The design and mechanism of synthetic homing endonuclease gene drives

Sebald Verkuijl

Linacre College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2023

0.1 Declaration

I declare that this thesis was composed by myself and that the work contained herein is my own, except where explicitly stated in the text. This work has not been submitted for any other degree or professional qualification.

Sebald Verkuijl

0.2 Abstract

When humans and insects come into conflict, this can lead to significant ecological and public health challenges. Traditional pest control methods often rely on chemical insecticides, which can adversely affect the environment and non-target species. Genetic biocontrol through the use of gene drive may offer a radically more efficient and species-specific method of addressing pest harm. Synthetic homing gene drives, particularly those based on the CRISPR-Cas9 system, have been the focus of intensive research in recent years. However, there is much we do not understand about the fundamental nuclease processes that mediate gene drive functioning and how these are affected by transgene design. Here, we investigate through computational modelling how unintended nuclease processes affect gene drive performance. We find that certain complex self-limited gene drive designs (daisy-chain gene drives) are especially sensitive to effects that cause separate gene drive elements to segregate prematurely. Through a meta-analysis of experimental gene drive crosses, we evaluate the effect of sex and deposition on different gene drive outcomes. We developed a companion web tool that allows users to find the highest-quality data evaluating specific aspects of gene drive designs and experimental conditions. Lastly, we analyse *Aedes aegypti* gene drive performance with different nuclease expression patterns. This shows that identical homing gene drives can vary in their underlying mechanism of inheritance bias.

0.3 Statement of Authorship

This thesis is submitted as an integrated thesis, with chapters 2-5 structured as collaborative free-standing manuscripts. Chapters 2, 3, and 5 have been published in article format, and in each case, I drafted the manuscript and was the first author or co-first in the case of Chapter 5. Each chapter includes a statement listing the individual author's contributions.

Chapter 2: Sebald A. N. Verkuijl, Joshua X. D. Ang, Luke Alphey, Michael B. Bonsall, and Michelle A. E. Anderson. The Challenges in Developing Efficient and Robust Synthetic Homing Endonuclease Gene Drives. *Frontiers in Bioengineering and Biotechnology*, 10, 3 2022. ISSN 2296-4185. doi: 10.3389/fbioe.2022.856981. URL `https://www.frontiersin.org/article/10.3389/fbioe.2022.856981`

Chapter 3: Sebald A N Verkuijl, Michelle A E Anderson, Luke Alphey, and Michael B Bonsall. Daisy-chain gene drives: The role of low cut-rate, resistance mutations, and maternal deposition. *PLOS Genetics*, 18(9):e1010370, 9 2022. doi: 10.1371/journal.pgen.1010370. URL `https://doi.org/10.1371/journal.pgen.1010370`

Chapter 5: Sebald A N Verkuijl, Estela Gonzalez, Ming Li, Joshua X. D. Ang, Nikolay P Kandul, Michelle A E Anderson, Omar S Akbari, Michael B Bonsall, and Luke Alphey. A CRISPR endonuclease gene drive reveals distinct mechanisms of inheritance bias. *Nature Communications*, 13(1):7145, 11 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-34739-y. URL `https://www.nature.com/articles/s41467-022-34739-y`

Chapter 4 is written in an article format but is not currently published. Chapters 1 and 6 were written by me for this thesis in a non-article format.

During my studentship, I contributed to the following additional publications outside of the scope of this thesis:

Joshua Xin De Ang, Katherine Nevard, Rebekah Ireland, Deepak-Kumar Purusothaman, Sebald A N Verkuijl, Lewis Shackleford, Estela Gonzalez, Michelle A E Anderson, and Luke Alphey. Considerations for homology-based DNA repair in mosquitoes: Impact of sequence heterology and donor template source. *PLOS Genetics*, 18(2):e1010060, 2 2022. ISSN 1553-7404. doi: 10.1371/journal.pgen.1010060. URL `https://doi.org/10.1371/journal.pgen.1010060`

Michelle A E Anderson, Estela Gonzalez, Joshua X D Ang, Lewis Shackleford, Katherine Nevard, Sebald A N Verkuijl, Matthew P Edgington, Tim Harvey-Samuel, and Luke Alphey. Closing the gap to effective gene drive in Aedes aegypti by exploiting germline regulatory elements. *Nature Communications*, 14(1):338, 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-36029-7. URL `https://doi.org/10.1038/s41467-023-36029-7`

0.4 Acknowledgements

My thanks to my supervisor Michael Bonsall for consistent support, patience, mentorship, and the freedom to pursue my specific scientific interests. My thanks to Luke Alphey and Michelle Anderson for their scientific input and for maintaining our collaboration through a pandemic.

My thanks to Dessislava, with whom I was able to share the daily struggles. My thanks to Joshua and Timothy for our longstanding scientific and non-scientific association. My thanks to Estella, Phillip, and Edward for our scientific collaboration.

Many thanks to Jeantine Lunshof for answering a cold email that started me on this gene drive journey.

My thanks to Stefan and Reza for our forays into the world of startups and Andrew for my time in Italy.

My thanks to my parents, brother, and sisters for their understanding and patience as I worked through this sometimes all-consuming undertaking. My thanks to Cleo for her support and our semiregular runs (and incidentally for naming Avon her successor in this). Most importantly, my thanks to Marisa, whose unwavering love and support saw me through this time. Our respective PhDs have spaned our entire relationship, and I look forward to what our future holds.

Lastly, many thanks to the funders and people who directly and indirectly made this thesis possible and who I failed to list above.

Contents

Chapter 1

Introduction

Around 80% of all known living animal species are arthropods, which includes insects[6]. They are generally small, have short generation times, and have large reproductive capacities. Unfortunately, the flourishing of certain species of insects and humans can be in conflict, which can present challenges in agriculture, infrastructure, public health, and conservation. More than half of the annual production of many crops is lost to arthropod pests before and after harvest, resulting in losses of almost half a trillion US-dollars[7–9]. Invasive species such as the fire ant (*Solenopsis invicta*) can be a major source of infrastructure damage by building nests under pavements and building foundations[10]. In public health, mosquitoes are the most harmful insects, with the *Anopheles* species complex and *Aedes aegypti* mosquito being responsible for over half a million deaths each year[11;12]. In addition to directly affecting humans, insect pests have contributed to the decline of many native species. For example, human introduction of invasive mosquito species that vector avian malaria has severely affected native bird populations in Hawaii[13].

Chemical pesticides are currently the most important tool for preventing and mitigating the harms of insect pests. However, pesticides are increasingly being regulated in terms of which chemicals can be used and under what circumstances[14]. These restrictions are in response to pesticide resistance, harmful effects on human health, and off-target effects on non-pest species[15]. Alternative approaches with higher efficacy and a more favourable safety profile are severely needed.

Biological control is an alternative to the use of chemical pesticides that aims to introduce predators, parasites, pathogens, or other biological agents into an ecosystem to suppress the population of a pest. One class of strategies aims to suppress wild populations by releasing sterile individuals of the same species as the pest (genetic

biocontrol[16]). If enough sterile individuals are released, they can compete with fertile wild individuals and reduce the number of mating events that lead to viable offspring. Genetic biocontrol has been shown to be an effective, fully biodegradable, and species-specific way to suppress wild pest populations[17;18].

Sterility is most commonly achieved by exposing laboratory-bred pests to irradiation, referred to as the sterile insect technique (SIT). The irradiation dose should be high enough to result in reproductive sterility but not too high that it causes other fitness defects[19]. However, a loss of mating competitiveness is often unavoidable. Releasing exclusively males, as opposed to a mixture of males and females, has been shown to increase the effectiveness of SIT several fold[20;21]. SIT has had high-impact successes, such as the eradication of the screwworm fly (*Cochliomyia hominivorax*) from North America[18]. However, it has not become a universally adopted pest control tool due to technical challenges in making the technique cost-effective at scale[22]. These challenges include fitness deficits caused by irradiation and mass-rearing methods, the labour required for sex sorting, and the cost of the biomass released. Potentially as important is that, in almost all cases, the approach is radically different from existing approaches, making adoption difficult and necessitating extensive stakeholder education and engagement to be able to present the merits of the approach.

Genetic biocontrol has significant potential benefits, and many researchers are actively working to address the technical challenges associated with this approach. Mechanisation is one of those strategies. For example, robot-assisted automatic rearing systems have been developed to aid in the rearing of *A. aegypti*, which is augmented by machine vision approaches to remove the need for manual sex sorting[23]. Complex parametric analyser and sorter (COPAS) flow cytometry machines are expensive but relatively standard laboratory equipment that can also eliminate the need for manual sex sorting[24]. However, such approaches require sex-linked fluorescent markers to be engineered into the genetic-biocontrol strain. Post-rearing, unmanned aerial vehicles (drones) are being used by the World Mosquito Programme to aid in mosquito dispersal[25;26]. These approaches can make currently marginally cost-effective genetic biocontrol applications viable and scalable, representing a substantial impact (particularly with *A. aegypti*). However, it is unlikely that these approaches will radically expand the applicability of genetic biocontrol to most pest species.

Several strain modification approaches have been developed to improve the efficacy and feasibility of genetic biocontrol. These modifications substantially alter how genetic biocontrol insects can be reared and their mechanism of action. As mentioned

above, sex-specific fluorescent markers can make highly efficient mechanical sex sorting approaches possible[27] and, in some cases, can eliminate the need for sex sorting altogether through sex-specific lethality or flightlessness[24].

Approaches have been developed to complement or replace irradiation as the effector of sterility. One such approach is the use of cytoplasmic incompatibility through *Wolbachia*-infection, which causes embryonic death in the offspring of infected males and uninfected females[28]. Another approach is drug-repressible lethality, where insects are engineered to depend on a specific supplement provided in the rearing facility[29]. When released, males that received the supplement can mate with wild females, but their offspring do not mature. This strategy can be further enhanced by making the transgene lethal only to females[30]. This sex-specific lethality provides a multigenerational effect in which males who inherit the transgene can pass it on and kill their daughters, increasing the impact per released male. The above strain modification approaches may be more practical and leave the released males fitter compared to irradiation[31;32]; however, in some cases, the actual experimental evidence for this is lacking.

The impact per individual released can be further increased by systems such as autosomal X-Shredders and Y-linked editors that carry molecular machinery to generate DNA damage in subsequent generations[33;34]. Despite the considerable number of proof-of-concept genetic biocontrol strategies published[35;36], actual field trials and real-world use of genetically modified insects have so far been limited[30;37–40]. The cost-effectiveness of transgenic methods[41] may, in many cases, still not be sufficient for the market or face additional challenges that are hard to overcome (e.g., higher regulatory requirements[42] and public scepticism).

A potentially radically more efficient implementation of genetic biocontrol lies in the use of gene drive technologies. Gene drives are genetic elements that increase their frequency in populations without a proportional benefit to fitness. Compared to the strategies listed above, gene drive organisms can have a much greater impact per released individual, making rearing challenges far less meaningful. Although they may be more powerful, regulatory, ethical and public concerns are substantially greater for any genetic engineering implementation that includes gene drive technology[43–45].

Gene drives can achieve inheritance bias through many different mechanisms, and synthetic homing endonuclease gene drives have seen the most rapid adoption. Homing drives are based on molecular tools and processes for site-specific genetic modification

that have been intensively studied and developed for decades[46]. Site-specific genetic modification generally relies on the generation of targeted DNA damage and the provision of an exogenous repair template in the form of plasmid DNA. Then, endogenous and well-conserved homology-directed repair processes will, with some frequency, repair the site of the DNA break to match the supplied repair template. Homing drives function through the same process but are designed so that the homologous chromosome that contains the gene drive is used as the repair template. Any full copy of the gene drive also carries the necessary nuclease genes to induce DNA damage, allowing the process to repeat itself in subsequent generations (self-perpetuating).

A gene drive can spread any of a number of effector modifications that result in the, for humans, useful change in the target pest species. A commonly pursued approach is the spread of a modification (usually the drive insertion itself) that results in female-specific recessive sterility or death[47–51]. If engineered correctly, the trait only affects females that inherit the effector modification from both parents. This allows the gene drive to spread to a high frequency and maximises the genetic load imposed on the target population. Other effector modifications, such as making vector insects refractory to certain diseases, are also being pursued[52;53].

The non-profit Target Malaria is the most prominent initiative that is evaluating the potential to release a self-perpetuating gene drive[54]. However, for all but a few cases, such as malaria mosquitoes, affecting populations beyond the release site is not necessary or desirable[55]. Insect pests are only harmful in places where they come into contact with humans. Specific drive implementations have been proposed as a middle ground, allowing for increased impact per individual released but with a limited ability to replicate (self-limiting drives) or confinement through threshold-dependent effects (confined drives). Many such drive systems with different attributes have been proposed and computationally modelled[56–59], but only a few different designs have been implemented in proof-of-principle studies[60–62]. Most research on synthetic gene drives has focused on overcoming the basic technical challenges of achieving efficient and robust inheritance bias rates[1].

A class of self-limiting drive systems that have received particular attention in scientific and public discourse are daisy-chain gene drives[59]. Daisy-chain gene drives function by linking multiple independent gene drive elements in an ordered open "chain". For example, gene drive element A has its inheritance biased by element B, and element B, in turn, is biased by element C. This results in copies of the gene drive system being imperfect and progressively failing to amplify the bottom-most element. An effector

modification housed on element A will experience multiple generations of exponential amplification but then replicate at a much slower rate.

Daisy-chain gene drives are a compelling self-limiting approach because they simply extend the basic inheritance biasing mechanism used in a self-perpetuating gene drive. As such, it may be expected that if we can engineer an efficient self-perpetuating gene drive, we can do the same for a daisy-chain. However, inefficiencies in the inheritance biasing process also increase in importance, and separate elements create emergent design constraints. As such, daisy-chain gene drives provide an interesting test case for probing the interaction of how homing gene drives actually function in practice and how we should accommodate these limitations in their design.

This thesis will focus on understanding how the fundamental nuclease process of homing gene drives can lead to unintended outcomes. Through computational modelling, we will evaluate how unintended outcomes interact with gene drive design. We will then investigate how to systematically use experimental data to inform computational modelling and how sex can influence gene drive performance. We end with an analysis of new experimental data that shows that the fundamental mechanism of homing gene drive inheritance bias is not fixed.

1.1 Thesis outline

Chapter 2 sets the foundation for this thesis. We introduce the history of research on synthetic-homing endonuclease gene drives and highlight the challenges facing their development. We highlight experimental results consisting of synthetic self-perpetuating and split-drive designs. We focus on unintended outcomes and how they interact with certain characteristics of transgene design. We also highlight strategies that have been implemented to mitigate unintended repair outcomes, including targeting genes with a high fitness cost.

In chapter 3, we perform computational modelling to analyse the population-scale effect of the molecular inefficiencies identified in chapter 2. Specifically, we explore how factors such as cut-rate, DNA repair, and deposition interact with daisy-chain gene drives. We highlight a phenomenon we term "phantom-cutting", in which a multi-element gene drive attempts to bias another element that has already segregated away. We show that inefficiencies such as a reduced cut-rate and the formation of non-lethal resistance alleles disproportionately affect the standard daisy-chain gene drive design but less so for a modified design. We perform extensive parameter sweep

simulations, testing the complete range of values that the parameters may take when experimentally measured.

In chapter 4, we develop a meta-analytic database and analysis pipeline to carry out a systematic investigation of published gene drive crosses. We also develop a companion web tool to allow users to find relevant experimental data to parameterise computational models and inform the molecular design of new gene drives. We perform a systematic literature search and data extraction on a subset of relevant publications. We apply this toolset to investigate the effect of sex on two gene drive outcomes, gene drive inheritance, and somatic phenotype rates.

In chapter 5, we analyse a set of previously published experiments and new experiments performed by a collaborator based on observations made in chapter 4. We study the efficiency and mechanisms of three Cas9 regulatory variants with a drive element that targets the *white* gene in *A. aegypti* mosquitoes. We find that each nuclease expression variant causes increased inheritance of the gene drive element. Using the fact that the sex-determining locus is linked to *white* gene allows us to determine that, in our crosses, the mechanism of inheritance bias is homing. This contrasts with previous work that found inheritance bias through meiotic drive with the same transgenic lines.

In chapter 6, we critically discuss the impact of this thesis and the developments within the gene drive field as a whole.

1.2 Glossary

Gene drive: Gene drive refers to both the process by which a genetic element biases its own inheritance, and the genetic element itself[63]. Frequently shortened to just 'drive' for either meaning. A gene drive can be composed of multiple distinct genetic elements.

Homing endonuclease gene (HEG) drive: A form of gene drive that functions through the cleaving of a specific genomic locus followed by homology-directed repair (HDR) using the homologous chromosome as the template. Inheritance of the drive allele is increased when the non-drive allele is converted to a drive allele in the germline lineage of an individual heterozygous for the drive. HEG-based drives are also commonly referred to as "homing drives".

Nuclease: An enzyme that can catalyse the cleavage of nucleic acids. For HEGs, sequence-specific endonucleases are used.

CRISPR-Cas9: Clustered Regularly Interspaced Short Palindromic Repeats Associated Protein 9. An endonuclease protein that can be programmed to specifically cut a DNA sequence through a guide RNA (gRNA). Cleavage requires an additional three base-pair protospacer adjacent motif ('NGG') specified by interactions of the Cas9 protein with the target DNA.

gRNA: A ≈100 nucleotide RNA that complexes with the Cas9 protein and specifies a ≈20 nucleotide long target sequence to be cut by RNA:DNA base-pairing.

Gene drive allele/element: Genetic components that contain all or a subset of the genetic parts needed to induce gene drive. Gene drive mechanisms rely on endogenous pathways (such as DNA repair), but these are not considered part of the gene drive itself.

Target allele: The target, often wild-type, allele that the HEG can cut and convert. It can also be a 'synthetic' target site specifically introduced to contain the HEG's recognition sequence. In research contexts, the target is often a (haplosufficient) gene that, when disrupted, gives a phenotypic readout.

Donor and Recipient chromosome: In a heterozygous individual, the HEG is designed to copy itself from the 'donor' chromosome to the 'recipient' chromosome. In some studies, the homologous chromosomes can be differentiated by a marker separate

from the drive allele. This allows the original donor alleles to be distinguished from the newly created drive alleles on the recipient chromosome.

Homing/Drive conversion rate: Conversion of a target allele to a drive allele through interchromosomal HDR. Commonly measured as the fraction of target alleles that have been converted to a drive allele in the germline lineage of a drive allele heterozygote. If either the target or the drive allele have a tightly linked marker, this can be calculated very precisely. If not, it is estimated from the inheritance of the drive compensated for the 50% expected inheritance from a heterozygote through non-drive Mendelian inheritance.

Type-1 resistance mutations: A functional classification of mutations in the HEG target gene that leave the target gene's function intact. This is target-dependent, and the same sequence changes can result in different functional outcomes for different genes and different sites within the same gene.

Type-2 resistance mutations: Mutations in the HEG target gene that disrupt the gene's normal function.

Unconverted target alleles: Target alleles that have not been mutated or converted by the HEG. It is possible that the unconverted target alleles were cut, but DNA repair did not introduce any mutations. For many crosses, without additional molecular assays, intact target alleles cannot be phenotypically differentiated from type-1 resistance mutations.

Deposition: HEG nuclease protein, mRNA, or gRNA that has been expressed in the parent and carried over to the progeny. Deposition does not require the progeny to have inherited the gene that expressed those HEG components in their parent(s). However, the inheritance of a HEG allele can influence the outcome of deposition.

Shadow drive: Biasing of a particular allele that occurs as a result of deposited HEG components.

Somatic phenotype: HEGs are commonly tested by targeting a gene that gives a somatic phenotype when disrupted, such as a yellow body or white eyes in insects. For many experimentally tested HEGs, heterozygous individuals do not present with a somatic phenotype unless nuclease activity has caused the wild-type target allele to be replaced by the drive allele or a type-2 resistance mutation in a substantial proportion of somatic cells.

Knock-out phenotype: Uniform phenotype which is consistent with the target gene having been disrupted in most or all cells of that tissue (e.g., completely white eyes).

Mosaic phenotype: An intermediate phenotype between wild-type and the knock-out phenotype consistent with some cells within the organism retaining function of the target gene and the target gene having been disrupted in other cells.

Somatic nuclease/drive activity: Activity of the HEG nuclease in somatic tissues. In almost all cases, this is not intended.

