 34 (1): 78–83.

———. 2016b. “A CRISPR-Cas9 Gene Drive System Targeting Female Reproduction in the Malaria Mosquito Vector Anopheles Gambiae.” *Nature Biotechnology* 34 (1): 78–83.

Hammond, Andrew M., Kyros Kyrou, Marco Bruttini, Ace North, Roberto Galizi, Xenia Karlsson, Nace Kranjc, et al. 2017a. “The Creation and Selection of Mutations Resistant to a Gene Drive over Multiple Generations in the Malaria Mosquito.” *PLoS Genetics* 13 (10): e1007039.

———. 2017b. “The Creation and Selection of Mutations Resistant to a Gene Drive over Multiple Generations in the Malaria Mosquito.” *PLoS Genetics* 13 (10): e1007039.

Isaacs, Alison T., Fengwu Li, Nijole Jasinskiene, Xiaoguang Chen, Xavier Nirmala, Osvaldo Marinotti, Joseph M. Vinetz, and Anthony A. James. 2011. “Engineered Resistance to Plasmodium Falciparum Development in Transgenic Anopheles Stephensi.” *PLoS Pathogens* 7 (4): e1002017.

Ito, Junitsu, Anil Ghosh, Luciano A. Moreira, Ernst A. Wimmer, and Marcelo Jacobs-Lorena. 2002a. “Transgenic Anopheline Mosquitoes Impaired in Transmission of a Malaria Parasite.” *Nature* 417 (6887): 452–55.

———. 2002b. “Transgenic Anopheline Mosquitoes Impaired in Transmission of a Malaria Parasite.” *Nature* 417 (6887): 452–55.

———. 2002c. “Transgenic Anopheline Mosquitoes Impaired in Transmission of a Malaria Parasite.” *Nature* 417 (6887): 452–55.

James, Stephanie, Frank H. Collins, Philip A. Welkhoff, Claudia Emerson, H. Charles J. Godfray, Michael Gottlieb, Brian Greenwood, 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.” *The American Journal of Tropical Medicine and Hygiene* 98 (6_Suppl): 1–49.

Jasinskiene, Nijole, Judy Coleman, Aurora Ashikyan, Michael Salampessy, Osvaldo Marinotti, and Anthony A. James. 2007. “Genetic Control of Malaria Parasite Transmission: Threshold Levels for Infection in an Avian Model System.” *The American Journal of Tropical Medicine and Hygiene* 76 (6): 1072–78.

KaramiNejadRanjbar, Mohammad, Kolja N. Eckermann, Hassan M. M. Ahmed, Héctor M. Sánchez C, Stefan Dippel, John M. Marshall, and Ernst A. Wimmer. 2018. “Consequences of Resistance Evolution in a Cas9-Based Sex Conversion-Suppression Gene Drive for Insect Pest Management.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (24): 6189–94.

Kiani, Samira, Jacob Beal, Mohammad R. Ebrahimkhani, Jin Huh, Richard N. Hall, Zhen Xie, Yinqing Li, and

Ron Weiss. 2014. “CRISPR Transcriptional Repression Devices and Layered Circuits in Mammalian Cells.” *Nature Methods* 11 (7): 723–26.

Kim, Won, Hyeyoung Koo, Adam M. Richman, Douglas Seeley, Jacopo Vizioli, Andrew D. Klocko, and David A. O’Brochta. 2004. “Ectopic Expression of a Cecropin Transgene in the Human Malaria Vector Mosquito Anopheles Gambiae (Diptera: Culicidae): Effects on Susceptibility to Plasmodium.” *Journal of Medical Entomology* 41 (3): 447–55.

Marshall, John M., Anna Buchman, Héctor M. Sánchez C., and Omar S. Akbari. 2017. “Overcoming Evolved Resistance to Population-Suppressing Homing-Based Gene Drives.” *Scientific Reports* 7 (1): 3776.

Min, John, Charleston Noble, Devora Najjar, and Kevin M. Esvelt. 2017. “Daisyfield Gene Drive Systems Harness Repeated Genomic Elements as a Generational Clock to Limit Spread.” https://doi.org/10.1101/104877.

Mordecai, Erin A., Krijn P. Paaijmans, Leah R. Johnson, Christian Balzer, Tal Ben-Horin, Emily de Moor, Amy McNally, et al. 2013. “Optimal Temperature for Malaria Transmission Is Dramatically Lower than Previously Predicted.” Edited by Peter Thrall. *Ecology Letters* 16 (1): 22–30.

Noble, Charleston, John Min, Jason Olejarz, Joanna Buchthal, Alejandro Chavez, Andrea L. Smidler, Erika A. DeBenedictis, George M. Church, Martin A. Nowak, and Kevin M. Esvelt. 2016. “Daisy-Chain Gene Drives for the Alteration of Local Populations.” *bioRxiv*. https://doi.org/10.1101/057307.

Noble, Charleston, Jason Olejarz, Kevin M. Esvelt, George M. Church, and Martin A. Nowak. 2017. “Evolutionary Dynamics of CRISPR Gene Drives.” *Science Advances* 3 (4): e1601964.

Oberhofer, Georg, Tobin Ivy, and Bruce A. Hay. 2018. “Behavior of Homing Endonuclease Gene Drives Targeting Genes Required for Viability or Female Fertility with Multiplexed Guide RNAs.” https://doi.org/10.1101/289546.

Papathanos, Philippos A., Nikolai Windbichler, Miriam Menichelli, Austin Burt, and Andrea Crisanti. 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 Molecular Biology* 10 (1): 65.

Rojas-Fernandez, Alejandro, Lina Herhaus, Thomas Macartney, Christophe Lachaud, Ronald T. Hay, and Gopal P. Sapkota. 2015. “Rapid Generation of Endogenously Driven Transcriptional Reporters in Cells through CRISPR/Cas9.” *Scientific Reports* 5 (April): 9811.