Germline restricted nuclease/drive activity: Activity of the HEG nuclease in cells of the germline. Characterised by changes to the gamete frequencies without generating a somatic phenotype.

Embryonic nuclease/drive activity: Activity of the drive in the early embryo. Often characterised by both (error-prone) germline and somatic drive activity. Associated with individuals who have experienced parental deposition.

Chapter 2

The challenges in developing efficient and robust synthetic homing endonuclease gene drives

Sebald A. N. Verkuijl[1,2], Joshua X. D. Ang[2,3], Luke Alphey[2,3], Michael B. Bonsall[1], Michelle A. E. Anderson[2,3]

1: Mathematical Ecology Research Group, Department of Biology, University of Oxford, 11a Mansfield Road, Oxford, OX13SZ, U.K.
2: Arthropod Genetics, The Pirbright Institute, Ash Road, Pirbright GU24 0NF, U.K.
3: Current Address- The Department of Biology, University of York, York, U.K.

2.1 Abstract

Making discrete and precise genetic changes to wild populations has been proposed as a means of addressing some of the world's most pressing ecological and public health challenges caused by insect pests. Technologies that would allow this, such as synthetic gene drives, have been under development for many decades. Recently, a new generation of programmable nucleases has dramatically accelerated technological development. CRISPR-Cas9 has improved the efficiency of genetic engineering and has been used as the principal effector nuclease in different gene drive inheritance biasing mechanisms. Of these nuclease-based gene drives, homing endonuclease gene drives have been the subject of the bulk of research efforts (particularly in insects), with many different iterations having been developed upon similar core designs. We chart the history of homing gene drive development, highlighting the emergence of challenges such as unintended repair outcomes, 'leaky' expression, and parental deposition. We conclude by discussing the progress made in developing strategies to increase the efficiency of homing endonuclease gene drives and mitigate or prevent unintended outcomes.

2.2 Introduction

Gene drive is the ability of a genetic element to bias its own inheritance. This allows gene drive elements to spread a genetic change through a population even while having a fitness disadvantage ('selfish-DNA'). Genetic engineering at scale through engineered/synthetic gene drives may allow many currently intractable public health challenges caused by pest species to be addressed. In particular, insect pests such as mosquitoes have life-history traits (e.g., sexual reproduction and short generation times) that may make them amenable to gene drive interventions. The feasibility of using gene drives to fix a particular trait (population replacement) or suppress wild populations are both being investigated for addressing the harm caused by insect pests, in some cases with the same ultimate goal (e.g., eradication of malaria).

There are many examples of gene drives occurring in nature, acting through many different mechanisms[64]. Some types of gene drive rely on the action of sequence-specific DNA nucleases (enzymes that create DNA breaks). These have recently received a lot of attention by researchers following the discovery and characterisation of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems[65]. The programmable CRISPR nucleases, of which CRISPR-associated protein 9 (Cas9) is the most widely used, have provided researchers with powerful new tools to both facilitate genetic engineering and as constituent parts of gene drive mechanisms. Nonetheless, many important fundamental insights into building synthetic gene drive systems were gained before the use of CRISPR nucleases.

Double-stranded DNA breaks are a common occurrence in cells, and a range of mechanisms exist to resolve them. Under specific conditions, cells can use a homologous DNA template to prevent the loss of genetic information. This can be from an identical sister chromatid that is present during the S and G2 phases of the cell cycle, or the near-identical homologous chromosome. Generally, in diploid organisms, each chromosome in a homologous pair is contributed by a different parent and contains the same content with minor sequence variation (sex chromosomes often are an exception). Therefore, interchromosomal repair within a homologous pair will result in loss of heterozygosity, but under most circumstances, results in the genomic region retaining its function after repair.

Homing endonuclease gene drives (HEGs) can induce their own switch from a heterozygote to a homozygote state by creating a DNA break in the 'recipient' homologous chromosome corresponding to the locus of the HEG genetic material on the 'donor' homologous chromosome (Fig.2.1A). In effect, the coding sequence for the HEG may then be identified as missing from the cut chromosome and the HEG and linked sequences are copied over during repair of the DNA break (Fig.2.1B). If the transformed cell is part of the organism's germline lineage, the gene drive element will be propagated to the next generation with a higher frequency than would be expected from normal Mendelian inheritance. This copying or 'homing' process can repeat itself in subsequent generations and allows the HEG element to increase in frequency in a population, along with any associated genetic modifications that affect the desired

change in the population.

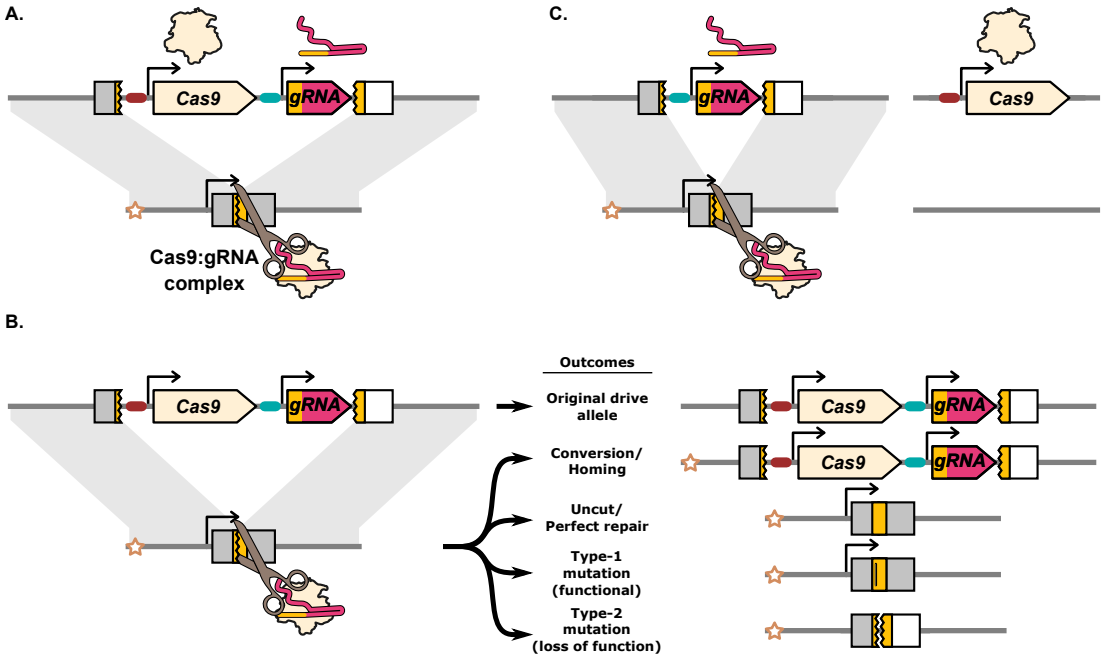


Figure 2.1: **Illustration of the Cas9:*gRNA* homing endonuclease gene drive inheritance biasing mechanism and potential DNA repair outcomes.** A. The drive expresses Cas9 and a gRNA which together form a complex that finds and cleaves the target allele. B. DNA breaks can be resolved by a range of different repair outcomes. Conversion occurs when HDR uses the homologous chromosome carrying the drive allele as a repair template. The star indicates the 'recipient' chromosome, and allows the original drive allele to be distinguished from a drive allele produced by homing. Alleles that are cut and repaired perfectly, as well as uncut alleles, remain unconverted. In addition to conversion, DNA repair can also result in mutations in the target gene. If the specific mutation(s) do not disrupt the function of the target gene they are classified as type-1. If the mutations do prevent normal function of the target gene they are classified as type-2. C. A component essential to the inheritance biasing process such as the gRNA or Cas9 gene can be located on a separate element producing a 'split-drive' configuration. The gRNA target sequence in the endogenous target gene is indicated by a yellow colouring. The target gene is shaded grey, unless disrupted by the drive or type-2 mutations at which point it is shaded white to indicate its loss of function. Type-1 mutations lose the gRNA target sequence (indicated by a vertical bar in the yellow target site) but remain functional.

In general, the HEG drives we describe here are designed and optimised for the homing inheritance bias mechanism. However, there are a number of ways through which nuclease-based drives have been described to bias their inheritance with seemingly subtle changes underlying the difference in mechanism. For almost all HEG studies, there is limited evidence on the actual underlying mechanism(s) giving rise to any observed inheritance bias and recent evidence suggests the mechanisms may be more heterogeneous than previously understood. An important hallmark of the homing process is the copying of the drive element onto the recipient chromosome. Many other nuclease drive mechanisms instead operate through decreasing the inheritance of the non-drive recipient chromosome. We will use the term inheritance bias or estimated homing when the specific experimental set-up was not strictly able to distinguish between inheritance bias through copying (homing) or exclusion of the chromosome

not carrying the drive allele.

Synthetic HEGs have, in almost all cases, been inserted into and targeted the sequence of an endogenous gene or targeted a separately inserted synthetic target gene (e.g., GFP). A principal reason for this is DNA sequence constraints. Many simultaneous DNA breaks in the genome may result in DNA damage-induced cell stress[66] and chromosomal rearrangement[67]. As such, synthetic HEGs are designed to only cut their specific target site and those targets are chosen to be unique within the genome. In addition, for HDR to occur, the region surrounding the DNA break must be (relatively) uniquely matched with the homologous chromosome, as homologous loci elsewhere in the genome may compete as evidenced by homing from non-paired sites[68;69]. Lastly, for the drive to affect a significant proportion of a population, its target must also be present in most individuals of the target population. These sequence constraints are generally only found in the (coding sequence) of highly conserved genes.

Beyond the sequence constraint, there are additional benefits that may come from placing HEGs in an endogenous gene. The 'effector' function of synthetic HEGs (e.g., female recessive sterility) may be most readily achieved by disrupting a specific endogenous gene directly with the drive element[70]. In research contexts, the target is often a gene that provides a phenotypic readout when disrupted. In addition, the chromatin environment associated with an endogenous (expressed) gene may be more permissive to the expression of the inserted transgenes[71–73], and an endogenous gene's promoter may even be directly used to express the drive genes[74;75]. The target gene's chromatin context may also influence Cas9 cutting efficiency and DNA repair[76]. Lastly, targeting highly conserved essential genes is one of the most important tools for addressing unintended repair outcomes which will be discussed more later.

Generally, the ability of a synthetic HEG to spread will depend on whether its efficiency at biasing its own inheritance can overcome its associated fitness costs. These costs depend on a number of factors: the particular application will matter, as population modification with a 'neutral' modification such as insecticide susceptibility or pathogen resistance will likely impose a far lower fitness cost than a modification designed to suppress the target population (cause a population decline). The actions of the drive machinery itself will also apply some fitness cost, and characteristics of the target species and population, such as size, density, gene flow, and density dependence will all factor into the drive requirements. More complex HEG designs required for 'self-limiting' drives[59] may also place higher requirements on the drive efficiency. Moreover, the HEG efficiency will also influence the required release frequencies, and the logistical costs and feasibility associated with the use of that particular system. As such, understanding and improving gene drive inheritance biasing efficiency and fidelity may allow for application in currently refractory species, and possibly decrease the cost of already feasible interventions.

To our knowledge, by the end of 2022 synthetic HEGs have, with varying inheritance biasing efficiencies (in some cases none), been reported in 9 species: *Saccharomyces cerevisiae*, *Candida albicans*, *Arabidopsis thaliana*, *Drosophila melanogaster*, *Plutella xylostella*, *Anopheles gambiae*, *Anopheles stephensi*, *Aedes aegypti*, and *Mus musculus*.

The research field has learnt much about the factors influencing HEG outcomes and explored different strategies for optimisation. However, much remains unknown, such as what specifically constitutes an efficient HEG, and what underlies the different outcomes observed with different drive designs. We will first present an overview of the field and the milestones achieved in the development of synthetic HEGs so far. Then we will discuss in detail specific technical challenges in developing efficient HEGs and end with potential solutions.

2.3 Milestones in the development of synthetic HEGs

Transferring a natural homing endonuclease gene drive

Natural HEGs were first identified in unicellular eukaryotes, fungi, and plants[64] and it was initially unknown if HEG based inheritance biasing mechanisms would function in animals. *S. cerevisiae* has a natural gene drive that relies on the I-*Sce*I meganuclease, which in its endogenous context cuts the large rRNA sub-unit of the biparentally inherited mitochondria[77]. Chan et al. integrated the I-*Sce*I nuclease into a synthetic docking site in *D. melanogaster* and separately inserted a fluorescent protein into the docking site to function as a recipient chromosome[68]. They for the first time demonstrated that a synthetic HEG could bias its own inheritance in animals. In that, and a follow-up study, they performed extensive tests of the I-*Sce*I drive with different regulatory sequences upstream (promoter and 5'UTR) and downstream (3'UTR) of the I-*Sce*I nuclease coding sequence[68;78]. The best-performing drive used the promoter of the Rcd-1r gene and the β-Tub56D 3'UTR aiming for spermatogenesis-specific expression. This drive converted 23% of the target alleles located on the recipient chromosome, achieving an overall inheritance of ≈62% (≈50% from donor chromosome + ≈12% converted recipients). The majority (63%) of target alleles on the recipient chromosome appeared to remain unmodified, likely uncut.

While a substantial number of promoter/3'UTR combinations achieved higher cut-rates than Rcd-1r/β-Tub56D, they resulted in lower or even no HEG conversion. Instead of copying the HEG element through HDR, alternative DNA repair pathways created mutations at the site of DNA cleavage. These results indicated that simply creating a DNA break was not enough for efficient homing, the timing of nuclease expression was seemingly essential for efficient conversion. In addition to potentially competing with the homing process, these mutations create cut-resistant alleles that are inherited by the next generation. Mathematical modelling and cage trials have indicated that these cut-resistant alleles can prevent a drive from reaching fixation, or even spreading effectively in real-world applications[79–84].

Work by Windbichler et al. demonstrated that the I-*Sce*I HEG could also function in the disease-relevant *A. gambiae* mosquito[85]. Expressed from a male-specific promoter, the HEG reached inheritance rates of 86%. Moreover, they showed using small-scale cage experiments the first evidence that a synthetic HEG could spread within a receptive population. Despite similar, and many additional promoter/3'UTR combinations

having been tested in *D. melanogaster*, the inheritance bias achieved with the *A. gambiae* drive was higher. This was the first suggestion that some organisms are more receptive to HEG based inheritance biasing than others. Yet, even in *A. gambiae* the inheritance bias was likely too low for most applications. Moreover, these HEGs would not function outside of a specifically modified lab strain: both *D. melanogaster* and *A. gambiae* do not naturally contain the I-*Sce*I target sequence and in each case, a synthetic target allele had to be created for the drives to function.

The cost of retargeting

To address the targeting limitations, Chan et al. used site-directed mutagenesis of the I-*Onu*I meganuclease to change its recognition sequence to allow it to target a closely related sequence naturally found in *Anopheles* mosquitoes[86]. They placed this *Anopheles* target in a GFP reporter and tested the inheritance bias in *D. melanogaster* males using the Rcd-1r promoter and β-Tub56D 3'UTR. While two I-*Onu*I variants biased their inheritance to the same degree as the I-*Sce*I drive[68], they only did this with far higher overall cut-rates (therefore generating more mutations). Moreover, there were indications that a mutation introduced into the I-*Onu*I catalytic site to achieve these higher cut-rates was causing reduced fertility, possibly due to sequence promiscuity. Furthermore, the *Anopheles* gene that the modified I-*Onu*I could target happened to relatively closely match the natural I-*Onu*I targeting sequence. Clearly, the creation of synthetic HEGs would greatly benefit from more programmable, yet specific, nucleases.

The first and second generations of programmable nucleases came in the form of zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs). The only reported use of ZFNs and TALENs in a HEG system was by Simoni et al.[87] with the Rcd-1r promoter and β-Tub56D 3'UTR in *D. melanogaster*. In males, ZFN and TALEN-based HEGs achieved homing rates of 34% and 49%, respectively. These homing rates were higher than with I-*Sce*I, but with equivalent or worse cut-to-homing ratios. While the TALEN HEGs had overall higher homing rates and better cut-to-homing efficiency than the ZFN HEGs, only the ZFN HEGs were able to spread significantly within small cage trials. It became apparent that the programmability of these nucleases came at a cost: repetitive genetic sequences. ZFNs and particularly TALENs are composed of large repeating DNA binding 'units' in which only a few amino acids are changed to specify the target sequence. This resulted in repetitive drive constructs, which in turn were found to make the drive unstable, losing function at a high rate due to internal recombination and/or partial homing. Only 40% of the TALEN and 75% of the ZFN inheriting progeny resulting from (partial) homing in the first generation could themselves home in the next generation.

Together, the above work demonstrated that the HEG mechanism could indeed work in animals and spread in small cage populations. However, the difficulty in programming meganucleases, and the shortcomings of ZFNs and TALENs meant that with the available tools it would be a momentous task to create an effective HEG drive system that would spread in non-ideal conditions. That is, until the discovery of the CRISPR

nucleases.

Programmable RNA guided gene drives

Starting in 2012, a new generation of molecular tools greatly accelerated our ability to perform gene editing. Central to this revolution has been the discovery of new easily programmable nucleases, the best-known version being Cas9 from the type-II CRISPR system of *Streptococcus pyogenes*[65]. Cas9 can, with few limitations, be targeted to almost any DNA sequence by straightforward RNA to DNA base pairing through a short 'guide' RNA (gRNA). If a sufficient match is found (not necessarily perfect), Cas9 will then create a double-stranded break. In a very short time-frame, CRISPR-based synthetic HEGs were reported in *S. cerevisiae*[60], *D. melanogaster*[88], *A. stephensi*[53], and *A. gambiae*[47].

The first CRISPR HEG reported in yeast demonstrated near perfect (>99%) inheritance over multiple generations, and in multiple strains[60]. Moreover, this work demonstrated the feasibility of using HEGs with more advanced modifications which had been previously proposed[70;89]. This included carrying a cut-resistant, but functional version of the target gene on the HEG allele, reversing the changes of one drive with another, and split-drives. In a split-drive, one component essential to the drive mechanism is housed on a separate locus, generally by separating Cas9 from its gRNA (Fig.2.1C). This allows the HEG that carries the gRNA to be safely tested, as it will only behave like a HEG in a lab strain that already expresses Cas9 and will not spread in wild populations. The synthetic target sites needed for earlier nucleases provided similar protection against unintended spread beyond the laboratory. The publication of CRISPR HEGs in yeast demonstrated that CRISPR gene drives are capable of extremely high conversion efficiencies and gave an initial indication that the CRISPR drives do not suffer from the same genetic instability issues seen with previous programmable nucleases. While these results were encouraging, the natural I-*Sce*I HEG also worked extremely well in yeast but failed to reach similar efficiencies in animals. Fortunately, the first CRISPR HEGs in *D. melanogaster*[88], *A. stephensi*[53], and *A. gambiae*[47], each using a *vasa* regulatory element, reported inheritance rates over 90%, a massive improvement over the non-CRISPR HEGs. However, each publication also laid bare challenges that could prevent the effective spread of CRISPR HEGs.

Gantz et al. reported the first CRISPR HEG in *D. melanogaster*[88]. The HEG was inserted in and disrupted the X-linked *yellow* gene, limiting drive to XX females only. Gantz observed loss of function of the yellow gene target in almost all progeny (97%) of gene drive heterozygous mothers. This also occurred in female progeny suggesting that the maternally inherited drive converted the paternally contributed functional yellow allele in the early embryo. However, later publications found substantially lower inheritance rates (76-85%) with a near-identical constructs but including a fluorescent marker[90;91]. It is probable that part of the seemingly super-Mendelian inheritance of the HEG observed in the earlier study[88] was due to the maternal 'deposition' of the Cas9:gRNA complex without inheritance of the drive expressing allele itself[91]. The

deposited nuclease and gRNA could cause the disruption of the *yellow* target gene in the absence of inheritance of the HEG itself. This could be problematic, as even with high rates of HDR the (repeated) cutting in individuals that did not inherit the drive allele will lead to mutations in the target gene. Moreover, the first HEG reported in *A. stephensi* showed that even when the drive is inherited to serve as an HDR template, deposition can have a negative effect.