Simoni, Alekos, Carla Siniscalchi, Yuk-Sang Chan, David S. Huen, Steven Russell, Nikolai Windbichler, and Andrea Crisanti. 2014. “Development of Synthetic Selfish Elements Based on Modular Nucleases in Drosophila Melanogaster.” *Nucleic Acids Research* 42 (11): 7461–72.

Windbichler, Nikolai, Miriam Menichelli, Philippos Aris Papathanos, Summer B. Thyme, Hui Li, Umut Y. Ulge, Blake T. Hovde, et al. 2011. “A Synthetic Homing Endonuclease-Based Gene Drive System in the Human Malaria Mosquito.” *Nature* 473 (7346): 212–15.

Windbichler, Nikolai, Philippos Aris Papathanos, and Andrea Crisanti. 2008. “Targeting the X Chromosome during Spermatogenesis Induces Y Chromosome Transmission Ratio Distortion and Early Dominant Embryo Lethality in Anopheles Gambiae.” *PLoS Genetics* 4 (12): e1000291.

Acknowledgements

This study was funded by the Bill & Melinda Gates Foundation grant OPP1158151 awarded to NW & GC. Funded by the European Research Council under the European Union's Seventh Framework Programme ERC grant no. 335724 awarded to N.W. A.N. is supported by the Biotechnology and Biological Sciences Research (BBSRC) Council-funded Doctoral Training Partnership award 1655064.

Figure legends

Figure 1. Schematic overview of the molecular constructs enabling both conventional (A) and integral (B) gene drive approaches for population replacement. The black triangle indicates the 2A ribosomal skipping signal and the gRNA locus is indicated as a circle.

Figure 2. A comparison of the steps and requirements for lab and field testing of both conventional (A) and integral (B) gene drive strategies for population replacement of malaria vectors.

Figure 3. 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 modelled behavior of the constituent components at the two loci. The dashed red line shows the proportionate reduction in vectorial capacity of the target population.

Figure 4. Dynamics of the 3-locus 2-effector model assuming the release of transgenics for all loci (locus 1: drive, locus 2 & 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 modelled behavior of the constituent components at the 3 loci. The dashed red line shows the proportionate reduction in vectorial capacity of the target population.

Figure 5. Evaluation of the allele frequency dynamics of limited release strategies. Release of strains carrying an IGD effector component only (upper panel) or of a limited-drive strain combining a non-driving Cas9 locus with an effector (middle panel) 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 panel).

Table 1. Parameters and Baseline Values for the IGD Model. 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. 2017) 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).

Table 2. List of Parameters and Baseline Values for 'Classical' Model. An exemplar set of parameter values, consistent with published work conducted in relation to classical replacement gene drives (Beaghton et al. 2017c).

Supplementary Figure 1. Schematic showing the expression of the endogenous and introduced genetic elements of an effector component allele at a hijacked gene locus. Expression of the gRNA occurs in the germline (yellow) whereas expression of the hijacked gene occurs in an infection-relevant somatic tissue (green). While expressing a straight fusion of endogenous protein (green) and the effector protein (grey) is possible the generation of separate polypeptides by using either (A) the 2A ribosome-skipping signal (black triangle) or (B) the use of intein sequences (black semicircle) are possible avenues to avoid interfering with the function of the hijacked gene.

Supplementary Figure 2. Duration of 67% and 95% protection (top and bottom lines in each graph) as each of the underlying parameters is varied, holding all others at their baseline values.

Supplementary Figure 3. Overview of allele frequency dynamics when releasing both classical drive (A) and IGD strains into field populations with 10% pre-existing resistance. For IGD, resistance is modelled at the drive component locus (B), the effector component locus (C), and at both loci (D).

Figure 1.

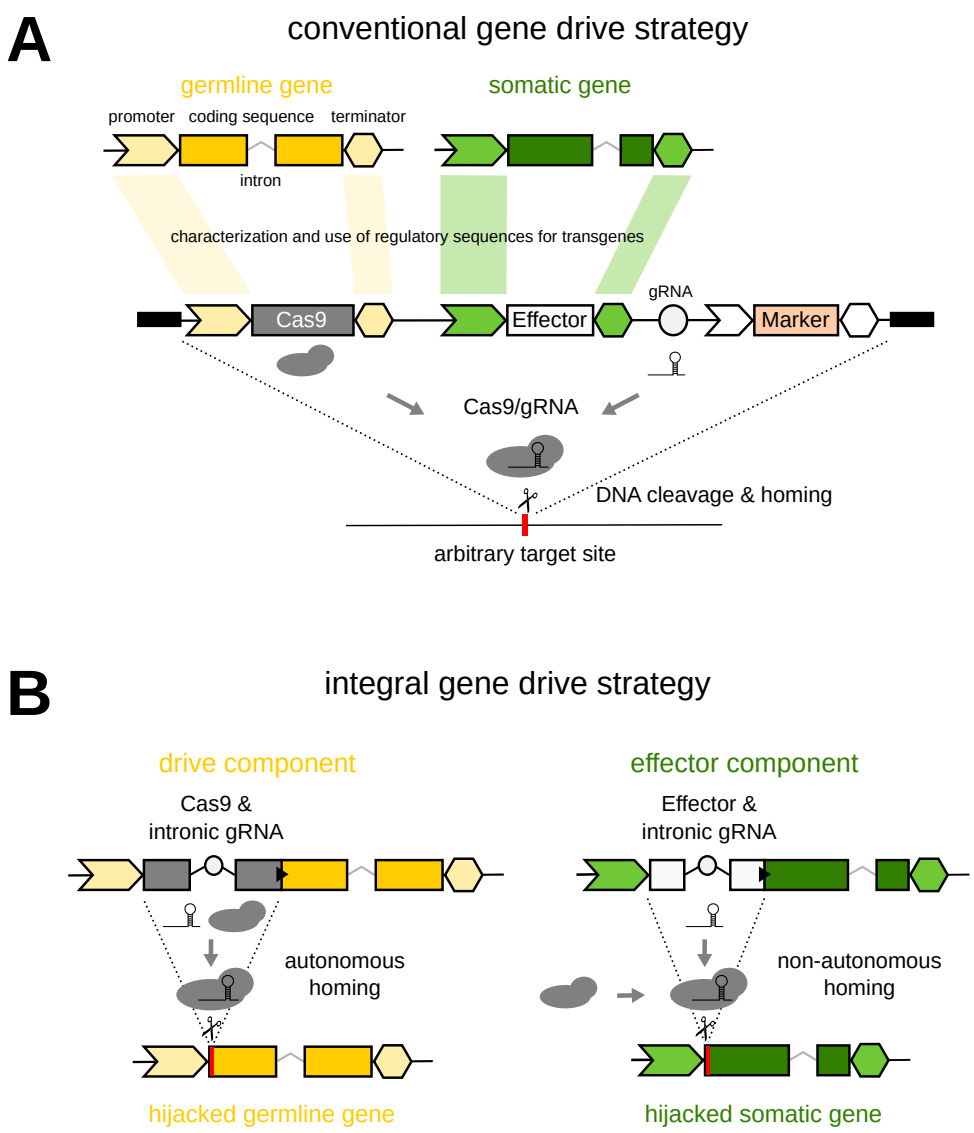

A
conventional gene drive strategy
germline gene
somatic gene
promoter
coding sequence
terminator
intron
characterization and use of regulatory sequences for transgenes
gRNA
Cas9
Effector
Marker
Cas9/gRNA
DNA cleavage & homing
arbitrary target site
B
integral gene drive strategy
drive component
effector component
Cas9 &
intronic gRNA
Effector &
intronic gRNA
autonomous
homing
non-autonomous
homing
hijacked germline gene
hijacked somatic gene

Figure 2.

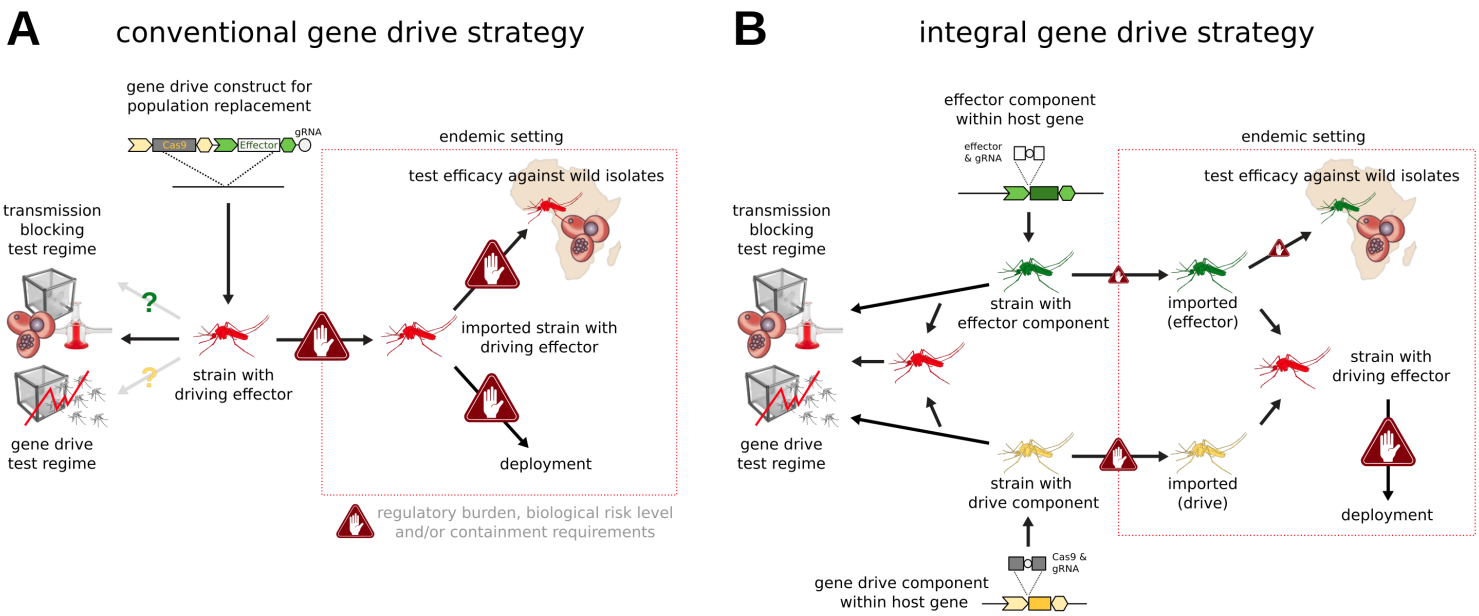

A
conventional gene drive strategy
gene drive construct for population replacement
Cas9
Effector
gRNA
endemic setting
test efficacy against wild isolates
transmission blocking test regime
?
?
strain with driving effector
imported strain with driving effector
gene drive test regime
deployment
regulatory burden, biological risk level and/or containment requirements
B
integral gene drive strategy
effector component within host gene
effector & gRNA
endemic setting
test efficacy against wild isolates
transmission blocking test regime
strain with effector component
imported (effector)
strain with driving effector
gene drive test regime
strain with drive component
imported (drive)
deployment
Cas9 & gRNA
gene drive component within host gene

Figure 3.