Gantz and Jasinskiene et al. reported the development of a synthetic CRISPR HEG in *A. stephensi*[53]. Drive inheritance was scored separately in the progeny of males and females that inherited the drive allele from their father, and in the progeny of males and females that inherited the drive allele from their mother. Heterozygous parents of either sex passed along the drive element to 98-99% of their own progeny when that parent inherited the drive allele via the paternal line (drive-carrying grandfather). However, strikingly, the maternal contribution of the drive allele (and accompanying maternal deposition) caused germline conversion rates to sharply drop, with only 56% of progeny from males and 62% of progeny from females having inherited the drive element. Moreover, unlike autonomous expression from the drive allele, the maternally deposited nuclease also affected somatic tissues. Maternal deposition was seemingly resulting in nuclease activity early in the embryo when HDR was not favoured, converting the target alleles to resistant alleles that could no longer be converted when expression occurs in the germline and HDR is favoured. Even if deposition-based conversion had been efficient, somatic drive activity has the possibility to cause its own issues. This was most strikingly highlighted by the first *A. gambiae* CRISPR HEG targeting candidate population suppression genes.

Hammond et al. identified three genes that, when disrupted, confer a recessive female sterility phenotype in *A. gambiae*[47]. They created Cas9 HEGs in each gene and demonstrated extremely high inheritance rates in both females and males (99%). However, the HEG heterozygous females unexpectedly produced only 0-9% of the number of larvae wild-type females did - a sterility effect that was intended to be limited to the drive homozygotes. They showed that *vasa*2-Cas9 expression was not fully germline restricted, and the nuclease was being expressed in somatic tissues. This lead to nuclease activity in some somatic cells, causing the remaining functional copy of the target gene to be lost, and the recessive phenotype to present in (initially) heterozygote individuals. Additionally, they identified mutations in the female fertility genes that, while preventing Cas9 cleavage, did not disrupt the normal function of the genes. They proposed that depending on the reproductive load of the drive, these functional resistant mutations could prevent the collapse of the target population and demonstrated this in a follow-up publication[92].

Together, this initial set of studies demonstrated that maximising inheritance bias, while minimising unintended fitness costs and the creation of inheritable resistance mutations, remains a challenge with Cas9 based HEGs. Moreover, subsequent studies with HEGs in *M. musculus*[75;93;94], *A. aegypti*[3;95;96], *P. xylostella*[97], and *Arabidopsis thaliana*[98] have generally proved less efficient than in *Drosophila* and the *Anopheles* mosquitoes. Unintended DNA repair outcomes, inopportune drive expression, and

deposition have emerged as the most important impediments to developing efficient HEGs.

2.4 The main technical challenges facing synthetic HEGs

Unintended DNA repair outcomes

In most cases, the highest possible inheritance biasing rate is desired when developing a HEG. However, target alleles that remain uncut may be converted in the following generations, whereas cut-resistant mutations cannot. As such, the ratio between drive conversion events and unintended outcomes can often be more important than the inheritance biasing rate alone. HEGs may be more susceptible to resistance than other gene drives because DNA damage and repair are directly involved in their inheritance biasing mechanisms. This means a HEG can directly create resistance to itself that was not already present in the target population. This 'induced' resistance is in the form of sequence changes to the target allele by unintended DNA repair pathways, often collectively described as non-homologous end joining repair (NHEJ), that prevent further cutting by the nuclease. Problematically, this resistance, if arising anywhere in the germline lineage, can, in addition to lowering the inheritance biasing efficiency of the drive, also be inherited and contribute a new resistant allele to the population's gene pool.

The consequence of the mutation depends both on the target gene and the nature of the DNA lesion. With 'type-1' resistance mutations preventing cutting by the nuclease, but otherwise leaving the function of the target gene intact, and 'type-2' mutations preventing cutting by the nuclease as well as preventing regular function of the gene[90]. Generally, type-1 resistance mutations are substitutions or small in-frame (n·3bp) insertions or deletions in the exons of protein-coding genes. Type-2 resistance mutations are more likely with mutations causing a frame-shift in the exons of protein-coding genes and large insertions and deletions. The importance of the amount and type of resistance mutations produced depends much on the particular application of the HEG. For a drive targeting a neutral locus, the distinction between type-1 and type-2 resistance will have no practical significance. In contrast, for a population suppression drive that aims to disrupt a particular (essential) gene, the ratio of type-1 and type-2 resistance alleles produced can be far more important than the overall amount of resistance alleles. Type-2 resistance alleles may slow the spread of the HEG, but they ultimately still contribute to the HEG effector function (disruption of the target gene). In contrast, even extremely rare type-1 alleles may allow for a population to quickly rebound or even be largely unaffected by a suppression HEG[92].

In addition to non-HDR outcomes, incomplete or internal HDR may also be a significant source of mutations. Alleles carrying parts, but not the whole drive element, have been reported in a number of publications studying CRISPR-Cas9 HEGs[47;48;81;82;90;99]. In some cases, this can be explained due to internal recombination similar to what

occurred with the repetitive ZFN and TALEN HEGs discussed earlier[87]. Oberhofer et al. tested a HEG with four gRNAs and found that the repetitive sequences introduced with multiple gRNAs likely caused the drive construct to internally recombine[48] (Fig.2.2A). It is not clear if internal recombination is fully independent of the homing mechanism or if a significant fraction of the incomplete drive alleles are created through partial HDR from an otherwise intact donor allele. In some cases, the recovered incomplete drive alleles more strongly suggest they emerged due to partial copying.

In contrast to I-*Sce*I, ZFN or TALENs, Cas9 identifies its target through Watson-Crick base pairing of ≈20 nucleotides of the gRNA with the genomic DNA. Some publications have specifically identified partial drive alleles that are consistent with the gRNA gene target sequence having been used as one of the homology 'arms' during the homing process[82;90]. This results in only part of the drive allele being identified as 'missing' from the recipient chromosome, generating a partial copy (Fig.2.2B). However, different repair processes may give the same ultimate product (Fig.2.2C). Internal recombination and partial homing may be more common than is currently recognised. In almost all cases, inheritance rates are determined by scoring the presence of a dominant fluorescent gene linked to the HEG which may also be a partial drive allele[48]. While in most cases incomplete HDR should result in a type-2 resistance mutation, these partial homing events may pose a problem for drives with distinct cargo genes or sequence changes. Of particular importance, HEGs have been developed that carry sequences that rescue the function of the gene that they disrupt and partial homing of only these rescue sequences may create type-1 resistance alleles.

There may be a set of DNA repair outcomes, such as mitotic recombination and meiotic drive, that are underappreciated because they do not leave a distinct mutational signature. With mitotic recombination, DNA repair causes a dividing cell to produce daughter cells, where one daughter cell has two copies of a paternal chromosome region, and the other has two copies of the maternal chromosome region (Fig.2.2D). The production of individual cells homozygous for a particular parental gene resembles the outcome of homing, however, mitotic recombination does not directly bias the inheritance of any allele as reciprocal cells homozygous for the other allele are also created. However, in one study, *D. melanogaster* females carrying a single copy of the dominant female sterility inducing ovo^{D1} transgene could nonetheless produce viable offspring due to mitotic recombination induced by nos-Cas9[100]. Mitotic recombination can seemingly be a substantial outcome of DNA damage, and may be relevant in a HEG with a dominant acting effector such as sex-conversion. Mitotic recombination is most commonly studied by targeting both homologous chromosomes[100;101], however it has been demonstrated to occur when only one homolog can be cut[102].

Finally, if DNA repair fails altogether, the cut recipient chromosome may instead be lost (Fig.2.2E). Loss of haploid cells or fertilised embryos carrying the recipient chromosome will in effect increase the relative inheritance of the donor chromosome, providing a potential separate mechanism of inheritance bias for HEG drives. If the recipient chromosome is marked, inheritance bias through homing or through the

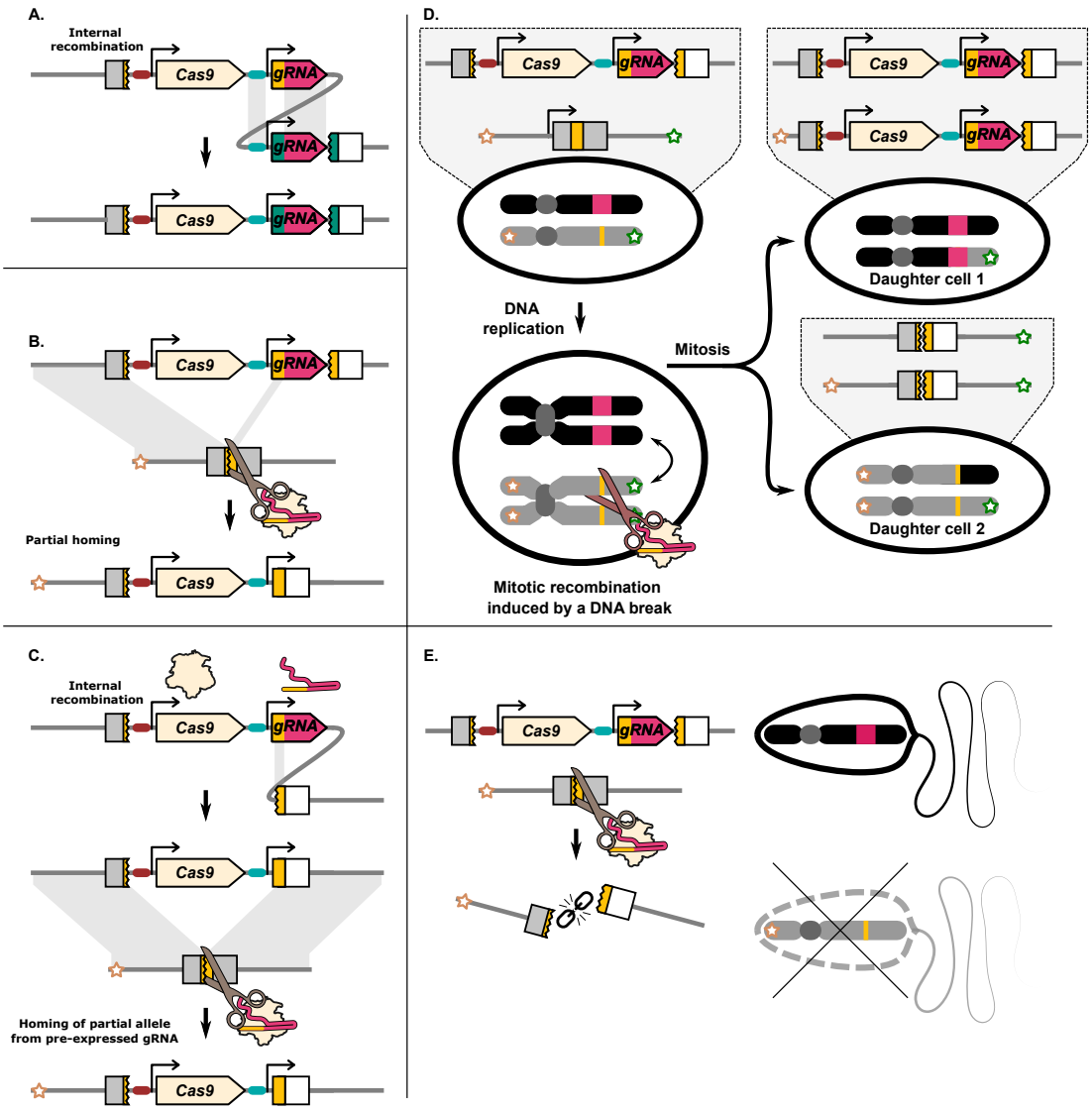


Figure 2.2: **Illustration of DNA repair outcomes that may be associated with a HEG** A. Repetitive sequences within the HEG may lead to internal recombination. The gRNA promoter and the constant gRNA 'backbone' sequence in multiplexing drives may be particularly susceptible to internal recombination. B. Partial homing by means of the gRNA target sequence located on the drive allele. The gRNA gene contains sequence homology to its target. This may allow for partial homing. C. Recombination of the gRNA target sequence with a partial target sequence adjacent to the drive element. If the resulting allele homes it may resemble the product of partial homing even if a chromosome marker is present. D. Cas9-induced DNA breaks have been reported to lead to mitotic recombination. This can produce two daughter cells that have loss of heterozygosity. Daughter cells generated by mitotic recombination can under some circumstances resemble products of homing. E. In some cases, a DNA break can lead to the loss of the target/recipient chromosome. This can result in inheritance bias through meiotic drive.

loss of the recipient chromosome can be distinguished. An under-representation of the recipient chromosome marker has been reported in multiple publications with an element otherwise expected to function through homing[50;91;103], with in some cases meiotic drive seemingly exclusively mediating the observed inheritance bias[3;95]. In one study, under-representation of a restriction enzyme site nearby the I-*Sce*I cut site

on the recipient chromosome was suggested to be due to DNA repair after homing also replacing the nearby marker (termed 'co-conversion' or 'copy-grafting')[85]. All the pre-CRISPR HEG studies we discussed performed crosses with marked chromosomes. However, for CRISPR HEGs only a small minority of studies have used a marked recipient chromosome, making it difficult to judge the extent of this phenomenon.

Maximising the efficiency of HDR after a Cas9-induced DNA break is a major topic of research because of its broad applicability to biological and medical sciences[104]. Most research into site-specific HDR has been with an exogenously supplied repair template, and less is known if or what the specific dependencies are for efficient interchromosomal HDR. In addition, many of the interventions that may be used to boost HDR may not be suitable for a gene drive context. Below we will discuss some specific alterations to HEG design that have been investigated in an attempt to steer the number of resistance alleles, the ratio between type-1 and type-2 alleles, and mitigate their effect once they do emerge. The most important of these has been to limit HEG expression to when HDR is more likely, which has a number of additional benefits. However, actually limiting drive activity to this 'ideal' window has been challenging.

Spatial and temporal restricted drive expression

For a drive to function as a mechanism for super Mendelian inheritance, homing need only occur in the relatively small number of cells that make up the germline lineage (Fig.2.3A). Editing in any other cell lineage forming somatic tissues does not contribute to the drive inheritance biasing rate and in most cases not its effector function. Indeed, 'somatic' cutting can be a significant source of additional fitness costs. In the most direct sense, many proposed population suppression drives will perform best when homing is tissue-restricted. These drives rely on heterozygote individuals being unaffected by a particular modification (e.g., disruption of a haplosufficent essential gene), yet passing the drive along at increased rates. This is achieved by restricting homing (and therefore induced homozygosity) to the germline. The effector modification remains heterozygous in tissues where it is required for normal function. An example of this is the *doublesex* targeting *A. gambiae* drive[51]. More generally, for any drive, the unnecessary activity of the drive in somatic cell lineages may contribute to an additional fitness cost of the drive such as from off-target effects[105]. This is compounded by the fact that the ability to perform HDR varies strongly by cell type and on-target resistance mutations that carry a significant fitness cost may be more likely to emerge in cells of somatic lineages.

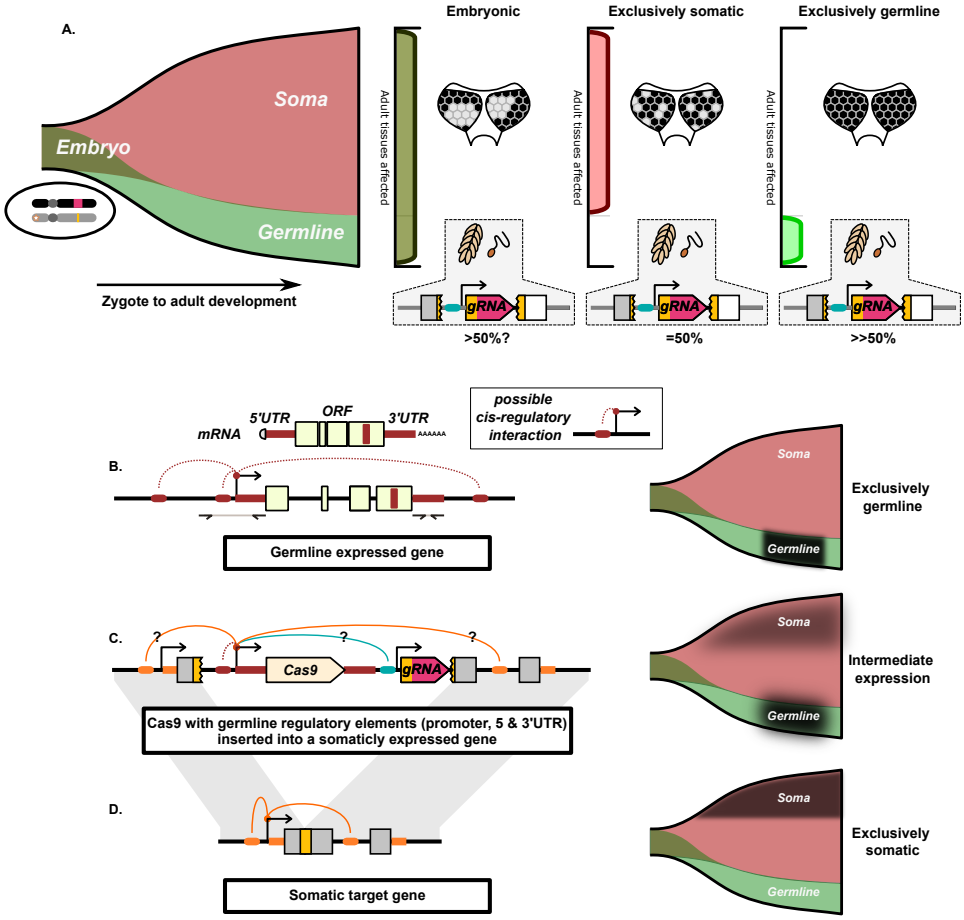


Figure 2.3: **Illustration of the interaction of spatial and temporal restricted drive expression.** A. Drive activity in the early embryo can simultaneously affect cells of the somatic and germline lineages. Drive activity later in development can independently affect the germline and somatic cells. The funnel represents the cells that compose an individual as it develops from a zygote to an adult (left-to-right). Early in development, cells are simultaneously part of both the germline and somatic cell lineages. Only later in development do the germline and somatic cell lineages diverge. The pattern of somatic mosaicism in adult tissue can under certain circumstances be indicative of the timing of cutting, here indicated by different patterns of loss of pigment in the eye. In these cases, the later in development the somatic gene is disrupted, the more fine-grained the mosaic pattern. B. Cas9 is commonly engineered to mimic the expression of endogenous germline-restricted genes. This is commonly approached by use of the promoter, 5'and 3' UTR of a germline-restricted gene. The 5' and 3'UTR can be identified from mRNA transcripts. Generally, a few Kb of sequence upstream of the 5'UTR are taken to capture the putative promoter. We have indicated additional speculative cis-regulatory interactions important for germline-restricted expression that are not captured by this approach. C. The Cas9 gene with the germline regulatory elements is inserted into the target gene's locus. Cas9's expression may be affected by cis-regulatory interactions with the target gene and with other drive components. We speculate that this may result in the Cas9 gene taking on an intermediate expression pattern resulting from the combination of different cis-regulatory interactions. D. The target gene commonly has a somatic expression pattern that may not be conducive to homing. For B. C. and D., the proposed expression pattern of the respective allele is indicated by a black overlay on the funnels.

A prominent hypothesis is that interchromosomal HDR after a DNA break is most likely if the DNA break coincides with meiosis I[49;50;64;75;81;93;94;97;106–110]. During

this time chromosomal homologs exchange information through crossing-over events. The alignment of the homologs in the cell and activation of particular DNA repair machinery[111–113] may make this timing more suited for interchromosomal copying of the HEG. As such, almost all synthetic HEGs have been designed to be active in the germline by flanking the nuclease transgene with the putative promoter, 5'UTR, and 3'UTR sequences of an endogenous germline-specific gene (Fig.2.3B). The most widely tested have been sequences from the *nanos* and *vasa* genes.

There are a number of examples where the locus from which the HEG components are expressed seems to affect otherwise identical drive elements[78;93;96;114,115] compared to[90;110]. One potential major challenge of achieving restricted expression is that a HEG inserted into an endogenous gene may be influenced by that gene's cis-regulatory elements and broader chromatin context (Fig.2.3C and 2.3D), which under normal circumstances facilitate the specific expression pattern of the target gene. Ironically, for drives inserted into essential genes these regulatory elements may prime the HEG to express in the cells that drive activity would be most undesirable. This in effect, can make that locus one of the worst possible places for the HEG to be inserted to prevent 'leaky' expression coinciding and interfering with the target gene's activity. Split-drives, where the Cas9 is expressed at an unrelated locus, can avoid this regulatory element mismatch, but this may also cause them to be behave differently if reconstituted to a single element drive[116]. The challenge of avoiding leaky expression is further compounded by some enhancers and other regulatory elements being located in the coding region of genes[117]. The presence of such elements in the target gene, and the absence of those elements from the germline genes, could alter the expression of Cas9 away from the germline-restricted expression pattern the HEG is trying to recapitulate. Finally, the regulatory components of other drive components (e.g., fluorescent marker or cargo genes) may also interfere with the intended expression pattern of the nuclease and gRNA genes[115].