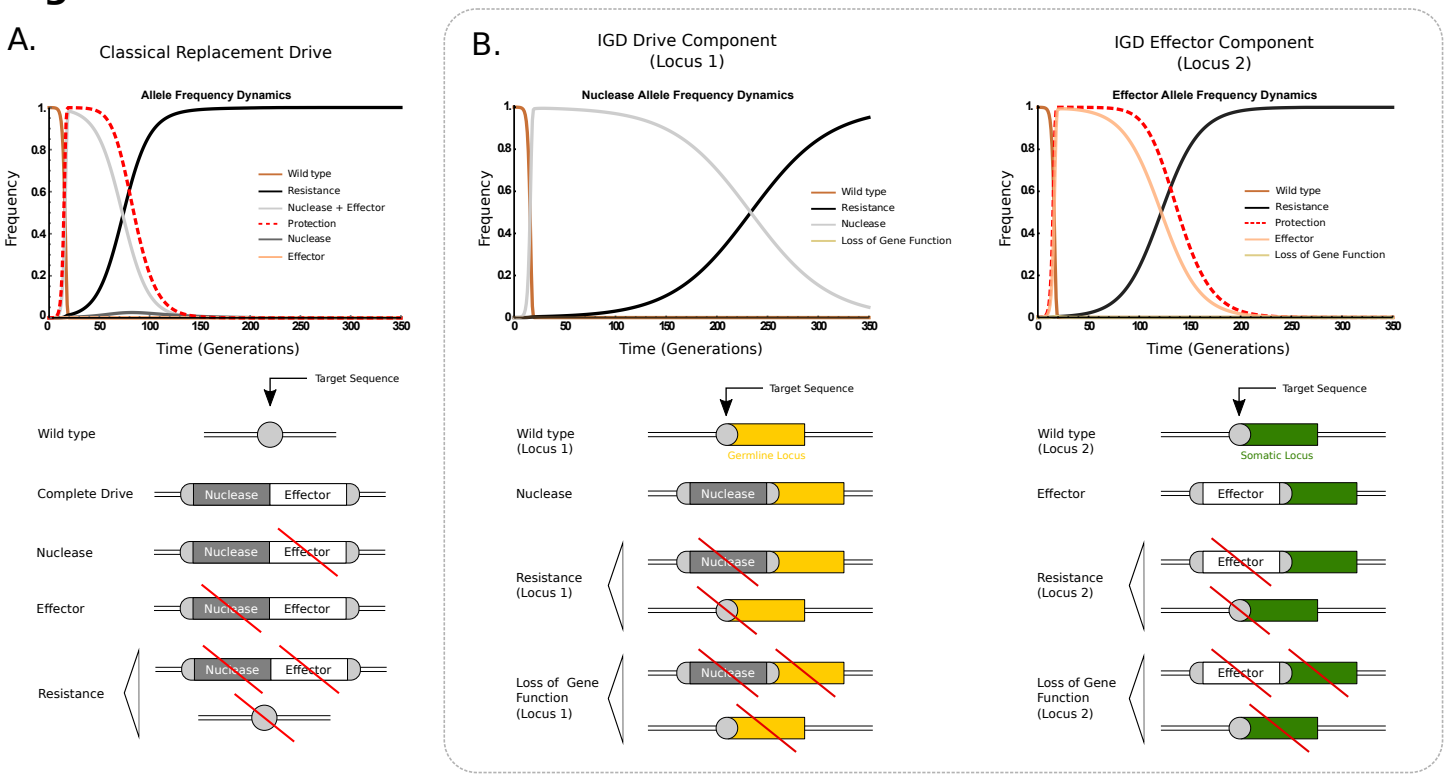

A.
Classical Replacement Drive
Allele Frequency Dynamics
Frequency
Time (Generations)
Wild type
Resistance
Nuclease + Effector
Protection
Nuclease
Effector
Target Sequence
Wild type
Complete Drive
Nuclease
Effector
Nuclease
Effector
Resistance
B.
IGD Drive Component
(Locus 1)
Nuclease Allele Frequency Dynamics
Frequency
Time (Generations)
Wild type
Resistance
Nuclease
Loss of Gene Function
Target Sequence
Wild type
(Locus 1)
Germline Locus
Nuclease
Resistance
(Locus 1)
Loss of Gene
Function
(Locus 1)
IGD Effector Component
(Locus 2)
Effector Allele Frequency Dynamics
Frequency
Time (Generations)
Wild type
Resistance
Protection
Effector
Loss of Gene Function
Target Sequence
Wild type
(Locus 2)
Somatic Locus
Effector
Resistance
(Locus 2)
Loss of Gene
Function
(Locus 2)

Figure 4.

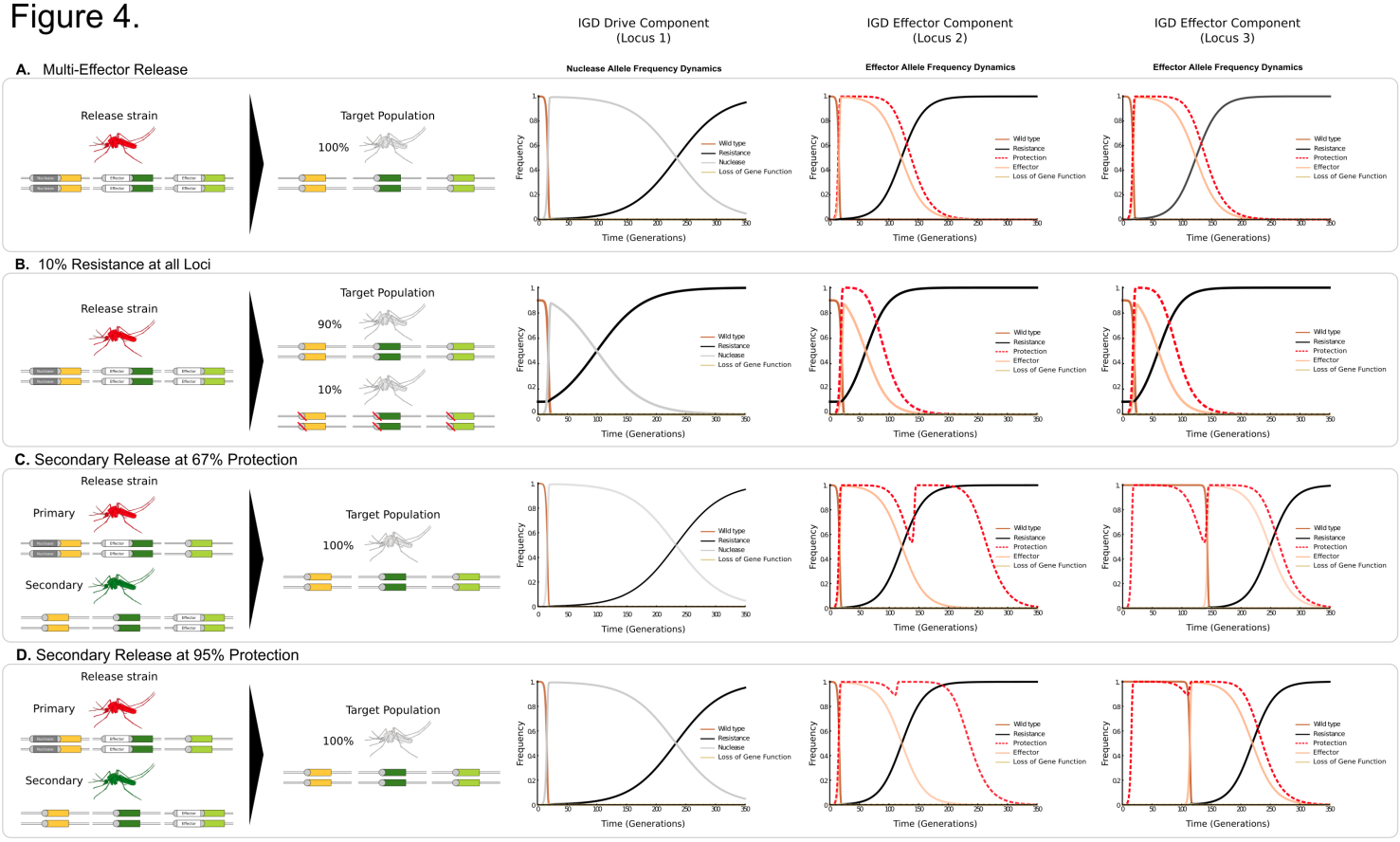
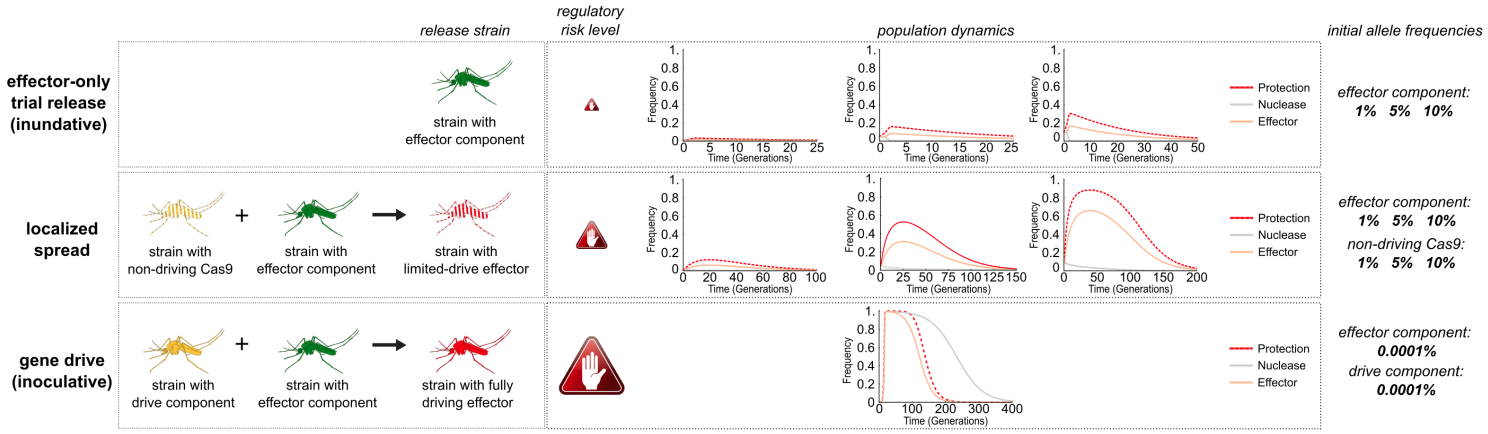

IGD Drive Component
(Locus 1)
IGD Effector Component
(Locus 2)
IGD Effector Component
(Locus 3)
A. Multi-Effector Release
Nuclease Allele Frequency Dynamics
Effector Allele Frequency Dynamics
Effector Allele Frequency Dynamics
Release strain
Target Population
100%
B. 10% Resistance at all Loci
Release strain
Target Population
90%
10%
C. Secondary Release at 67% Protection
Release strain
Primary
Secondary
Target Population
100%
D. Secondary Release at 95% Protection
Release strain
Primary
Secondary
Target Population
100%
Frequency
Time (Generations)
Wild type
Resistance
Nuclease
Protection
Effector
Loss of Gene Function