Somatic cutting, in the absence of maternal deposition, has been reported for many Cas9 expression regulatory elements[3;53;81;95;118]. As described above, this is generally detrimental, however, there are a limited set of cases where somatic conversion can be an intended part of the drive effector mechanism. Carrami et al. aimed to develop a sex-conversion drive disrupting the autosomal transformer (*tra*) gene[99]. By deliberately selecting promoters that would be active in somatic tissues, homozygous disruption of *tra* by the drive and somatic mutations would, in the medfly, convert XX females into fertile males[119]. They performed their experiments in *D melanogaster* in which *tra* disruption leads XX individuals to develop into infertile pseudo males. However, in practice, the XX individuals displayed an intermediate intersex phenotype (and were infertile). Males were unaffected by the somatic disruption of *tra* and displayed modest estimated homing rates (≈56%) in their germline. This work highlights that even in cases where somatic activity is desired, achieving a uniform disruption of the target gene in all cells can be a challenge. The outcome of such intermediate conversion, with some cells converted and others not, is called mosaicism and in many reports, this is how somatic HEG activity presents.

While theoretically somatic and germline activity can be fully distinct processes, in some cases, the drive activity that gave rise to the somatic conversion did not necessarily (only) happen in the wrong cell lineages, but also at the wrong developmental time (Fig.2.3A). Drive activity in the early embryo can convert cells that go on to give rise to both the germline and somatic tissue, producing the associated somatic phenotype later in development. In these cases, preventing leaky expression early in development may simultaneously lower the production of resistance alleles (if HDR is indeed not favoured) and decrease disruption of the target gene in somatic tissues.

Another potential source of resistance mutations occurs at the other end of differentiation, post-meiosis. Any recipient chromosomes that escaped cutting in the germline will be separated from the donor chromosome once meiosis has occurred. If transcribed or translated drive components persist, cutting may occur after this point, resulting in repair by interchromosomal HDR being impossible. DNA damage near or post-meiosis may therefore be a source for resistance alleles[90;115].

It is currently not clear if significant cutting occurs after meiosis. However, that expressed/translated HEG components persist into haploid cells that do not contain the HEG genes has been established for many drives. This is because the HEG components can go on to affect the fertilised zygote. This is the phenomenon of deposition we introduced earlier, and it has similar consequences to that of early embryonic leaky expression. However, there are important differences between the two processes.

Parental effects (Deposition)

Many publications studying HEGs have noted that genetically identical individuals will show different somatic phenotypes and inheritance biasing efficiencies depending on which parent contributed the Cas9 and gRNA genes[48;81;90;99;103;115]. These types of parental effects have even been observed in individuals that did not inherit any genetic components of the drive, indicating the deposition of already expressed drive components. In almost all cases, an exclusive maternal effect is observed, where a female carrying the HEG transgene(s) is thought to contribute the gRNA and/or nuclease protein/mRNA to her haploid eggs. While these parental effects are commonly referred to as deposition, it is important to note that for some HEGs an alternative or additional mechanism such as imprinting has not necessarily been excluded.

A key difference between 'leaky' embryonic expression and embryonic cutting by deposition is that in the case of deposition, cutting can occur in the absence of inheritance of the drive. As such, even if interchromosomal HDR were favoured, deposition may result in the target allele being cut when not paired with a HEG allele. In addition, the activity of the deposited drive components may be expected to be early in development, affecting both the somatic and germline cell lineages. Cas9 protein half-life in cells and embryos has been estimated to be (substantially) less than 24 hours[120;121]. In *Drosophila*, the first meiotic divisions occur in third instar larvae (>2–days) in males, and early pupal stages (>3–days) in females[122]. This suggests

that if Cas9 protein stability is limited to hours, it would not persist long enough to overlap with meiosis I in many species. Moreover, even if Cas9 protein persists to when meiotic divisions occur, most target alleles may already have been cleaved earlier in development. The activity window of deposited Cas9 mRNA is harder to predict. The mRNA's translation in the embryo would likely delay and extend Cas9's window of activity, while 5' and 3'UTR sequences in the Cas9 mRNA copied from a germline gene may specifically limit the timing and location of translation. Moreover, it is possible that if mRNA deposition were to occur, the translated Cas9 would only become active in progeny that inherited the gRNA-expressing gene, as the gRNA is highly unstable when not in complex with Cas9 protein[123–125].

In general, deposition can be a substantial issue for effective homing, but it should be noted that the fitness costs associated with deposition-induced drive activity may be less detrimental than that of equivalent leaky expression drive activity[126]. This is because deposition affects the progeny independently of if they have inherited the drive or not. In contrast, the fitness cost of leaky expression is limited to those individuals carrying the drive element. However, the possibility of creating inheritable (type-1) resistance mutations in non-drive inheriting progeny may still make deposition substantially more problematic.

Isolating and quantifying the resistance allele contribution of deposition can be difficult. Maternal deposition is most readily identified by disruption of the paternally contributed allele in the absence of inheritance of the drive allele. However, the maternally contributed allele may have been disrupted at the same time, or instead at any point in the mother's development. As such, it is not possible to directly distinguish between germline resistance mutations and those that arise in the early embryo due to deposition; but there are two key pieces of evidence that may suggest a particular timing.

Firstly, particular DNA lesions have been found repeatedly in multiple offspring from the same parent[81;90;99;114]. This can suggest that the specific mutation arose in the HEG drive parent's germline, was replicated by cell divisions, and was passed along to multiple offspring. Interestingly, the fraction of offspring inheriting the same mutation may provide evidence for when it occurred in the parent's germline[114]. However, it should be noted that while NHEJ mutations are variable, their scope can be heavily influenced by the sequence context of the DNA break increasing the likelihood of identical mutations arising independently[127;128]. Nonetheless, statistical analyses can detect if particular mutations co-occurred more in progeny of the same parent than between individuals with different parents[81].

The second observation is mosaicism. If there is a mix of mutant alleles found within an individual progeny, this indicates they arose after the first genome replication and did not arise in the parent's germline. However, this analysis only works if the experiment is performed in such a way that the paternally contributed allele is cut-resistant and cannot in itself generate a mosaic outcome. As a caveat to this, Chan et al. has suggested that a mosaic phenotype observed in progeny that did not inherit their I-*Sce*I HEG allele was due to the inheritance and replication of an unrepaired DNA

break instead of parental deposition[68]. However, this was not further investigated.

Champer et al. reported that for a single drive element (comprising both *nanos*-Cas9 and a gRNA), the degree of embryonic cutting they saw in progeny subjected to maternal deposition and inheriting the drive was similar in the cases where the mother was heterozygous or homozygous for the drive[90]. They offered that this "implies that most maternal Cas9 persisting to the [progeny's] embryo stage was expressed after drive conversion events" in the mother. This was further supported by evidence that the rate of embryonic cleavage from deposition was lower when drive conversion was less likely, such as when mothers carried mostly resistance alleles instead of an additional wild-type target allele. Interestingly, in the progeny from these crosses the fraction of progeny with a somatic phenotype was similar regardless of whether a drive allele was inherited or not. This implied that most Cas9 persisted through to the embryo after maternal expression in diploid cells, rather than being expressed after meiosis and correlating with drive inheritance. These experiments provide some insight into the complex ways in which deposition can manifest for a particular expression pattern. However, it should be noted that deposition, or its interpretation, can be context-dependent, as is well illustrated by the *zpg*-Cas9 *A. gambiae* drive. Crosses with maternal[83], paternal[51], or no deposition[129] have all been described for this same drive. Moreover, in *Drosophila* it has been reported that the embryo resistance rates can be strongly affected by the background genetics of crossed individuals[130].

As introduced earlier, early embryonic nuclease activity associated with maternal deposition can cause a drive to affect somatic tissues (with the accompanying issues described above). While DNA repair associated with deposition-based cutting seems to be more error-prone, interchromosomal HDR does occur. The use of split-drives has allowed deposition to be investigated in more detail. Many HEGs have been tested in the form of a split-drive, where the Cas9 is located at an unlinked locus. Generally, this results in 50% of progeny inheriting the Cas9 gene independent of inheritance of the main drive allele carrying the gRNA(s). A number of publications have noted that individuals that only inherit the gRNA element can nonetheless pass it along at increased rates and this has been termed 'shadow drive' (homing through deposited factors)[50;103;115;118] (Fig.2.4). The activity window of deposited nuclease may be expected to be early in development with limited persistence. As such, examples of efficient shadow drive may provide a counterpoint to the hypothesis that homing is limited to meiotic cells that emerge later in development. Important in this interpretation is that deposition patterns can seemingly differ for the nuclease, gRNA, and Cas9:gRNA complex.

Kandul et al. created a split-drive in *D. melanogaster* targeting the *white* gene, and carrying a gRNA targeting the *yellow* gene in trans[133]. They tested four regulatory elements expressing Cas9 from a separate locus, and mediated significant inheritance bias with an average estimated homing rate of 73%. Interestingly, when the nuclease was carried by the grandmother, the estimated homing rate in the parent's germline was roughly the same (69%), even when the Cas9 gene had not been inherited. Shadow drive through maternal deposition with the four Cas9 regulatory elements they tested

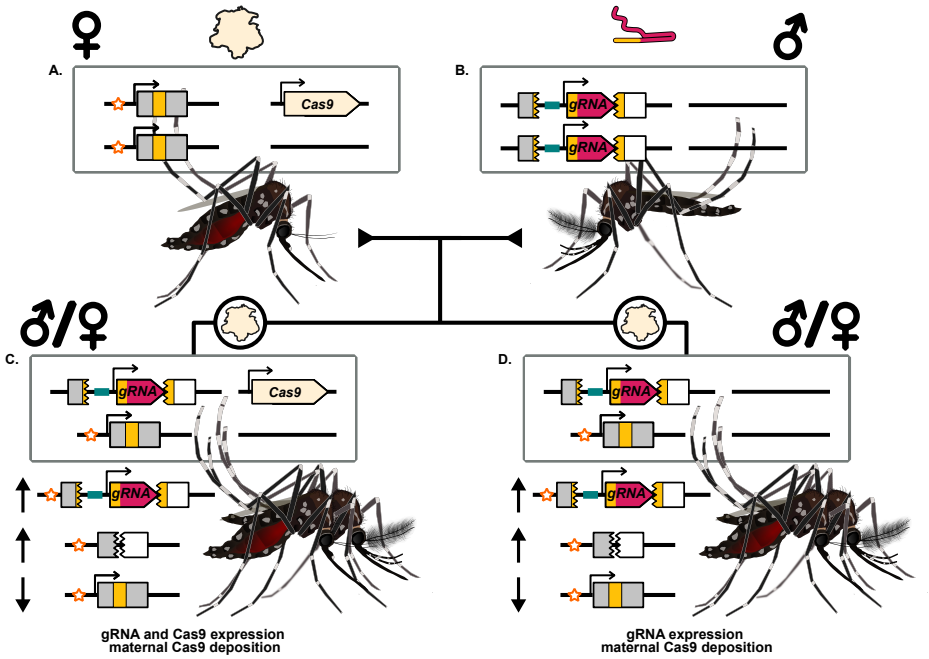


Figure 2.4: **Illustration of maternal deposition and shadow drive in the context of a split-drive.** A. A female carrying a Cas9 split-drive element and depositing Cas9 protein into her eggs. B. A male that carries the gRNA drive element and does not deposit the expressed gRNA. C. Progeny of either sex that inherited the two transgenes from each parent separately. Expression of the Cas9 and gRNA genes can allow for inheritance bias. D. Progeny of either sex that inherited the gRNA gene from their father but not the Cas9 gene carried by their mother. Maternal deposition can nonetheless provide a source of Cas9 that can complex with the expressed gRNA and mediate inheritance bias of the gRNA drive element. This is termed shadow drive. The recipient chromosome marker (star symbol) is used to indicate the target alleles that have been replaced by the drive element. Inheritance biasing mechanism of drive expression and/or deposition could operate through a non-copying mechanism. The arrows indicate a relative increase (upward arrow) or decrease (downward arrow) of the particular allele. While mosquitoes are illustrated[131;132], these results have primarily been documented in *Drosophila*.

was seemingly just as efficient at biasing the inheritance of the gRNA HEG element as germline expression of the nuclease was. However, in these first crosses the gRNA gene was contributed by the grandfather, providing no opportunity for the Cas9:gRNA to complex before being deposited. In a subsequent cross with the grandmother carrying both the Cas9 and gRNA genes, the estimated homing rate dropped sharply to 9.2% in trans-heterozygotes (gRNA + Cas9) and to 6% in heterozygotes (gRNA only). When both Cas9 and gRNA were maternally deposited, cleavage could occur in the early embryo, forming resistance alleles, which then prevented drive conversion at a more opportune stage. This did not occur with maternal deposition of only Cas9 (not gRNA) into an individual that can nonetheless express the gRNA. gRNA expression, thought to be constitutive, and subsequent complex formation with the deposited Cas9 protein seemingly limited cutting to a more opportune stage for inheritance bias even in the absence of Cas9 expression. A similar result was reported by López Del Amo et al., with Cas9 and gRNA carrying mothers having a detrimental effect on inheritance bias by their progeny while contribution of the gRNA from the father and Cas9 from the mother did not[114].

2.5 What strategies do we have to combat these challenges?

Achieving restricted nuclease expression

Restricting nuclease activity to cells and developmental stages where HDR is expected to be favoured and somatic tissues are unaffected is an optimisation strategy commonly pursued within the field. This is especially important in population suppression drive systems which rely heavily on the fitness of drive-carrying heterozygous (female) individuals[84;126;134–136]. To achieve this restricted activity, the field has largely relied on identifying and testing multiple genes which are predicted to have the desired expression/activity profile. The putative regulatory sequences of these genes are then isolated and used to express Cas9. This strategy has shown success such as the improvements achieved by *zpg* expressed Cas9 in *A. gambiae* compared to Cas9 expressed with *vasa*2, *nanos*, or *exu*[129]. This change of expression resulted in the reduction of both somatic drive activity and deposition of the nuclease while maintaining the inheritance biasing efficiency. While this trial-and-error approach has yielded improved HEGs in multiple species, insights into what underlies any improvement in performance are very limited as many changes are made simultaneously that cannot be deconvoluted.

The future design of HEGs may be aided by studying the effect of 'stacking' multiple limited regulatory mechanisms. In addition to the use of promoter/5'UTR and 3'UTR sequences, other endogenous regulation mechanisms can be included, such as tissue-specific splicing[137–139], modulation of protein degradation[140], sub-cellular localisation[141], and inclusion of miRNA binding sites[142]. Ideally, each regulatory system should make as limited and well-defined a change as possible. Decoupling of expression timing from expression levels may be a useful first candidate as the stacking of regulatory mechanisms may be expected to cause a cumulative decrease in activity levels due to imperfect removal of inhibition. Grunwald et al. demonstrated this principle with a HEG by using a strong and constitutive promoter to express Cas9 that could only be translated once a stop codon had been excised by separate and germline-restricted expression of a recombinase[93]. An important downside of Grunwald's approach was that once activated, the Cas9 could no longer be shut-off. A similar approach, but using a transcription factor intermediate such as the GAL4–UAS system would allow for reversible activation[143]. Such a system may particularly benefit 'integral' gene drives.

Integral gene drives make direct use of an endogenous gene for their expression[74]. Hoermann et al. demonstrated, in three different genomic loci of *A. gambiae*, that an artificial intron can be used to express both an effector protein and the host gene product[144]. A gRNA targeting the unmodified locus was also included, which allowed for efficient inheritance bias of the whole element with a separately expressed Cas9. Inclusion of the large Cas9 gene (>4Kb) in an integral gene drive context has been recently demonstrated in mice, where Cas9 was integrated at the very end of the

coding region of the *Spo11* gene[75]. This design resulted in Cas9 being co-translated with the endogenous gene, with the aim of restricting Cas9's activity to match that of *Spo11* which is involved in meiotic recombination. Using this approach, Weitzel et al. demonstrated for the first time homing in male mice despite earlier efforts with other regulatory systems[75;93]. However, the majority of target alleles remained uncut, indicating that improvements in regulating the timing of Cas9 activity came at the cost of its activity level. An intermediate amplifier of expression may address these expression issues and, if smaller than Cas9, interfere less with the endogenous gene's function (the endogenous *Spo11* gene was impaired by the Cas9 insertion).

Finally, there may be specific interventions that can address the effect of Cas9 deposition. In a study in *A. gambiae*, the I-*Ppo*I homing endonuclease was expressed by the testis-specific promoter *β-tubulin* to establish a synthetic sex ratio distortion system by shredding the X chromosome in the paternal germline[145]. However, no viable embryos were produced because paternally deposited I-*Ppo*I also shredded the maternally contributed X chromosome in the zygote. A subsequent study was carried out to reduce the half-life of the endonuclease by systematically introducing point mutations into the protein[146]. Strains with high levels (95-97.4%) of male-biased sex distortion and fertility rates similar to controls were eventually generated using this approach. Interestingly, the modified I-*Ppo*I was recently combined with a HEG drive system into a 'sex-distorter gene drive'[147]. When tested at three new loci, expressed with the identical *β-tubulin* promoter, male sterility was reestablished to varying degrees, presumably due to locus-dependent changes in expression of the I-*Ppo*I endonuclease, causing sufficient protein to persist into the embryo. By introducing a 100-bp GC-rich DNA sequences into positions -271, -244, and -355 upstream of the start codon, respectively, transcriptional activity of the *β-tubulin* promoter was reduced to 0.5%, 8.1%, and 16.2%. The promoter variant with 8.1% transcriptional activity, coupled with a destabilised I-*Ppo*I, was inserted into the *dsx* locus and was found to have no detectable sterility in drive heterozygous males. Similar approaches may also address deposition of Cas9 in a HEG context.

Multiplexing

The targeting of multiple sequences ('multiplexing') has been proposed as a means of addressing one of the most significant impediments to HEG drives - resistance[89]. If an initial attempt at homing fails and induces a mutation, multiplexing may still allow for homing through cleavage at an alternate cut site. Moreover, multiplexing would also allow the HEG to drive in individuals that have preexisting sequence variation in a subset of cut sites. Another benefit, specifically when targeting high-fitness cost genes, is that for complete resistance (resistance at all target sites) more extensive sequence changes would need to occur and this reduces the likelihood of the formation of type-1 resistance mutations. In terms of feasibility, multiplexing is particularly convenient with CRISPR-Cas9 nuclease as it only requires expressing additional gRNAs. Under these assumptions, computational modelling has indicated multiplexing to be an effective strategy to reduce the formation and accumulation of

resistant alleles[79;80;110;148;149]. However, some practical challenges have emerged with this 'classical' multiplexing approach.