Figure 5.


release strain
regulatory risk level
population dynamics
initial allele frequencies
effector-only trial release (inundative)
strain with effector component
localized spread
strain with non-driving Cas9
strain with effector component
strain with limited-drive effector
gene drive (inoculative)
strain with drive component
strain with effector component
strain with fully driving effector
Frequency
Time (Generations)
Protection
Nuclease
Effector
effector component:
1% 5% 10%
non-driving Cas9:
1% 5% 10%
0.0001%
drive component:
0.0001%

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
rr_0	Initial frequency of resistance	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 (I)	0.99
d_{3b}	Transmission rate of drive (locus 3) if drive component present in homozygosity (I)	0.99
l_1	Resistance arising from HR at locus 1	10^{-4} x (1/3)
l_2	Resistance arising from HR at locus 2	10^{-4} x (1/3)
l_3	Resistance arising from HR at locus 3 (I)	10^{-4} x (1/3)
u_1	Resistance arising from NHEJ at locus 1	0.5 x (1/3)
u_2	Resistance arising from NHEJ at locus 2	0.5 x (1/3)
u_3	Resistance arising from NHEJ at locus 3 (I)	0.5 x (1/3)
L_1	Loss of gene function arising from HR at locus 1	10^{-4} x (2/3)
L_2	Loss of gene function arising from HR at locus 2	10^{-4} x (2/3)
L_3	Loss of gene function arising from HR at locus 3 (I)	10^{-4} x (2/3)
U_1	Loss of gene function arising from NHEJ at locus 1	0.5 x (2/3)
U_2	Loss of gene function arising from NHEJ at locus 2	0.5 x (2/3)
U_3	Loss of gene function arising from NHEJ at locus 3 (I)	0.5 x (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) (I)	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 (I)	0.1
h_m	Dominance coef. for loss of gene function	0.2
h_{d1}	Dominance coef. for hijacking at locus 1	0.5
h_{d2}	Dominance coef. for hijacking at locus 2	0.5
h_{d3}	Dominance coef. for hijacking at locus 3 (I)	0.5
h_n	Dominance coef. for expressing nuclease	0.5
h_{e2}	Dominance coef. for expressing effector (locus 2)	0.5
h_{e3}	Dominance coef. for expressing effector (locus 3) (I)	0.5
h_{rc1}	Dominance coef. for refractoriness (1 effector allele)	1
h_{rc2}	Dominance coef. for refractoriness (2 effector alleles) (I)	1
h_{rc3}	Dominance coef. for refractoriness (3 effector alleles) (I)	1
r_c	Homozygous degree of refractoriness	1

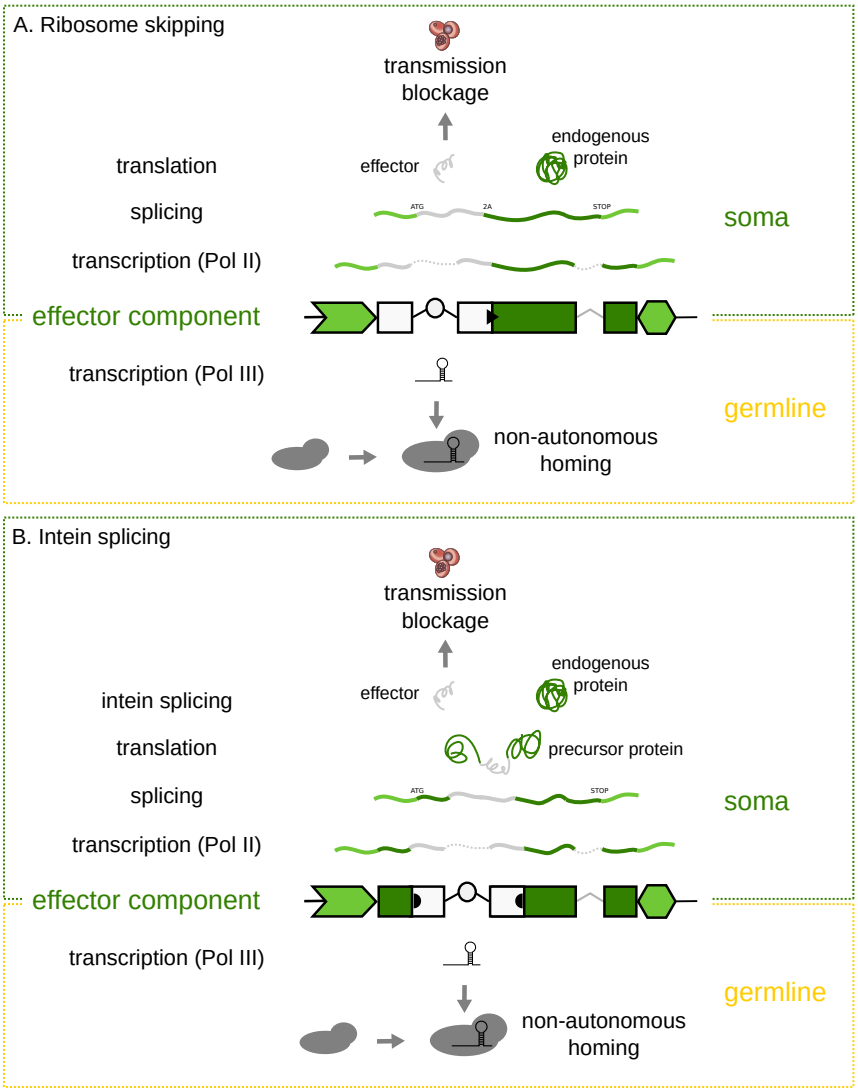
[+] *I – Three locus model only*

Table 1.