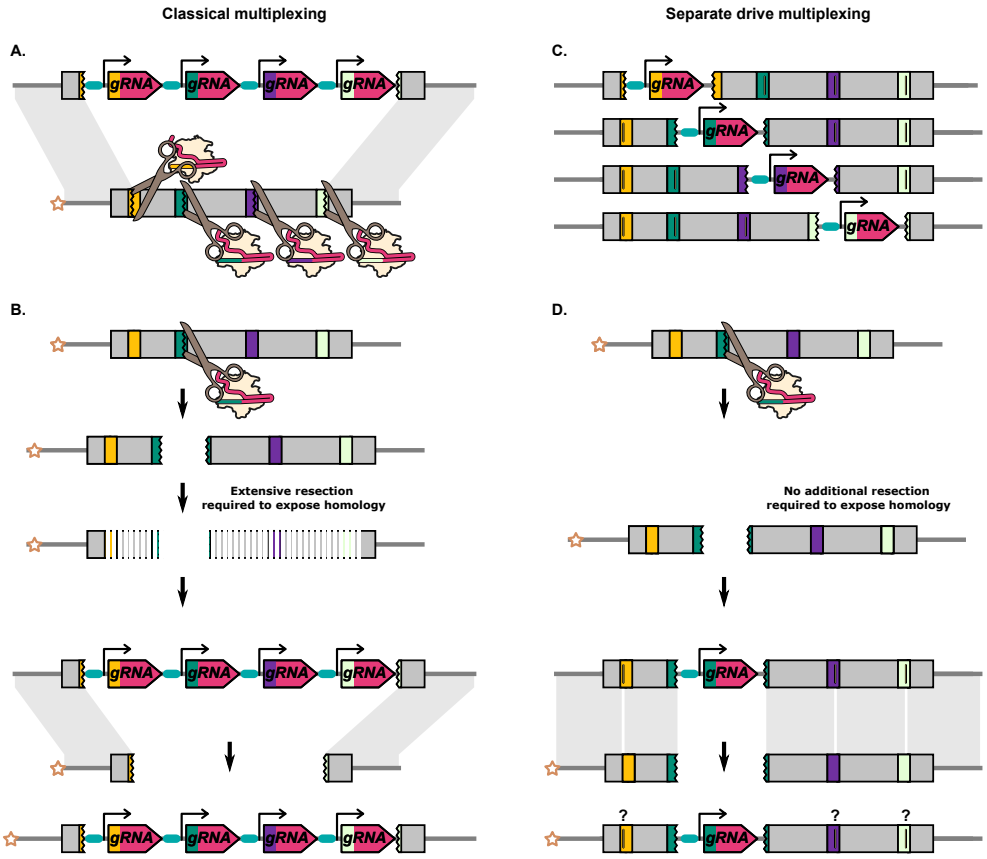


Figure 2.5: **Differences between classical multiplexing and separate drive multiplexing.** A. Classical multiplexing construct targeting adjacent sequences within a single gene. Four gRNAs are encoded from a single construct. B. When the outermost gRNAs do not cleave the recipient chromosome simultaneously, the cut caused by any one of the gRNAs will inevitably result in a region of extraneous sequence on the recipient chromosome which is 'unmatched' to the donor chromosome. As further 5'-3' and 3'-5' resections have to occur prior to HDR, this might reduce HDR efficiency and favour NHEJ. C. Four separate drive elements targeting adjacent sequences within a single gene. These are independent modifications of the same target gene and are not present in the same individual. Alternatively, separate drive constructs could be used to target multiple loci at distinct sites within the genome. D. The cut caused by the gRNA on the recipient chromosome is repaired by using the 'matched' homology arm on the donor chromosome as its template. The separate drive elements can include recoded target sequences for the other gRNAs to prevent cutting between elements. It is unclear if and how these types of sequence changes would affect homing efficiency. In A, one gRNA target is shown to be cut in the opposite orientation to the others. The different orientation is to do with the asymmetrical position of the cut site within the standard 20bp gRNA binding site (17bp//3bp). After a cut with a single gRNA, one end of DNA break will have at least 17bp of homology with the gRNA gene, and the other end will have at least 3bp. With a multiplex design, the outermost gRNAs can be oriented opposite to each other such that both DNA ends carry either the smaller or the larger region of homology to the gRNA gene. A DNA end with only 3bp of homology to the gRNA gene will presumably minimise the risk of partial homing/internal recombination. In contrast, if resection after a single cut (see B) is to be minimised, the opposite gRNA orientation may be desirable. For gRNA target sites that are not in the outermost position, there may be no overall 'optimal' orientation as it varies with the particular combination of gRNA targets that have been cut.

In the classical approach to multiplexing, multiple gRNAs targeting closely linked adjacent sequences of a single gene are expressed in a single drive transgene (Fig.2.5A). Additional sequences need to be removed or replaced on the donor chromosome to prevent the HEG from cutting itself at these additional sites. For a single target HEG, the ends of the cut site can be perfectly homologous to the donor chromosome. However, in a multiplexing system, any individual cut site can no longer generate two DNA strands that are perfectly homologous to the donor chromosome (Fig.2.5B). There are indications that these extraneous, 'unmatched' sequences could reduce the homing efficiency[81;110;114], presumably due to the additional resection that would not need to occur prior to HDR with a perfectly homologous repair template[4;150]. López Del Amo et al. introduced 20bp truncations in the homology arms either side of a *D. melanogaster* HEG[114]. These truncations result in 20bp of unpaired sequences on the recipient chromosome that would normally be homologous to the sequence directly adjacent to the HEG. The inheritance biasing rate of the HEG was significantly reduced with truncation on both sides of the HEG. Consistent with this, two HEGs, each with four gRNAs, targeting sites spread over a large region (>2Kb)[48] performed worse than similar drives with one gRNA[90] or two gRNAs targeting a smaller region[81]. Moreover, additional gRNAs may compete to complex with a limited amount of Cas9 protein lowering the cut-rate at any one site[110]. These results indicate that any individual cut site in classical multiplexing may be less efficient than a drive element optimised for only one cut site. Nonetheless, additional studies have indicated that multiplexing can increase the overall efficiency of a HEG, albeit in some cases with diminishing returns for additional gRNAs[81;110;151], and[152] compared to[153].

A novel challenge introduced by multiplexing is expressing different gRNAs simultaneously at similar concentrations without introducing repetitive sequences. Multiple strategies have been proposed to achieve this, many of which involve the excising of individual gRNA from a single long transcript such as with tRNAs[154;155]. However, these excising approaches are frequently not perfectly efficient or leave scars in the form of additional nucleotides attached to the gRNA that can reduce their activity. A separate approach that may be effective is for each gRNA using a set of different (minimal) promoters that have been characterised to have similar expression levels[156]. Using different experimentally validated 'backbones' for the non-targeting sequences of the gRNA may further reduce the likelihood of internal recombination[59].

Cas9 is known to remain bound to its target even after making a double-strand break[157], and DNA repair occurs more slowly than with other sources of DNA damage[158]. This may provide an opportunity for DNA breaks to occur at different sites before DNA repair is completed, leading to frequent deletions between independent target sites[158], as has been observed for some multiplexing HEGs[48;81;110]. In addition, CRISPR-Cas9 has been found to frequently induce large deletions (>250) at single cut sites[67;104;159]. Large deletions, or simultaneous cutting of at least the outermost target sites of a multiplex drive, could remove all unmatched sequences from a HEG recipient chromosome, potentially restoring homing efficiency to the level of a single target drive. However, this would also negate some of the benefits of multiplexing, as it may cause the simultaneous loss of all gRNA recognition sites on the chromosome.

Recently, a separate drive or multi-locus multiplex strategy has been proposed that may avoid some of the diminishing returns of classical multiplexing.

Multi-locus multiplexing consists of multiple 'parallel' single-target HEGs generated as separate lines, each targeting an adjacent site in the target gene (Figures 2.5C and 2.5D)[149]. Compared to the classical multiplexing strategy, the separate drive approach will be logistically more onerous but benefits from not being able to generate deletions of all target sites by simultaneous cutting. However, experimental validation of this approach has yet to be reported.

Targeting essential genes

The targeting of essential genes with HEGs has been reported in many studies, but depending on the goal of the system (i.e., suppression or modification), the approaches may diverge. For a suppression drive, the goal is generally to disrupt the essential gene in individuals of the target population. To achieve this, a functionally constrained sequence can be targeted such that few, if any, cut-resistant mutations generated will have a fitness advantage over the drive allele. In *A. gambiae*, an ultra-conserved region of a female-specific isoform of the *doublesex* gene was targeted to impede formation of resistant alleles[51]. HEG drives based on this target site in three cage trials were able to cause a complete population crash, and no type-1 resistance alleles were recovered[51;147;160]. A similar drive, targeting an ultra-conserved exon in a different gene did lead to the emergence of resistance to the HEG in the form of a single nucleotide silent mutation[83]. The target site to which no resistance emerged was located at an intron-exon junction, potentially making mutations liable to disrupt crucial mRNA secondary structures.

Targeting an essential gene can also be used for non-suppression drives as a general approach for lowering the viability of mutations ('home-and-rescue'/rescue HEG). By providing a rescue copy within the HEG drive construct (Fig.2.6A), resistance can be mitigated as the HEG drive will now have a fitness advantage over type-2 resistance alleles. The role of the rescue sequence is well illustrated by the first HEG drive developed in *A. stephensi*. This drive was inserted into an eye pigmentation gene, *kmo*[53], which was later found to have a recessive fitness cost in females[82]. This reduced fitness, coupled with a reduction in inheritance biasing efficiency in individuals experiencing maternal deposition caused the drive to fail to reach fixation when released at a 1:10 drive:wild-type ratio and even performed poorly at a 1:1 ratio[82]. A new version of the HEG was developed that included recoded parts of the *kmo* gene resulting in the drive allele no longer disrupting its function. A subsequent cage trial demonstrated this improved version of the HEG could effectively spread and reach fixation[161]. In addition to biasing its inheritance by cutting target alleles, the improved HEG could also increase in frequency by positive selection when the frequency of type-2 alleles accumulated in the population.

The most common approach for rescuing the function of the target gene is providing a 'recoded' version of the sequence the drive allele is disrupting (Fig.2.6B). Like type-1

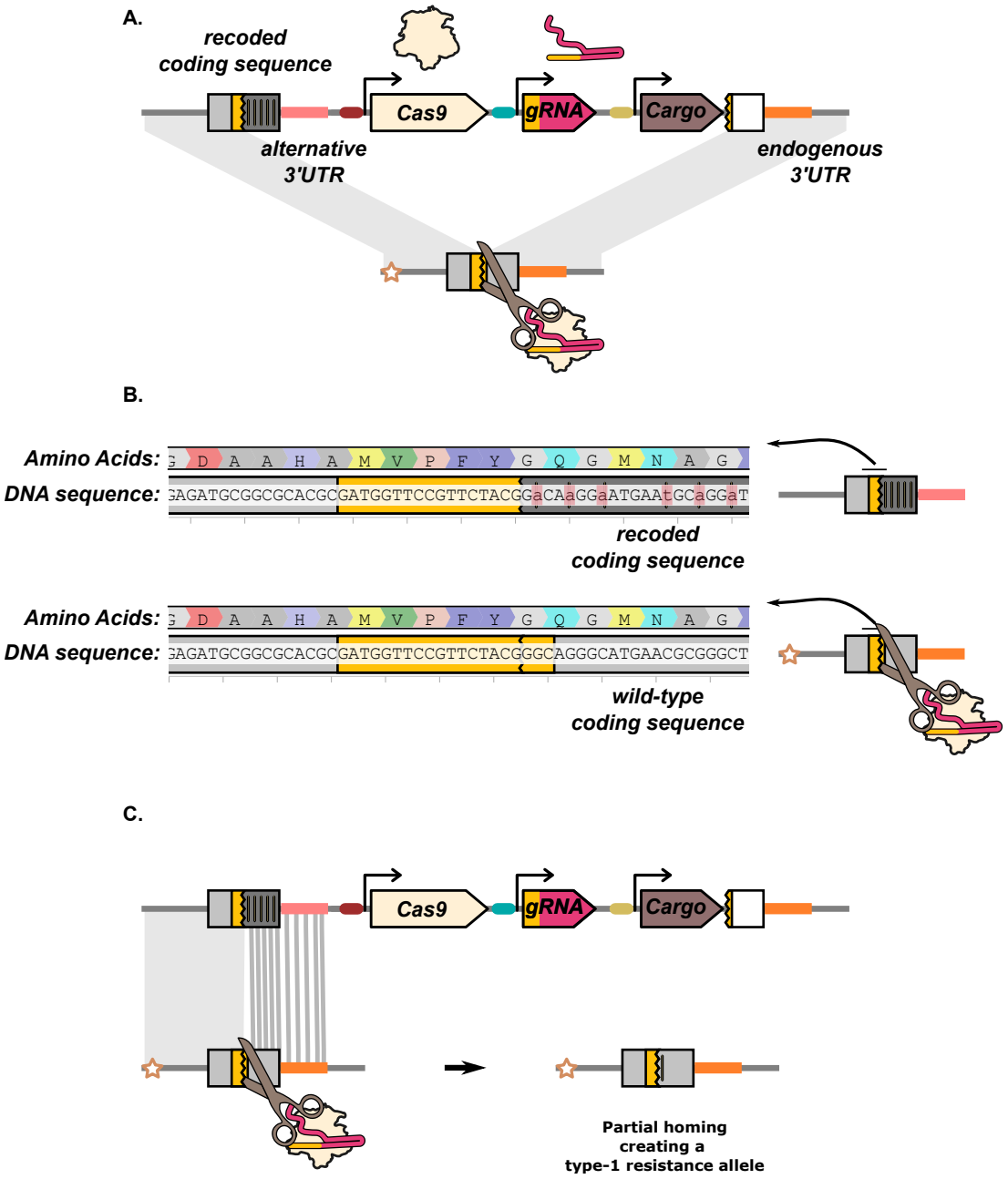


Figure 2.6: **Rescue HEG targeting an essential gene.** A. In addition to the nuclease components (Cas9 and gRNA), the drive element carries an effector/cargo gene and a 'rescue' sequence that restores the function of the gene the HEG is disrupting. The rescue sequence allows the drive element to be more fit than type-2 resistance alleles. The effector function of the rescue HEG is mediated by a separate cargo gene. In all other illustrations in this manuscript, the effector function of the drive is assumed to be mediated by disruption of the target gene and the HEG does not carry a rescue. B. In this example, the rescue is a recoded version of an endogenous gene. Synonymous codon changes have been made to prevent partial homing (and recognition by the gRNA). C. The recoded gRNA target sequence is located at the start of the rescue sequence. Partial homing by means of any part of the recoded sequence would generate a type-1 resistance allele. The recoding and gRNA target sequence is taken from [161].

resistance alleles, these are sequence changes (such as the swapping of synonymous codons) that prevent recognition by the gRNA but leave the target gene functionally intact. While in some cases a single nucleotide substitution can be sufficient to prevent Cas9:gRNA binding, much more extensive recoding is performed to reduce the risk of partial homing (Fig.2.6C). Recoding is further complicated by the need to include noncoding sequences such as the 3'UTR, for which no straightforward synonymous sequence substitution rules exist. In the case of the *kmo* targeting *A. stephensi* drive, the 3'UTR from the *A. gambiae kmo* gene was used in the rescue element[161].

The manner in which positive selection is conferred to the recoded rescue allele is similar to that expected in a Cas9-based toxin-antidote system where Cas9-induced mutations cause lethality or sterility allowing a cut-resistant rescue/antidote gene to spread in the population[61;162]. Moreover, deposition can increase the effective inheritance rate of the drive by culling individuals that have inherited a type-2 allele from their drive-carrying mother and a functional target allele from their father. Disruption of the paternally contributed target allele by maternally deposited nuclease can make these individuals no longer viable. Progeny that inherited the drive allele will be protected as they carry the (dominant acting) recoded antidote. However, these drive-inheriting individuals will likely have severely reduced homing rates. These rescue HEG systems have been reported in several studies targeting haplolethal (*RpL35A*[163]) or haplosufficient (*rab5*, *rab11*, *spo11*, *prosalpha2*, and *PolG2*[49;50] genes. All three of these studies showed increased inheritance of the rescue alleles to varying degrees and demonstrated this strategy to be successful in mitigating the negative effects of type-2 alleles to the drive system.

In the study by Champer et al., multiplexing (two gRNAs) was combined with targeting the haplolethal *RpL35A* gene[163]. When the target gene is haplolethal instead of haplosufficent, a single rescue gene may not be sufficient to protect from the lethal effects of deposition. This makes deposition a much more substantial hurdle for these types of systems. Shadow drive may theoretically be expected to create viable progeny by homing an inherited rescue element or a type-1 resistance allele. However, rescue through shadow drive may be very unlikely as deposition often results in mosaic outcomes. Depending on the target gene, any individual progeny may have a significant proportion of cells that have not been rescued by shadow drive and therefore nonetheless become inviable ('lethal mosaicism'). Similarly, with a haplosufficient target, individuals inheriting a maternally contributed type-2 resistance allele may not be rescued by mosaic type-1 alleles produced by deposition-induced cutting.

2.6 Concluding remarks

Synthetic homing endonuclease gene drives have been actively researched for over a decade. In this time, the field has developed and characterised a range of designs and applied these to a diverse set of species. Insects, and specifically *Drosophila* and the *Anopheles* mosquitoes, have so far proved substantially more amenable to HEG-mediated inheritance bias than other animals and plants. Additional work will

prove if the optimisation approaches developed in these insects can be successfully applied in these other species. This effort may be aided by a more systematic and high-throughput HEG test and design approach.

While efforts have been made to experimentally validate the intended transgene expression pattern[47;75;108;118], there are fewer cases where this has been done throughout development[164]. The result of this is that any hypothesis about the ideal expression/activity pattern for HEGs is currently essentially unfalsifiable as any exception to a proposed hypothesis can easily be explained away by the many ways a drive may fail to recapitulate the intended expression pattern, at the needed expression level. Improved methods to test different activity patterns (e.g., drug inducible Cas9[153;165]) and high-throughput methods to track HEG expression and nuclease activity (e.g., Cas9-based lineage tracing[166]) are sorely needed to validate our assumptions about the underlying factors influencing interchromosomal HDR. This becomes increasingly important as evidence emerges that interchromosomal HDR can occur before the formation of the mature germline[106;114;118;167].

Many approaches have been developed to control the expression and activity of transgenes. However, the use of modular systems such as GAL4–UAS to enforce a new HEG activity pattern will likely be more challenging than the current trial-and-error approach of identifying and testing new promoter/5'UTR and 3'UTR regulatory sequences. Yet, we expect that this type of modular approach will enable high-throughput design-build-test cycles. The modularity gained with such an approach will in turn improve our ability to draw conclusions about the underlying biology affecting HEG efficiency and increase the robustness of new designs going forward.

While multiplexing may have diminished returns in improving homing efficiency with standard approaches, it may nonetheless greatly diminish the likelihood of type-1 resistance alleles emerging. The targeting of highly conserved sequences in essential genes has proved beneficial for reducing the impact of resistance alleles. If approved, we expect that the current 'state-of-the-art' HEGs in *Anopheles* mosquitoes may progress on to field trials without substantial additional changes to their core design. If this is the case, the complexity of a real-world release will be the ultimate test of the HEG technology. Our constantly expanding genetic 'toolbox' and optimisation strategies provide hope that HEGs may be a highly effective tool for combating the harms caused by a broader set of medically and agriculturally relevant (homing refractory) insect pests.

This review has laid out the remaining challenges and open questions in the homing gene drive field. In the next chapters, we will investigate these through the use of computational modelling and analysis approaches. We will specifically revisit many of the qualitative relationships we have highlighted here, and provide quantitative measures of their importance. We will also critically evaluate the actual evidence base for assumed relationships and transgene design effects.

Conflict of Interest Statement

The authors declare that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

SV wrote the initial draft. JA wrote sections of the manuscript. All authors contributed to the manuscript revision, read, and approved the submitted version.

Funding

Biotechnology and Biological Sciences Research Council (BBSRC) supported S.A.N.V. (BB/M011224/1), L.A. (BBS/E/I/00007033, BBS/E/I/00007038, and BBS/E/I/00007039) and M.B.B. (BB/H01814X/1, BB/L00948X/1, BB/V008110/1). M.A.E.A., J.A., and L.A. were supported through an award from DARPA's Safe Genes program [N66001-17-2-4054]. The views, opinions, and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the U.S. Government. The funders had no role in study design, data collection and analysis, the decision to publish, or the preparation of the manuscript.

Chapter 3

Daisy-chain gene drives: the role of low cut-rate, resistance mutations, and maternal deposition

Sebald A. N. Verkuijl[1,2], Michelle A. E. Anderson[2,3], Luke Alphey[2,3], Michael B. Bonsall[1]

1: Mathematical Ecology Research Group, Department of Biology, University of Oxford, 11a Mansfield Road, Oxford, OX13SZ, U.K.
2: Arthropod Genetics, The Pirbright Institute, Ash Road, Pirbright GU24 0NF, U.K.
3: Current Address- The Department of Biology, University of York, York, U.K.

In the previous chapter, we reviewed the challenges facing the design of effective gene drives: unintended repair outcomes, nuclease activity in the wrong place and time, and deposition. This work has focused on data from self-perpetuating and split-drive systems. We have drawn attention to many relationships and phenomena that affect their performance. The question remains how they actually quantitatively affect gene drive performance on the population scale and on metrics that matter for real-world use. In addition, it is not clear how the challenges identified in the simpler gene drive designs affect the more complex implementations. We will evaluate this in this chapter and test how gene drive designs can be altered to reduce the impact of some of these issues.

3.1 Abstract

The introgression of genetic traits through gene drive may serve as a powerful and widely applicable method of biological control. However, for many applications, a self-perpetuating gene drive that can spread beyond the specific target population may be undesirable and preclude use. Daisy-chain gene drives have been proposed as a means of tuning the invasiveness of a gene drive, allowing it to spread efficiently into the target population, but be self-limiting beyond that. Daisy-chain gene drives are made up of multiple independent drive elements, where each element, except

one, biases the inheritance of another, forming a chain. Under ideal inheritance biasing conditions, the released drive elements remain linked in the same configuration, generating copies of most of their elements except for the last remaining link in the chain. Through mathematical modelling of populations connected by migration, we have evaluated the effect of resistance alleles, different fitness costs, reduction in the cut-rate, and maternal deposition on two alternative daisy-chain gene drive designs. We find that the self-limiting nature of daisy-chain gene drives makes their spread highly dependent on the efficiency and fidelity of the inheritance biasing mechanism. In particular, reductions in the cut-rate and the formation of non-lethal resistance alleles can cause drive elements to lose their linked configuration. This severely reduces the invasiveness of the drives and allows for phantom cutting, where an upstream drive element cuts a downstream target locus despite the corresponding drive element being absent, creating and biasing the inheritance of additional resistance alleles. This phantom cutting can be mitigated by an alternative indirect daisy-chain design. We further find that while dominant fitness costs and maternal deposition reduce daisy-chain invasiveness, if overcome with an increased release frequency, they can reduce the spread of the drive into a neighbouring population.

Author Summary

Reducing the harm of pest species by the introgression of traits into a wild population is often limited by the difficulties of mass rearing and release of modified individuals. Gene drives present an opportunity to substantially reduce the release frequencies required to spread a particular modification. However, uniform modification of a target species is, with a few specific exceptions, not necessary or desirable. Self-limiting gene drives, such as daisy-chain gene drives, have been widely discussed as a potential solution, allowing the invasiveness of a drive release to be tuned to the target population. Here, we investigate through computational modelling how daisy-chain gene drives perform when subjected to commonly observed inefficiencies associated with CRISPR-Cas9-based inheritance biasing. Compared to a self-perpetuating drive, daisy-chain gene drives are sensitive to factors that cause their separate elements to segregate prematurely. In particular, a reduction in the DNA cut-rate and an increase in the formation of resistance alleles. We find that the effect of inefficiencies in the drive mechanism is generally more pronounced when the drive is at low frequencies. With low rates of migration, this substantially reduces daisy-chain gene drives spread into a neighbouring non-target population.

3.2 Introduction

Synthetic gene drive methods are potentially powerful approaches in the management of agricultural pests and disease vectors. However, for many applications, the uniform modification of a species by a fully self-perpetuating drive is unnecessary and likely undesirable[55]. From an ecological, regulatory, and public acceptance perspective, a drive that is spatially and temporally limited may be more appropriate. Gene drives

can be inherently self-limiting, but this is generally due to low inheritance biasing efficiency, high fitness costs, or frequency-dependent effects, all of which can reduce their practical use. Some self-limiting systems have been proposed that are restricted in specific ways that maintain efficient but local transgene invasion[168]. These include targeting specific DNA sequences found only in subpopulations[58], dependence on supplementation with a synthetic inducer molecule[153;165], and split-drive systems.

In a split-drive system, at least one genetic component essential for the gene drive inheritance biasing mechanism is removed from the biased locus and is located elsewhere in the genome[60]. The split-drive will bias, or drive, a primary locus; however, those copies cannot bias themselves in subsequent generations without the component(s) from the unbiased (non-driving) secondary locus. As the split-drive spreads into a large population, the lack of amplification of the secondary locus will limit the rate of further amplification of the primary drive element. Compared to a cargo-only release (inundative release), split-drives provide a modest reduction in the number of modified individuals that need to be released to spread a trait among a wild population. Researchers have sought to develop variations on the basic split-drive design with increased invasiveness, but that can still be tuned to specific target populations and applications.

For a meiotic drive system that biases chromosomal regions proportional to their recombination distance, the split-drive elements can be located an intermediate recombination distance from each other. This intermediate linkage causes most, but crucially not all, progeny that inherit the main element to also inherit the secondary element, increasing invasiveness, tunable by the recombination distance[169]. However, the most extensively studied gene drives are homing endonuclease gene drives (HEGs) that, with poorly understood exceptions[1], only bias a very limited genomic region. Therefore, it is unlikely that the intermediate inheritance biasing of a separate split-drive element can be incorporated in the same way. For HEGs, other split-drive strategies have been proposed that would allow the spread to be tuned (and are also applicable to a meiotic drive system); one of these is the daisy-chain gene drive[59].

Daisy-chain gene drives (DCDs) sequentially link drive elements to prolong inheritance bias on a main effector modification, which is the final element in the chain. Each element biases the inheritance of another element apart from the final element, and each element is biased except for the first element. Sequentially reduced copies (reduced by one element) of the released drive configuration will continue to bias an effector modification until copies are produced where it is the only element remaining. Increasing the number of elements in the chain prolongs the generations of inheritance bias experienced by the main effector modification. DCDs are a seemingly safer alternative to most applications for which a self-perpetuating drive would be considered. However, the increased complexity of their design raises questions about the efficacy of the molecular machinery required for practical implementation.

Engineered HEGs with varying inheritance biasing efficiencies have been reported in various species[53;60;68;85;93;95;97;170], most commonly with the CRISPR-Cas9 nuclease. CRISPR-Cas9 is a two-component endonuclease formed by the Cas9 protein and a

guide RNA (gRNA) that specifies the sequence to be cut[65]. CRISPR-Cas9-based homing split-drives can be engineered by moving either the gRNA gene(s) or the Cas9 nuclease gene, as both are essential for a cut to occur. DNA breaks that Cas9 makes can, under specific circumstances, induce cellular repair pathways to copy a gene drive from one homologous chromosome to another (homing). This results in the biased genetic element converting itself from a heterozygous state to a homozygous state.

There are several ways in which the HEG inheritance biasing mechanism may not work as intended. The most straightforward is a less-than-saturating cut-rate of the target allele. A common optimisation strategy for HEGs is to test endogenous regulatory sequences that limit nuclease expression to meiotic cells, which are expected to perform homing more readily after a DNA break. There may be cases where regulatory elements with the best expression window may not have the optimal expression level. An example of this is the report that in mice, the cotranslation of Cas9 with an endogenous gene resulted in a more favourable DNA repair profile compared to previous efforts with an autonomously expressed Cas9[93;94]; however, this improved DNA repair profile came at the cost of the overall cut-rate[75]. When fitness costs are low, self-perpetuating drives are generally robust to lower cut-rates as they can simply try again in a subsequent generation[171]. However, for DCDs, their self-limiting mechanism continues to operate, and the unbiased drive elements can be expected to randomly segregate, leading to only a subset of progeny co-inheriting all the drive elements. As such, even if the overall allele frequency in the population remains unchanged, there will be less co-occurrence of the multiple drive elements necessary for inheritance bias. Moreover, for DCDs, individuals may emerge with one or more upstream elements without the corresponding downstream drive element. Certain DCD designs will continue to cut and bias their downstream locus even when the associated drive element is not present to serve as the repair template on the homologous chromosome (phantom cutting).

The inheritance biasing of HEGs relies on homology-directed repair (HDR) to mediate copying of the drive element(s). However, cells have multiple competing pathways that can resolve a DNA break. Sequence changes made by alternative DNA repair pathways (and specific undesired HDR events) can be grouped by their consequences into two categories: type-1 and type-2 resistance mutations[90]. Type-1 resistance mutations (r1) are changes in DNA sequence that disrupt the gRNA binding site while preserving the normal function of the target gene (silent mutations). Type-2 resistance mutations (r2) also prevent the nuclease from cutting the site and, in addition, disrupt the normal function of the target gene (nonsense/missense mutations). Generally, target genes and specific target sites within those genes are chosen such that type-2 resistance alleles impose the same or higher fitness costs than the drive element. Type-2 resistance alleles will be selected against and will not prevent the drive element from approaching fixation. However, type-1 resistance mutations are expected to impose no or substantially lower fitness costs than the drive element. Even if type-1 resistance alleles are produced only very rarely, positive selection can increase their frequency and prevent the drive element from reaching the frequency needed for its intended function[92]. The segregation of drive elements with downstream resistance

alleles and additional biasing of these resistance alleles presents a unique challenge to DCDs that has not been investigated.

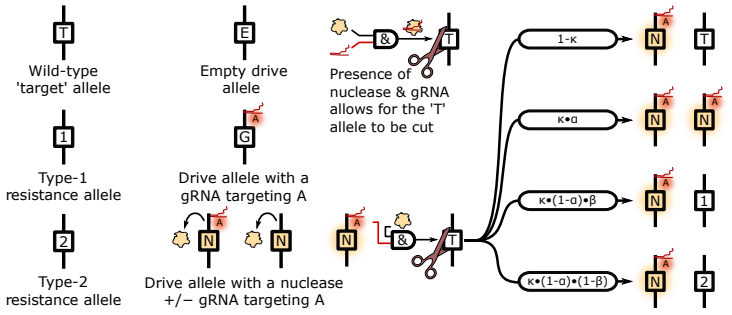
Lastly, another common HEG attribute is deposition, which may also be expected to have unique interactions with DCDs. Deposition is the phenomenon in which drive components expressed or translated in a parent can persist into their offspring. This can occur even if the offspring does not inherit the gene from which those components were expressed in the parent. Deposition occurs almost exclusively from female drive carriers (maternal deposition), with the expression pattern of the nuclease found to influence the degree and pattern of deposition observed[3;48–50;53;82;90;92;94;95;97;114;118;129;161;163;172]. Deposition can result in nuclease activity at times when the inheritance biasing process is not favoured and in individuals who do not carry the drive, generating additional resistance mutations. Furthermore, nuclease activity early in development can cause the conversion of somatic tissues creating (mosaic) individuals with additional fitness costs. In contrast, if those genotypes with additional fitness costs are generated in the germline by autonomous expression of the drive, these costs generally only manifest in the next generation.

Although some modelling has been done of DCDs[57;59;173;174], limited research has been done on understanding their interaction with functional resistance alleles, recessive and dominant fitness costs, reduced cut-rates, and maternal deposition. Here, we investigate, through theoretical modelling, the performance of two alternative homing endonuclease DCD designs when subjected to the common drive imperfections introduced above. In each case, we compare their performance to a self-perpetuating single-element drive allowing us to highlight unique interactions that affect DCDs. Furthermore, we evaluate the degree to which DCDs remain limited to the target population and do not spread substantially to a secondary population linked by migration. We find that DCDs place much higher requirements on the fidelity and efficiency of the drive mechanism for their spread. In addition, we find that phantom cutting can severely increase the production of type-1 resistance alleles, but this can be partially mitigated by using an indirect DCD design.

3.3 Results

3.3.1 Phantom cutting by DCDs can bias type-1 resistance alleles

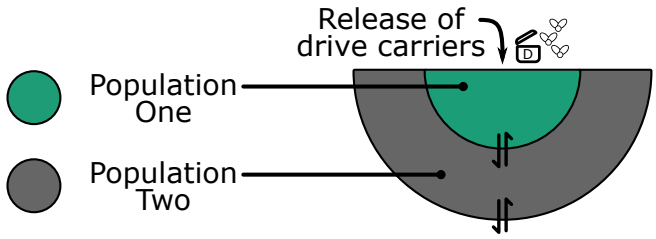
We simulate homing endonuclease gene drives in two populations with up to three unlinked diploid loci designated by a subscript: $_{A}$, $_{B}$, or $_{C}$. Briefly, a wild-type target allele (T) can be cut by the combined product of a nuclease allele (N) and a gRNA. The nuclease is target independent, whereas the gRNA always matches a specific wild-type target allele. The presence of a gRNA gene in an allele is indicated by a superscript of the locus to which it is targeted and can be together with a nuclease gene (N^{A}) or only a gRNA gene without the nuclease gene (G^{A}). A drive element can also lack a Cas9 or gRNA gene (E), being empty apart from the cargo gene/effector modification



(a) (b)

Parameters	Default values
Loci/Elements	$_A/_B/_C$
Cut-rate (κ)	100%
HDR rate (α)	95%
A locus type-1:type-2 ratio ($_A\beta$)	1:9
B locus type-1:type-2 ratio ($_B\beta$)	1:0
Fitness $_A$T / $_B$T	100%
Fitness $_A$D ($_AN^A/_AE/_AG^A$)	90%
Fitness $_B$D ($_AN^A/_BN$)	95%
Fitness $_C$D ($_CG^B$)	95%
Fitness $_A$1 / $_B$1	100%
Fitness $_A$2	0%

(c)



(d)

Parameters	Default values
Generations	t, Discrete
Simulation variability	Deterministic
Simulation length	100 generations
Release type	Single-sex
Release zygosity	Homozygotes
DCD at release	25% of males at t=1
SPD at release	0.25% of males at t=1
Release site	Population one
Tracked populations	Two
Migration	1% per generation

(e)

Figure 3.1: **Drive nomenclature system and overview of simulation parameters. (a)** Different alleles considered in our computational simulations: Wild-type target (T), type-1 resistance (1), type-2 resistance (2), and drive alleles (E, G, and N: collectively D). **(b)** When a target allele, a gRNA targeting that specific allele, and the nuclease are present in the same genotype, there is a certain likelihood that a cut occurs (by default 100%). The cut allele can be repaired by interhomolog HDR, or to a type-1 resistance allele or type-2 resistance allele. **(c)** Default nuclease and fitness parameters. The A drive allele has a fitness of 90% to account for general fitness costs (e.g., metabolic) and imperfect rescue of the haploinsufficient target gene by the drive element. This is the only fitness cost of the self-perpetuating gene drive. For DCDs, the B and C drive alleles have a fitness of 95%. Each fitness cost is dominant, with no additional recessive cost unless specifically listed for that simulation. Individuals carrying one or more of each different DCD allele are therefore assumed to have a fitness of 81% (90%·95%·95%). Type-2 resistance alleles at the A locus are assumed to be dominant lethal (fitness 0%). The B element is in a neutral locus; as such, its resistance alleles are all neutral and classified as type-1. The C locus is never cut, and resistance alleles are not generated at that locus. **(d)** Illustration of the two linked populations for which we report the allele frequency. Drive individuals are released into population one at the start of the simulation. Each generation, 1% of the surviving adults migrate to the neighbouring population. Population one receives 1% of population two, while population two receives 0.5% of population one and 0.5% of exclusively wild-types. **(e)** Default release frequency and simulation length parameters. DCD: Daisy-Chain Drive. SPD: Self-Perpetuating Drive.

that is being spread. The nomenclature system we use is further expanded in Fig 3.1a.

Nuclease deposition into the embryo, differences in fitness, and nuclease expression in the germline can cause a change in the genotype frequency from one discrete generation to the next (Figs 3.1b and S3.1). The fitness and drive conversion rates for the many unique combinations of alleles are extrapolated from a small set of parameters listed in Fig 3.1c. The inheritance biasing parameters most closely resemble those of HEGs developed in the *Anopheles* mosquito[129;161]. However, we assume that type-2 resistance alleles are dominant lethal at the A locus (haplolethal target gene) and the drive

elements at that locus carry a rescue element[60;161;163]. The creation of a rescue element is generally achieved by including a recoded version (many synonymous codon swaps) of the target site and downstream portions of the target gene. Essentially, engineering the drive allele to carry extensive type-1 resistance mutations. As such, we assume that a relatively high fraction of non-HDR repair results in type-1 resistance alleles (10% of 5%) when compared to what has been achieved targeting a highly conserved locus (for which creating a viable rescue element may not be possible)[160].

For each simulation, we report the zygote genotype frequency before fitness, deposition, and expression-mediated changes in genotype frequency occur in that generation. We track genotype frequencies in two separate populations which have a 1% migration rate, with population two receiving a constant migration of wildtypes from the general population in addition to migration to and from population one (Fig 3.1d). Alleles for a particular population are indicated with an additional number in the subscript (e.g., $_{A1}$). Unless otherwise specified, our default release scenario for daisy-chain drives is such that at the start of the simulation, 25% of males (12.5% overall allele frequency) in population one are homozygous for all drive elements (Fig 3.1e). The self-perpetuating drive is initiated with a frequency of male drive carriers 100 times lower (0.25%) compared to the release scenario for DCDs.

The three different HEG designs that we investigate are a single-element self-perpetuating drive ($_{A}N^{A}$) shown in Fig 3.2a and two different three-locus daisy-chain drives. For each drive, the A drive element is inserted into and rescues a haploinsufficient endogenous gene. For DCDs, the B and C elements are located at neutral sites. The first DCD design we call a direct DCD ($_{A}E/_{B}N^{A}/_{C}G^{B}$), shown in Fig 3.2b and is the most similar to the design modelled by Noble et al.[59]. The C element ($_{C}G^{B}$) carries a gRNA targeting $_{B}T$. The B element ($_{B}N^{A}$) carries both a nuclease and a gRNA targeting $_{A}T$. The drive element for A would, in practice, contain the desired effector modification, but does not contain any components (gRNA or nuclease) that are directly involved in inheritance biasing (denoted as $_{A}E$ without a superscript).

The second DCD design we consider is an indirect DCD ($_{A}G^{A}/_{B}N/_{C}G^{B}$), shown in Fig 3.2c, which has the gRNA targeting the A element expressed from the A element itself ($_{A}G^{A}$). The chain structure is maintained as the A element will only function in the presence of the nuclease expressed from the B element ($_{B}N$). The C element is identical in both DCD designs. The indirect DCD has several practical benefits, including that A and B element transgenic lines can be developed independently, while for the direct DCD design, the A targeting gRNA(s) located on B limits interoperability. A DCD design functionally similar to this (using orthogonal nucleases) was proposed in Noble 2019 et al., but not explicitly investigated[59].