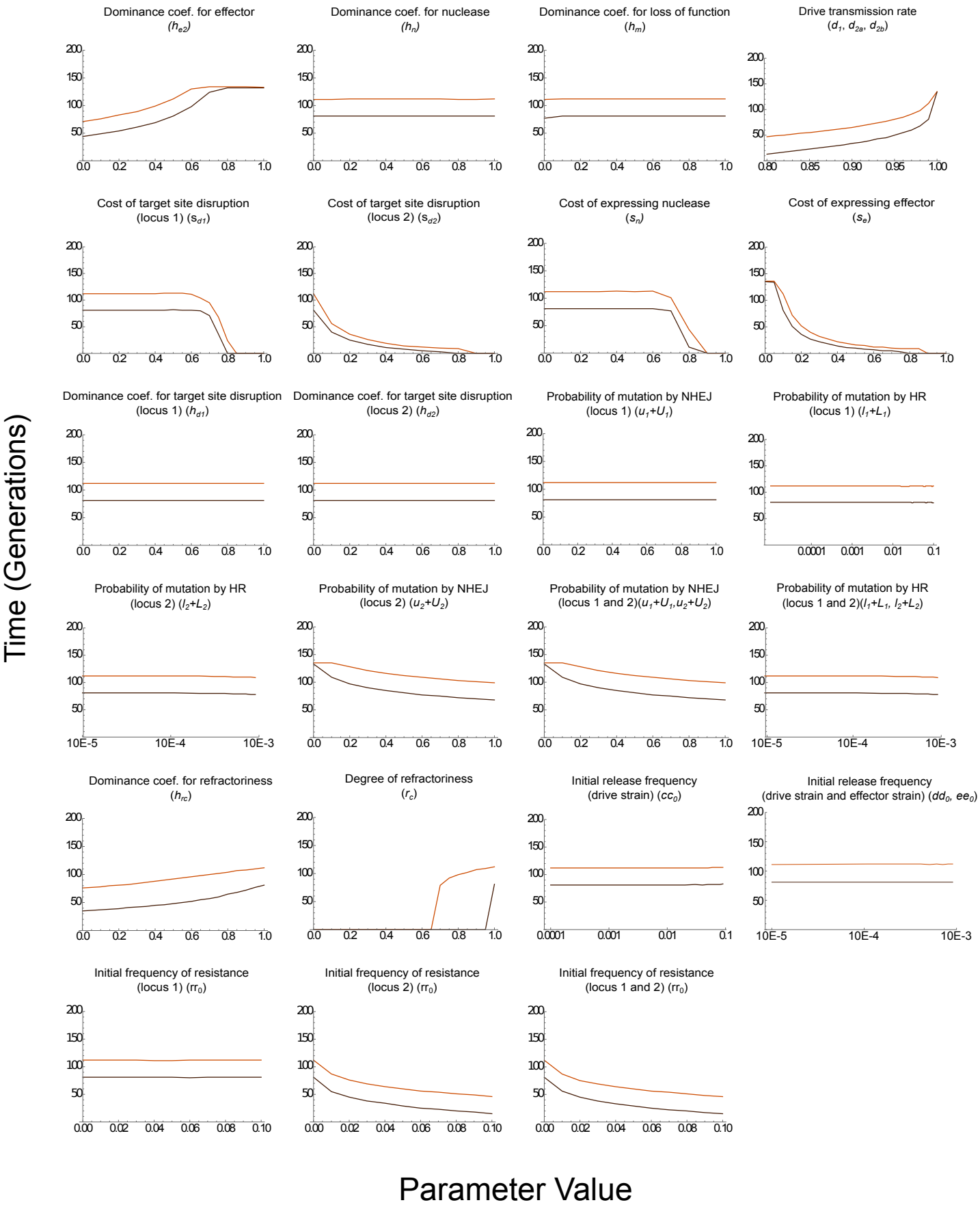
Symbol	Parameter	Baseline Value
cc_0	Initial release frequency of homozygous construct	10^{-4}
d	Transmission rate of drive	0.99
u	Probability of resistance (functional)	0.5
l_n	Probability nuclease gene lost during homing	10^{-4}
l_e	Probability effector lost during homing	10^{-4}
l_{ne}	Probability nuclease & effector lost 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 coef. for target site disruption	0.5
h_n	Dominance coef. for nuclease expression	0.5
h_e	Dominance coef. for effector expression	0.5
h_{rc}	Dominance coefficient for refractoriness	1
r_c	Homozygous degree of refractoriness	1

Table 2.

Supplementary Figure 1.

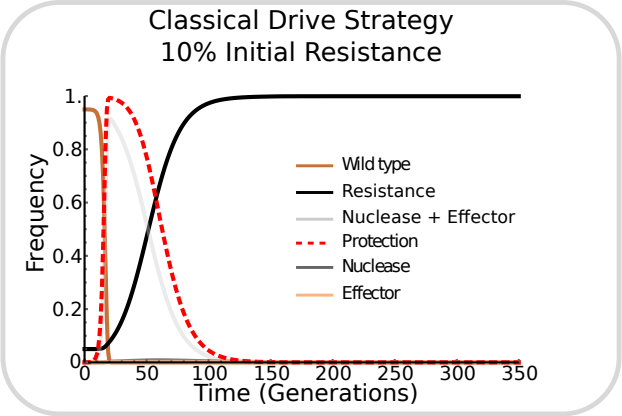


Supplementary Figure 2.



Supplementary Figure 3.