The allele dynamics of the self-perpetuating drive ($_{A}N^{A}$) shown in Fig 3.2d represents the maximally efficient implementation of an HEG, albeit with unlimited spread. The A drive element (the only drive element) in population one ($_{A1}D$) reaches a maximum allele frequency of 99.4% at generation 23. The fitter type-1 alleles slowly outcompete the drive allele, but this process will take many generations, since at generation 23 the type-1 resistance allele ($_{A1}1$) frequency is only 0.6%. The drive

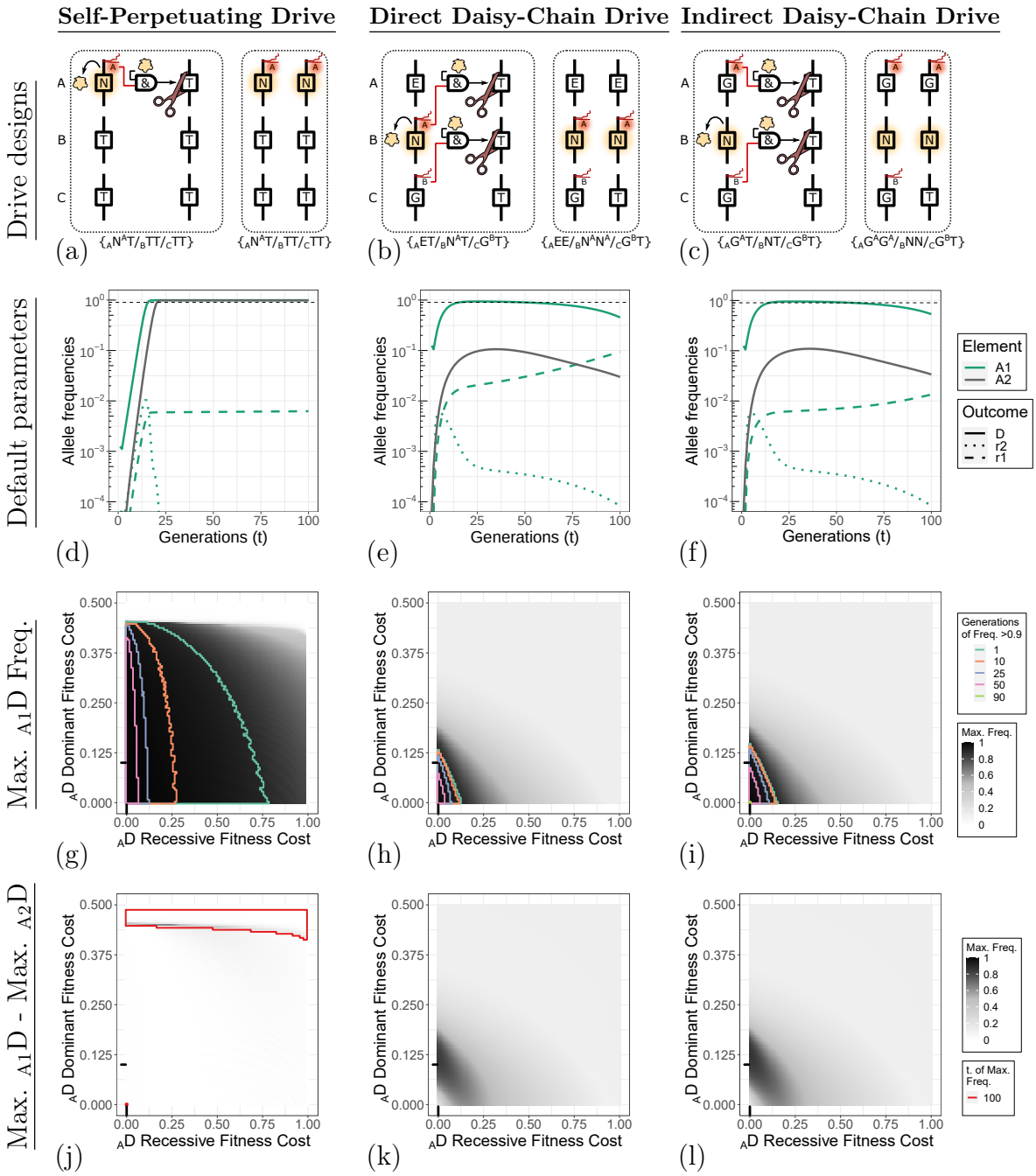


Figure 3.2: **Illustration of drive designs, default behaviour, and performance when subjected to different recessive and dominant $_{A}$D fitness costs. Column 1.** Self-Perpetuating Drive. **Column 2.** Direct Daisy-Chain Drive. **Column 3.** Indirect Daisy-Chain Drive. **Row 1 (a-c).** Illustration of the molecular designs of the three drives we simulate. Each design is shown as heterozygote, with on the right the product resulting from inheritance bias. **Row 2 (d-f).** A locus allele dynamics under the default conditions. The thin dashed line indicates a frequency of 90%. An allele at locus A in population one is indicated by $_{A1}$ (green lines), and a drive allele at locus A in population two is indicated by $_{A2}$ (grey line). Type 1 and 2 resistance alleles are not shown for population two in these panels. The allele dynamics for all loci are shown in S3.2 Fig. **Row 3 (g-i).** Parameter sweep for different recessive and dominant fitness costs applied to the $_{A}$D ($_{A}N^{A}/_{A}E/_{A}G^{A}$) allele. The shading of the heat map indicates the maximum $_{A}$D allele frequency in population one reached within the 100 generations simulated. The maximum frequency of $_{A}1$ alleles and outcomes for population two are shown in S3.7 Fig. An immediate decline after release results in the maximum frequency being the release frequency. Regions are boxed with coloured lines that indicate threshold values for the number of generations in which the $_{A1}$D allele frequency was greater than 90%. The black lines on each axis indicate the default value for the parameter being varied. **Row 4 (j-l).** Difference in the maximum $_{A}$D allele frequency between population one and two. Regions boxed with a red line are simulations in which the maximum allele frequency of population one or two was at generation 100 and the outcomes may change if additional generations are simulated.

element in population two reaches a maximum frequency of 98.8% at generation 24. While the initial spread of the drive element into population two lags a few generations behind that of population one, migration means that population one only reaches its maximum drive frequency when population two is also (almost) at its maximum drive frequency. The frequency of wild-type alleles in population two quickly drops and stabilises at just over 0.55%. This is due to the 0.5% migration of wild-type individuals from the general population each generation and the reduced fitness of drive-carrying individuals (90% of wildtype).

While the allele frequency dynamics of the B and C alleles are almost identical for the two DCD designs (S3.2 Fig), the A allele dynamics, especially the type-1 resistance allele dynamics, differs substantially (Fig 3.2e and 3.2f). The direct DCD reaches a maximum ${}_{A1}$E frequency of 93.8% at generation 25, which coincides with a type-1 resistance allele frequency of 2.1%. The maximum ${}_{A2}$E frequency in population two is 10.6% at generation 35. The indirect DCD reaches a maximum ${}_{A1}G^{A}$ frequency of 95.3% at generation 27, which coincides with a type-1 resistance allele frequency of 0.6%. The maximum ${}_{A2}G^{A}$ frequency in population two is 11% at generation 36.

The difference in performance between the two DCD designs occurs due to a phenomenon we term phantom cutting, illustrated in S3.3 Fig. This occurs when an upstream element, such as the direct DCD B element(${}_{B}N^{A}$), attempts to bias the inheritance of the downstream A element, even when the downstream element is not present. This is possible in our simulation as with a homing rate less than 100% the B drive element is occasionally co-inherited with a resistance allele generated at the A locus. Phantom cutting is only possible for the direct DCD as both the A gRNA and the nuclease are on B (${}_{B}N^{A}$) and it does not occur with the indirect DCD, as the nuclease on the B element (${}_{B}$N) requires the gRNA expressed from the A element drive allele (${}_{A}G^{A}$).

We have assumed an HDR rate of 95% and a type-1 to type-2 resistance allele repair ratio of 1:9 at the A locus, resulting in 0.5% of DNA breaks generating type-1 alleles. In the indirect design, these type-1 resistance alleles persist and slowly outcompete the drive alleles but are not otherwise biased in any way. In the direct DCD, these type-1 resistance alleles can be biased in the same way that the A drive allele can (e.g., ${}_{A}1T/{}_{B}N^{A}T \rightarrow {}_{A}11/{}_{B}N^{A}T$), and therefore reach much higher frequencies. In our simulations, type-2 resistance alleles at the A locus cannot be biased by phantom cutting as even one copy is dominant lethal. Phantom cutting can only occur when there are ${}_{A}$T alleles remaining in the population. As the A drive and resistance alleles replace the ${}_{A}$T alleles, the frequency of phantom cutting events fades away.

The default single-sex drive frequency at release of 25% allows the daisy-chains to spread to high frequencies in population one, but not in population two. As such, the release frequency can be considered to be roughly tuned to the drive's inheritance biasing efficiency, fidelity, and fitness. The self-perpetuating drive cannot be tuned and will spread to high frequencies in both populations. The A locus drive and type-1 allele frequency dynamics are shown for different HDR and migration rates in S3.4 and S3.5 Figs, respectively.

It is important to note that the differences between the two DCD designs depend on the assumptions we have made compared to previous work. Using the parameters and assumptions of Noble 2019 et al.[59] (100% cut-rate, no type-1 alleles, and type-2 alleles at A are dominant lethal), the direct and indirect DCD designs behave identically to each other (S3.6 Fig, columns one and two). Phantom cutting cannot occur because no progeny survive in which the B element segregates away from the A element. Moreover, the DCD designs that we evaluate differ from other models of DCDs[57;59] in that they only have a nuclease expressed by the B element, not by both the B and C element (i.e., ${}_{\mathrm{A}}\mathrm{E}/{}_{\mathrm{B}}\mathrm{N}^{\mathrm{A}}/{}_{\mathrm{C}}\mathrm{N}^{\mathrm{B}}$). With a nuclease on the C element, the direct design simulations match those reported by Noble et al. (S3.6 Fig, columns three and four). For the indirect DCD design, the nuclease expressed from the C element helps increase the spread of ${}_{\mathrm{A}}\mathrm{G}^{\mathrm{A}}$, but also generates many more resistance alleles at the B locus due to phantom cutting. All other simulations in this study use DCD designs without a nuclease on the C element.

Fig 3.2g-3.2i show the maximum frequency of the A drive element, and the number of generations it remains above a 90% allele frequency under different recessive and dominant ${}_{\mathrm{A}}\mathrm{D}$ fitness costs. The recessive fitness costs apply to individuals with two copies of the drive, the dominant fitness costs apply equally to individuals with one or two copies. Obviously, minimisation of the fitness costs is the most conducive to DCD spread. However, with a constant release frequency, this also allows the drive allele to spread beyond population one and reach high frequencies in population two (S3.7 Fig). In Fig 3.2j-3.2l we contrast the two populations by subtracting the maximum drive frequency in population two from the maximum drive frequency in population one. This reveals a more narrow parameter space in which the DCDs can spread to high frequencies in the target population, yet do not substantially invade population two. We have highlighted cases where the maximum drive frequency in either population coincides with the end of the simulation, indicating that simulating more generations may change the outcomes.

With a fixed release frequency, dominant fitness costs appear to be more desirable when evaluating the differential spread in the two populations. Unlike recessive fitness costs, dominant fitness costs do not additionally penalise the drive for spreading to a high frequency in population one and immediately apply to the drive alleles that have migrated to population two, even when they are at low frequencies. Furthermore, dominant fitness costs can apply to heterozygous individuals carrying a type-1 resistance allele. Individuals with a type-1 allele are never affected by a drive with an exclusively recessive fitness cost, increasing the relative benefit for an individual carrying such a functional resistance allele.

The relatively low default fitness costs of the A drive rescue element make these simulations representative of low fitness cost population modification drives (e.g., spreading an antibody gene in a vector of human disease). Another class of drives is aimed at population suppression, with the goal of reducing a target species's population density. We have performed simulations at different release frequencies for a recessive lethal (S3.8 Fig), and a sex-specific recessive lethal (S3.9 Fig) A element target gene. In

both cases, the A element does not contain a rescue gene and imposes the same fitness cost as a type-2 resistance allele (i.e., disrupts the target gene as it homes). As may be expected, the self-perpetuating drive approaches fixation under the broadest set of conditions, largely independent of the release frequency. All drives rapidly decrease in frequency after reaching their maximum, with type-1 resistance alleles driven to near-fixation by the self-perpetuating drive (type-1 alleles would reach fixation were it not for the low levels of migration of wildtypes from the general population). The DCDs only approach a maximum drive frequencies similar to the self-perpetuating drive with upward of 70% of males carrying the DCD at the start of the simulation. In all cases, the indirect design is burdened by fewer type-1 mutations and spreads to higher frequencies.

Together, these results show that even low rates of type-1 resistance allele generation cause the two DCDs to perform differently from each other. Furthermore, the maximum frequency achieved by a DCD is highly dependent on its fitness profile, with dominant fitness costs providing the largest difference between the spread in population one and two.

3.3.2 DCDs are severely impacted by reductions in cut-rate

Like the generation of type-1 resistance alleles, Cas9:gRNA cut-rates that are less than saturating (<100%) have been reported for multiple HEGs[75;93;94;103;118;129;172]. In Fig 3.3 we show A allele dynamics of simulations with a cut-rate of 85%, with the allele dynamics of all loci shown in S3.10 Fig. The spread of the self-perpetuating drive is delayed, but otherwise largely unaffected by a lowering of the cut-rate as each element can simply try again in the next generation (Fig 3.3a). With a cut-rate of 85% the two DCDs reach maximum A element frequencies of 79.7% and 92% for the direct DCD and indirect DCD, respectively (Fig 3.3b and 3.3c). Interestingly, despite the lowered cut-rate, type-1 alleles are still produced in large numbers with the direct DCD design, even exceeding the drive allele frequency by the end of the simulation. Simulations of a range of cut-rates are presented in Figs 3.3d-3.3f and S3.11.

With a reduction in the cut-rate the DCD B element can now be passed along together with an uncut ${}_{\mathrm{A}}\mathrm{T}$ allele (as opposed to necessarily a drive element or a resistance allele when the cut-rate is 100%). This expands the possible genotypes under which phantom cutting can occur to include genotypes like ${}_{\mathrm{A}}\mathrm{TT}/{}_{\mathrm{B}}\mathrm{N}^{\mathrm{A}}\mathrm{T}$ and creates the possibility of simultaneous cleavage of homozygous target alleles at the A locus. This complicates the nuclease conversion scheme as HDR could, in principle, recursively regenerate a T allele from the homologous T allele, opening up the opportunity of repeated and simultaneous cutting events. The simulation presented in Fig 3.3 were performed under the assumption that HDR does not occur when a TT genotype is cleaved, instead, additional resistance alleles are generated. In S3.12 Fig we present simulations with various cut-rates assuming HDR is possible, and individual T alleles in a TT genotype are cut sequentially. Regardless of the approach taken, the indirect DCD is more invasive than direct DCD at equivalent cut-rates and produces fewer type-1 resistance alleles.

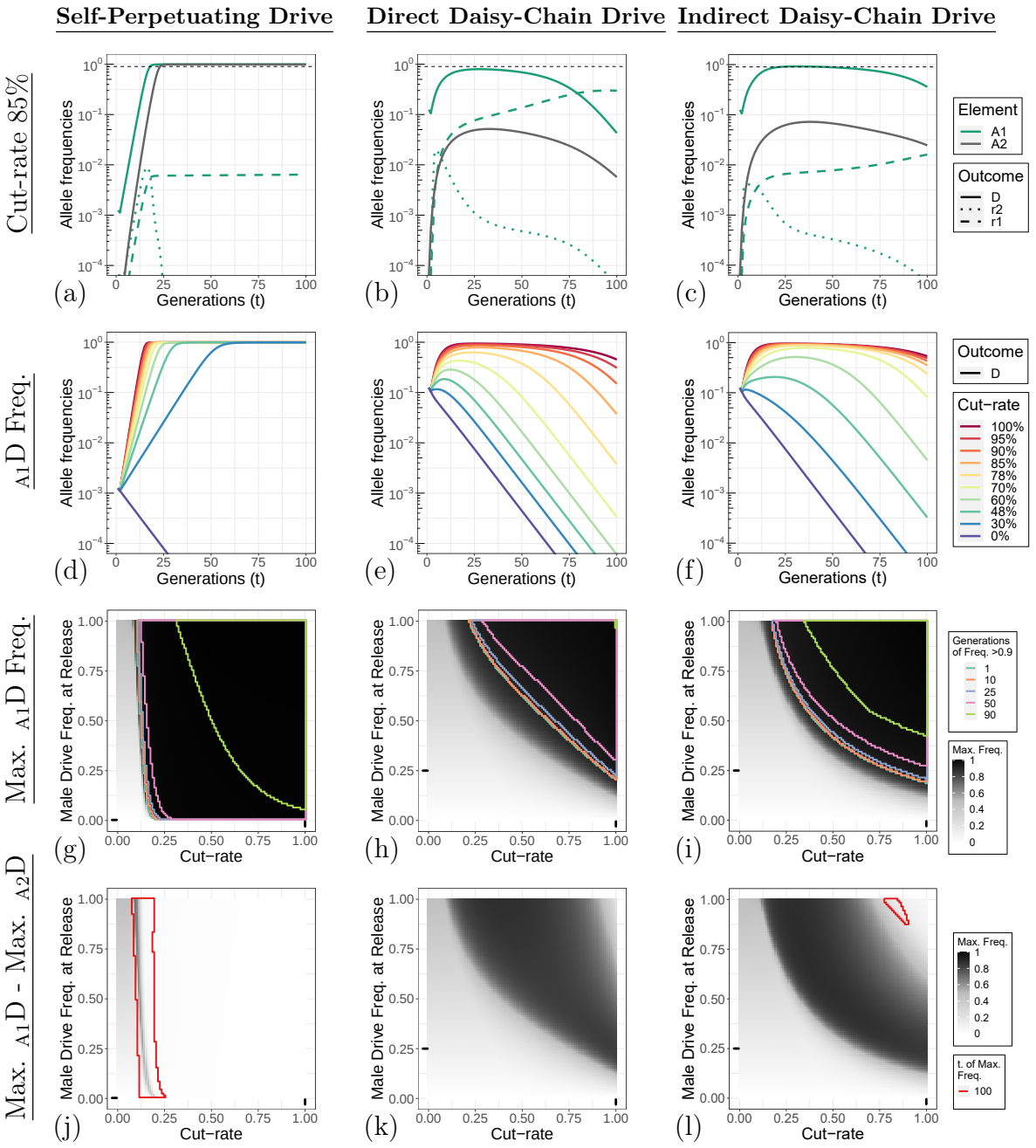


Figure 3.3: **Daisy-chain gene drive element prematurely segregate with reduced cut-rates. Column 1.** Self-Perpetuating Drive. **Column 2.** Direct Daisy-Chain Drive. **Column 3.** Indirect Daisy-Chain Drive. **Row 1 (a-c).** Individual allele dynamics with a cut-rate of 85%. The thin dashed line indicates a frequency of 90%. An allele at locus A in population one is indicated by $_{A1}$ (green lines), and a drive allele at locus A in population two is indicated by $_{A2}$ (grey line). Type 1 and 2 resistance alleles are not shown for population two in these panels. The allele dynamics for all loci are shown in S3.10 Fig. **Row 2 (d-f).** A element drive allele dynamics for population one with different cut-rates. Allele dynamics of the A locus drive allele in population two, and type-1 resistance alleles for both populations are shown in S3.11 Fig. **Row 3 (g-i).** Parameter sweep for the male drive frequency at release and germline cut-rate. The shading of the heat map indicates the maximum $_{A}$D allele frequency in population one reached within the 100 generations simulated. The maximum frequency of $_{A}$1 alleles and outcomes for population two are shown in S3.13 Fig. An immediate decline after release results in the maximum frequency being the release frequency. Regions are boxed with coloured lines that indicate threshold values for the number of generations in which the $_{A1}$D allele frequency was greater than 90%. The black lines on each axis indicate the default value for the parameter being varied. **Row 4 (j-l).** Difference in the maximum $_{A}$D allele frequency between population one and two. Regions boxed with a red line are simulations in which the maximum allele frequency of population one or two was at generation 100 and the outcomes may change if additional generations are simulated.

The most straightforward way to compensate for a reduction in invasiveness may be to increase the drive frequency at release. However, for the self-perpetuating drive, this has a very limited effect when the reduction in invasiveness is due to an extremely low cut-rate. Under these conditions (cut-rates lower than≈15%) the drive element remains unable to overcome its fitness costs independent of the drive frequency at release (Fig 3.3g). In contrast, any reduction in cut-rate affects the DCD much earlier through the loss of co-inheritance of the different drive elements (Fig 3.3h-3.3i). For these conditions, an increased release frequency can allow DCDs to reach frequencies where orphaned drive elements are likely to be reacquainted by random mating. For DCDs, the frequency-dependent consequence of a reduction in cut-rate limits the DCDs ability to invade population two when the cut-rate is reduced (S3.13 Fig). This frequency-dependent effect does not apply to the self-perpetuating drive as only a single-element needs to be inherited to affect a cutting reaction at the target locus. Fig 3.3j-3.3l shows this consequently provides a broad range of cut-rate and release frequency parameter combinations under which the DCDs spread to high frequencies in population one, but not population two.

The production of type-1 resistance alleles through phantom cutting from isolated B elements causes the direct DCD to be more severely affected by reductions in cut-rates compared to indirect DCD. In addition to failing to spread with low cut-rates, the type-1 resistance alleles produced by DCDs may persist and theoretically impact a future release of an otherwise appropriately tuned drive release targeting the same locus. Next, we set out to determine how deposition would affect DCDs and potentially reveal additional differences between the drive designs.

3.3.3 Maternal deposition can aid daisy-chain containment

The inclusion of deposition greatly expands the genotypes that can be exposed to cutting by the drive. We assume exclusively maternal deposition without any paternal deposition, and deposition occurs uniformly into each progeny, with the deposition cut-rate determining the fraction of target alleles that are cut in each progeny. The most straightforward deposition scenario is the simultaneous deposition of the Cas9 protein and gRNA as a complex. However, gRNAs in CRISPR HEGs are generally expressed from (putatively) constitutive pol III promoters. In line with experimental data[50;103;115;118], in our simulations embryonic cutting can also occur when the Cas9 protein is deposited into an individual that has inherited a gRNA expressing transgene from their father. In contrast, due to the reduced stability of the isolated gRNA [123–125], we assume that the pre-complexing of Cas9 and a gRNA in the mother is necessary for the gRNA to be deposited. Generally, cutting by the deposited nuclease is thought to occur much earlier in development, when HDR rates are much lower and HDR may in some cases be impossible due to physical isolation of homologous chromosomes[159;175;176]. As such, we assume that HDR repair does not occur in the early embryo and all cuts result in resistance alleles.