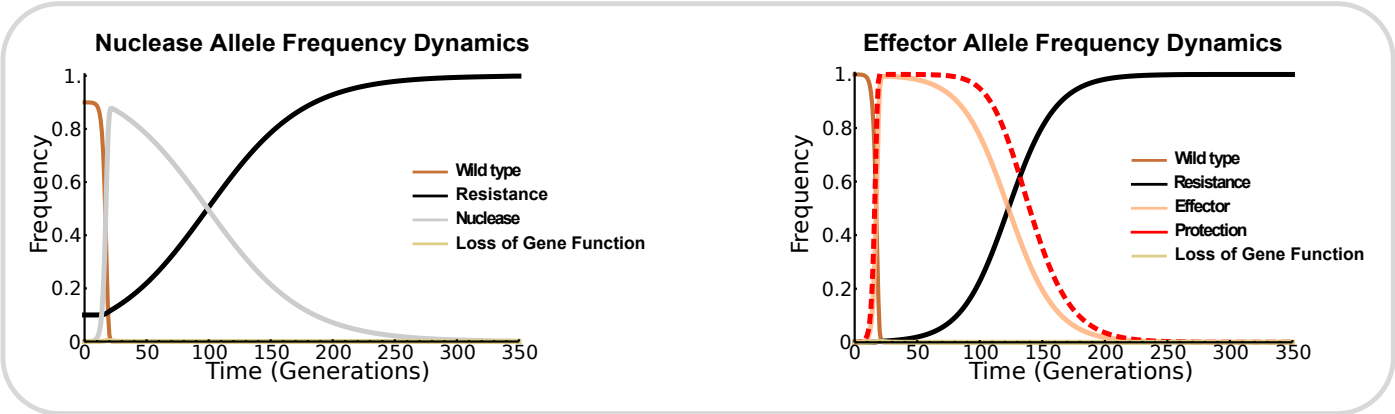
A. Classical Drive - 10% Pre-existing Resistance



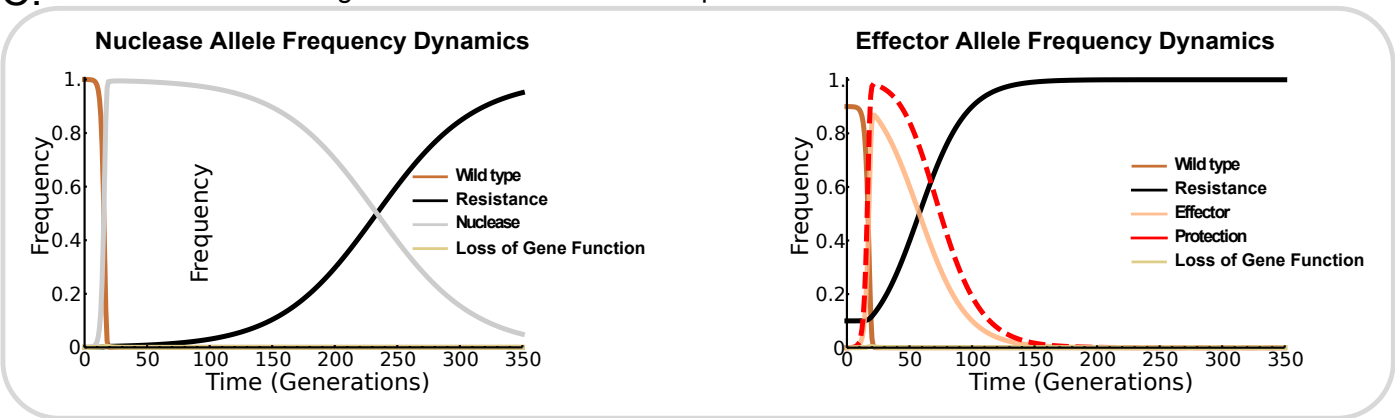
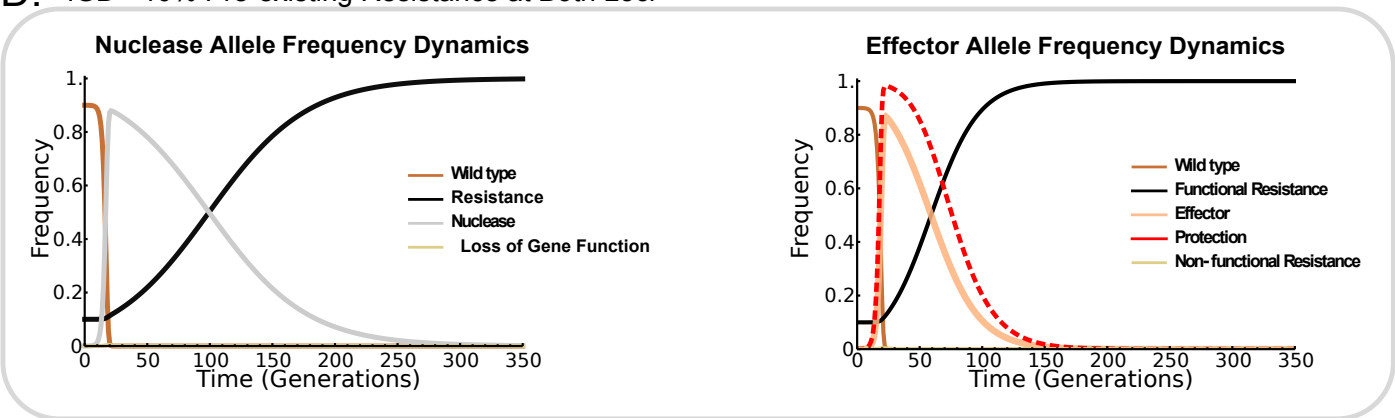
IGD Drive Component (Locus 1)

IGD Effector Component (Locus 2)

B. IGD - 10% Pre-existing Resistance at Drive Component Locus



C. IGD - 10% Pre-existing Resistance at Effector Component Locus



D. IGD - 10% Pre-existing Resistance at Both Loci