In Fig 3.4a-3.4c we show the A locus allele dynamics of simulations with a maternal deposition cut-rate of 25%. In contrast to the outcomes when the cut-rate is changed,

maternal deposition affects the two DCDs very similarly. The two DCDs reach maximum $_{A1}$D element frequencies of 64.5% and 66.6% for the direct DCD and indirect DCD, respectively. For both DCDs, the B drive element carrying the nuclease drops in frequency rapidly due to its association with deposition-induced fitness costs at the A locus (S3.14 Fig).

In the previous simulations, the fitness cost of type-2 alleles were only experienced by those individuals that inherited it, not the individual in whose germline it was generated. To account for somatic genotype conversion by embryonic cutting, fitness costs are applied after maternal deposition-based gene conversion occurs (but before expression-based conversion in the germline). Any individual in which its $_{A}$T alleles are cut due to deposition has its fitness lowered in proportion to the combined fitness of the individual genotypes of the now mosaic individual. Although cutting of $_{B}$ T alleles can occur due to deposition, resistance mutations are neutral at this locus. However, deposition still reduces the proportion of uncut target alleles available for subsequent homing in the germline. It is important to note that despite the A drive element being a rescue, the drive-inheriting individual are fully affected by resistance alleles that emerge from deposition as the target gene is haplolethal.

Fig 3.4d-3.4f show the $_{A1}$D allele dynamics for different maternal deposition rates. The $_{A2}$D and $_{A}$1 allele dynamics are shown in S3.15 Fig. The same simulations were performed without the possibility of gRNA deposition. Restricting deposition in such a way may, on the face of it, be expected to aid the drives. However, it actually makes the fitness costs associated with deposition exclusively tied to inheritance of a gRNA-expressing drive element, favouring individuals that do not inherit the gRNA-expressing drive element. There may be more complex interactions with nuclease-only deposition for the two DCD designs; however, we find no substantial differences in the overall outcomes (S3.16 Fig).

There have been several reports of high rates of inheritance bias from maternally deposited nuclease (shadow drive)[50;103;115;118]. This appears to be observed mainly in cases where Cas9 is provided by the mother and the gRNA gene from the father[1]. S3.17 Fig shows simulations of different maternal deposition rates with the possibility of shadow drive. We find that this specific scenario is so rare (and impossible for the self-perpetuating drive) that it has a negligible impact on the drive outcomes.

Finally, we tested how the drive frequency can be adapted to compensate for increasing rates of maternal deposition (Figs 3.4g-3.4i and S3.18). This shows that major changes in the deposition rate can be compensated for with relatively minor changes in the release frequency. This is in contrast to the same analysis described for changes to the germline-based cutting rate, which displayed a much more linear relationship. Strikingly, spread into population two was much more limited for DCD with maternal deposition, even at high release frequencies (S3.18 Fig). This provides a large parameter space with substantially higher drive spread into population one compared to population two (Fig 3.4j-3.4l) for the DCDs. The explanation for this is that once all target alleles have been cut, deposition ceases to have any consequence. However, at low frequencies, such as with the migration of drive-carrying individuals

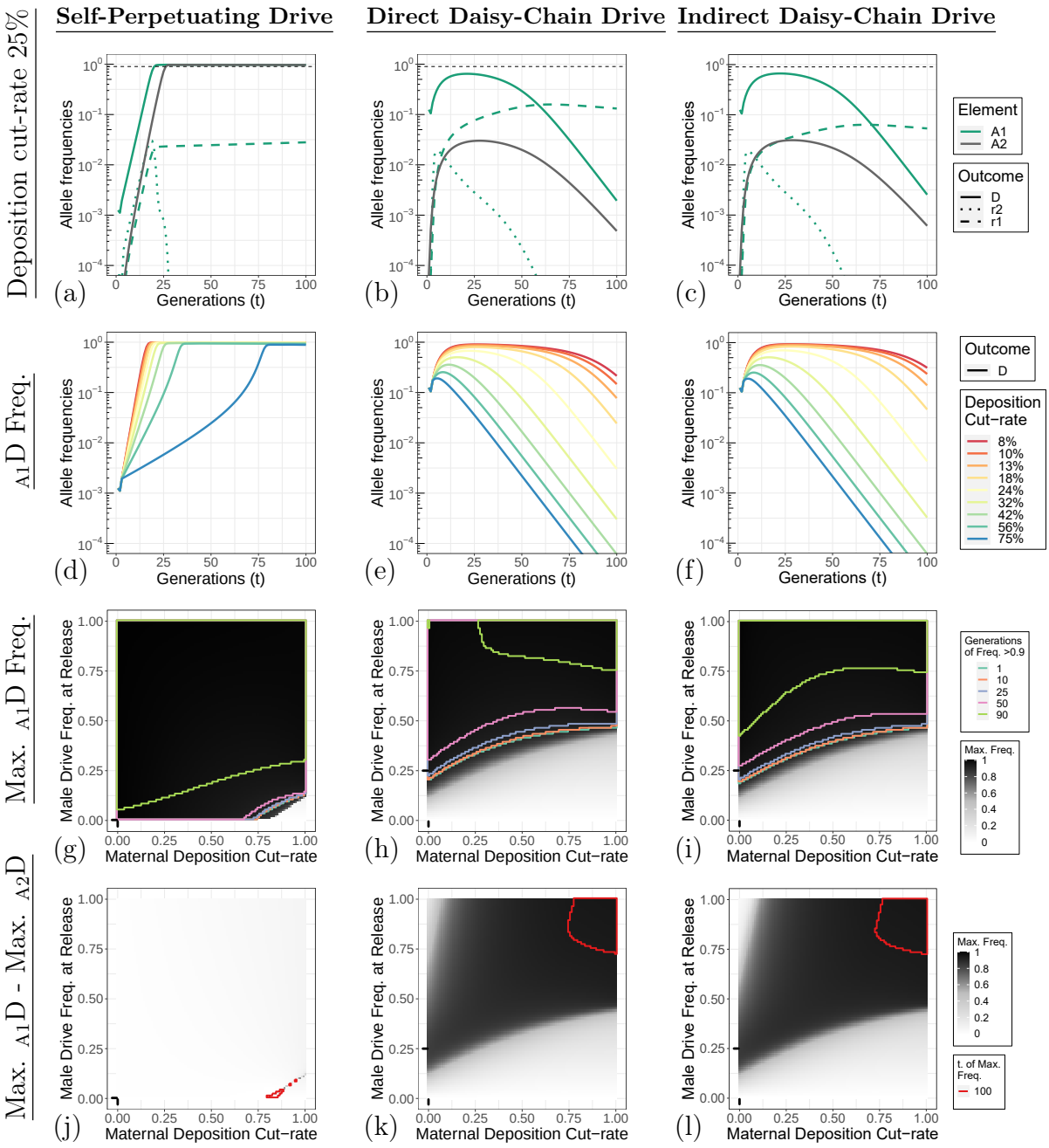


Figure 3.4: **Maternal deposition affects the drives when they are at low frequencies. Column 1.** Self-Perpetuating Drive. **Column 2.** Direct Daisy-Chain Drive. **Column 3.** Indirect Daisy-Chain Drive. **Row 1 (a-c).** Individual allele dynamics with a maternal deposition cut-rate of 25%. The thin dashed line indicates a frequency of 90%. An allele at locus A in population one is indicated by $_{A1}$ (green lines), and a drive allele at locus A in population two is indicated by $_{A2}$ (grey line). Type 1 and 2 resistance alleles are not shown for population two in these panels. The allele dynamics for all loci are shown in S3.14 Fig. **Row 2 (d-f).** A locus drive allele dynamics for population one with different maternal deposition cut-rates. Allele dynamics of the A locus drive allele in population two, and type-1 resistance alleles for both populations are shown in S3.15 Fig. **Row 3 (g-i).** Parameter sweep for the male drive frequency at release and maternal deposition rates. The shading of the heat map indicates the maximum $_{A}$D allele frequency in population one reached within the 100 generations simulated. The maximum frequency of $_{A}$1 alleles and outcomes for population two are shown in S3.18 Fig. An immediate decline after release results in the maximum frequency being the release frequency. Regions are boxed with coloured lines that indicate threshold values for the number of generations in which the $_{A1}$D allele frequency was greater than 90%. The black lines on each axis indicate the default value for the parameter being varied. **Row 4 (j-l).** Difference in the maximum $_{A}$D allele frequency between population one and two. Regions boxed with a red line are simulations in which the maximum allele frequency of population one or two was at generation 100 and the outcomes may change if additional generations are simulated.

to a population of mostly wild-types, the deposition-induced fitness costs still occur and limit the drive's ability to spread.

3.4 Discussion

Here we have investigated the effect of drive element fitness, type-1 resistance alleles, cut-rate, and maternal deposition on two alternative daisy-chain gene drives (DCDs). The DCDs that we model are the direct DCD ($_{A}E/_{B}N^{A}/_{C}G^{B}$), and indirect DCD ($_{A}G^{A}/_{B}N/_{C}G^{B}$), each composed of three elements (downstream to upstream: A, B, C). In each case, we have contrasted our results with the behaviour of a self-perpetuating drive ($_{A}N^{A}$) for which more experimental and other computational results have been reported on the effect of fitness costs[126;171], resistance alleles[79;126;171], cut-rate[171], and deposition[126;135]. Despite using the same underlying inheritance biasing mechanism, we find that the performance of the DCDs can differ from a single-element self-perpetuating drive in important ways.

There are important limitations to our model and the approach taken. Foremost among them is that we capture very little of the complexity of the life-history traits of the target species. Furthermore, our deterministic simulations of panmictic populations with simple migration will not adequately reflect any real-world populations. However, we expect that the broader relationships we have highlighted will remain true and can inform the design and optimisation of split and daisy-chain gene drives.

We find that the exponential nature of drive replication has different implications for the self-perpetuating drive and daisy-chain gene drives. In our simulations of replacement drives, the self-perpetuating drive generally has binary outcomes. When the drive can replicate itself faster than the drive is lost by fitness defects, the exponential replication of the drive will allow it to spread to high frequencies in both populations. If the drive is extremely inefficient, it will not spread at all. There is only a narrow set of parameters where the replication of the drive element is closely matched by the loss of drive elements, and the drive does spread but is substantially slowed. While the drive is spreading, more drive carriers are lost due to generational fitness effects, requiring more copying events (and therefore cutting) to reach equivalent frequencies compared to a faster spreading drive. The increased number of cutting events results in the generation of more resistance alleles. Whereas a self-perpetuating drive gains independence from the starting conditions and nuclease efficiencies from its sustained exponential spread, the daisy-chain gene drive is affected in the opposite way. The limited window of exponential replication experienced by DCDs causes their maximum frequency to be highly dependent on the starting conditions. Minor increases or decreases in efficiency will compound generation upon generation and can substantially alter the maximum frequency the drive can achieve, and its potential to spread beyond the target population. In our simulations, this compounding difference in efficiency is most apparent with changes to the drive allele fitness. However, type-1 resistance alleles, cut-rate, and maternal deposition also affect the balance of drive allele generation to loss and additionally have more specific interactions with the DCD

design.

The spread of DCDs is substantially impacted by factors that can increase the effective linkage disequilibrium between different elements. Under ideal conditions, upstream daisy-chain elements are always inherited together with their downstream drive element (A for B and B for C) because those upstream elements cause the downstream element to switch from a heterozygous to homozygous state. However, reductions in cut-rate and the generation of resistance alleles in the germline or in the early embryo through deposition can cause upstream elements to be inherited without the corresponding downstream element. In addition to lowering the inheritance biasing rate, this also lowers the probability that in a subsequent generation, the DCD elements are in a configuration that allows for inheritance biasing to occur at all. For a self-perpetuating drive, all drive components are housed on the same element and cannot segregate away from each other.

The downside of linkage disequilibrium between daisy-chain elements can be further exacerbated by phantom cutting. This occurs with the direct daisy-chain B element ($_{\mathrm{B}}\mathrm{N}^{\mathrm{A}}$), which can cut a wild-type A locus even when the B and A drive elements have segregated away from each other, generating additional resistance alleles. In our simulations, the production of additional dominant lethal type-2 resistance alleles at the A locus by phantom cutting, and the loss of B elements by association, has only a limited impact. This is because while an indirect DCD B element isolated from its A element cannot phantom cut, it also no longer contributes to the spread of the A element unless it is reunited with an A element by random mating. Reunification of B and A elements is only common if the drive is already at a high frequency, and by then additional B elements have been created, reducing the significance of the direct DCD B elements that have been lost. However, the effect of phantom cutting on the type-1 resistance allele frequency is more problematic. The additional production and biasing of type-1 resistance alleles by direct DCD phantom cutting substantially increase their frequency compared to the indirect DCD and self-perpetuating drive.

Throughout this study, we have exclusively discussed three-element DCDs. Incorporating additional DCD elements can increase invasiveness, as has been shown by modelling by others[59;173]. This can be achieved by adding upstream gRNA elements (e.g., $_{\mathrm{D}}\mathrm{G}^{\mathrm{C}}$), however, this now provides additional opportunities for phantom cutting (e.g., $_{\mathrm{B}}\mathrm{N}/_{\mathrm{C}}\mathrm{T}/_{\mathrm{D}}\mathrm{G}^{\mathrm{C}}$). Upstream drive element may be expected to reach lower maximum frequencies than downstream elements, limiting the impact of resistance alleles generated by phantom cutting at these sites. Nonetheless, relatively complex additional modifications (e.g., orthogonal nucleases) would likely be required to avoid phantom cutting with longer DCDs.

The default release frequency was chosen to allow for high-frequency spread of the two DCDs, and should be considered roughly tuned to our chosen drive efficiency and fidelity parameters. As a consequence, it is not a surprise that additional inefficiencies in the drive mechanism immediately affect the DCDs ability to approach fixation. Nevertheless, we show that the DCDs spread can be severely impacted by relatively minor changes to the drive efficiency. For a real-world release, the efficiency of the

drive must be known to a high degree of precision to be able to tune a DCD release to the intended application. Relatively small errors in estimation or differences between laboratory and field conditions may allow the daisy-chain drive to spread substantially less or more than predicted. As we have shown, higher release frequencies can buffer the DCD against additional inefficiencies, but this may be undesirable if spread beyond the intended population is to be minimised. However, for many applications, there may be a large difference between the minimal release frequency needed and a release frequency that could allow for unacceptable spread beyond the target population. For those cases, the minimal required DCD release frequency can be safely exceeded to buffer against shortcomings in the drive efficiency. In addition, the risk of undesired additional spread by a DCD may also be mitigated by a slow ramp-up of release frequencies and continuous monitoring of spread. Nonetheless, our results indicate that high inheritance biasing efficiency and fidelity will be required for a DCD to achieve the design's potential as useful middle ground between a self-perpetuating drive and non-drive inundative releases.

One possible way to address the increased nuclease efficiency and fidelity sensitivity we have found with DCDs is to tightly link individual elements by reducing their recombination distance. The elements should be close enough on the same chromosome, so segregation by meiotic recombination is rare but far enough that the homing events at each locus are still independent. This would negate the issues we find of premature dissociation of drive elements and phantom cutting; upstream elements should very rarely segregate away from downstream elements, even with low inheritance biasing efficiency and fidelity. However, there are a number of reports of drives designed to function through homing, also affecting the inheritance of a separate sequence located on the same or homologous chromosome[3;50;85;91;95;103]. This collateral inheritance bias could, in a daisy-chain with tightly linked elements, cause two or more elements to bias themselves (more) like a self-perpetuating drive[169]. Daisy-chain drives in such a configuration may require higher safety assessments to ensure that it remains self-limiting in phased trials from the laboratory to the field.

An important consequence of DCD design that we have not considered in our simulations, that may be addressed in future work, is the potential interference between different drives. Cutting by the direct DCD A element is dependent on nuclease expression from another element for its function. However, the nuclease could be supplied by a different drive altogether. A rogue Cas9-based self-perpetuating drive could cause an indirect DCD drive A element to piggyback on the Cas9 expression and spread together with the self-perpetuating drive. This is similar to the behaviour of some reversal drives[177]. This piggybacking behaviour is not possible for the otherwise inferior direct DCD.

To date, inheritance biasing efficiency and fidelity exceeding our default parameters (cut-rate of 100%, HDR rate of 95%) has been achieved in *Anopheles gambiae* mosquitos in a single-element construction[51]. These results are encouraging for developing a DCD in the same species; however, it is unclear how the results with a single-element extend to a DCD. Multiple simultaneous DNA breaks have been reported to

cause additional cellular stress[66] and interchromosomal abnormalities[178–182]. While some drives that bias multiple loci simultaneously have been reported[114;183], to our knowledge, no multigenerational assessment of a DCD has been published. It remains to be demonstrated that a DCD can achieve similar efficiency and fidelity as a single-element drive. Furthermore, to date, drive efficiencies have been low in other potential target species such as *Aedes aegypti*[3;95;172] and *Mus musculus*[75;93;94], or high, but severely affected by maternal deposition, in *Anopheles stephensi*[161]. We expect that substantial improvements need to be made to allow HEGs in these species to function effectively with a DCD design.

It is important to note that there are many ways in which deposition can manifest (e.g., paternal vs. maternal deposition rates, deposition patterns for the gRNA, Cas9, and Cas9;gRNA complex, mRNA or protein deposition, and deposition rates from drive homozygous or heterozygous parents, repair rates through development) and many factors that influence the consequences of deposition (e.g., recessive and dominant fitness costs, life stage of fitness cost manifestation, threshold dependence of somatic conversion). As such, the deposition simulations performed here necessarily present a small subset of possible conditions and highlight that the consequence of deposition is context- and design-dependent. As more experimental data become available, further work can specifically test the deposition conditions observed for daisy-chain gene drives.

The work we have presented here demonstrates a complex interplay between gene drive design and sensitivity to molecular inefficiencies. This shows that to guide gene drive design, we need to have a robust understanding of the actual values that certain parameters take on in practice. In the next chapter, we will systematically investigate the experimental performance of published homing gene drives. We use this to better understand the role of sex in gene drive performance and develop a web tool to investigate the influence of individual design characteristics.

3.5 Methods

Overview

The model we present here is capable of tracking the frequency of different genotypes through discrete life stages and generations. To facilitate that, we devised a nomenclature system capable of defining the range of possible genotypes and parental deposition states that individuals can take on for daisy-chain gene drives (the most complex case we consider). This is required to model a range of possible different drive designs and the many different individual states that a single drive can give rise to during its spread. During the embryonic and germline life stages, drive-induced genotype conversion occurs, changing the relative frequency of the genotypes. We use non-overlapping discrete-generation recursion equations for genotype frequencies, treating males and females separately. This method is applied in a simple deterministic allele frequency simulation of a single release of gene drive individuals into a population

of all wildtypes.

Stages and genotype nomenclature

The model tracks the frequency of different genotypes through different life stages. Each discrete generation (t), the same stages are repeated. The stages and their corresponding frequency vectors are: zygote ($\mathbf{z}$), embryonic ($\mathbf{e}$), germline ($\mathbf{g}$), and gamete ($\mathbf{p}$) stages. All vectors apart from the zygote vector are sex-specific, and male versions are indicated with a bar below the symbol (e.g., $\underline{\mathbf{p}}$). The overall genome frequency may vary through these stages; however, each generation the frequency is normalised to 1. For the particular drive system we are considering, d is the number of unique diploid genotypes, while h is the number of possible haploid genotypes. As such, $\mathbf{z}_{\mathrm{i}}$ is the frequency of the i^{th} zygotic genotype, with i ranging from 1 to d.

Throughout each stage, the genomes maintain their genotype unless they are exposed to a gRNA:nuclease complex and contain a target allele (T) that can be cut. The other alleles we consider are type-1 (1), functional, and type-2 (2) non-functional resistance mutations and drive alleles. The symbol for the drive alleles is distinguished by whether that allele contains a nuclease (N), no nuclease but a gRNA (G), or neither (E). The N allele can carry a gRNA, but does not necessarily have to. None of the alleles except T can be cut but, if present, can influence the repair outcome of the T allele on the homologous chromosome. The drive elements can, depending on the specifics of the drive, provide the components necessary for a cut to occur at another locus but do not influence the repair outcomes at that other locus.

To model daisy-chain drives, we must further distinguish genotypes at different genomic loci. The defining feature of the persistence of inheritance bias for a daisy-chain element is its priority in the chain[59]. Sequentially going from A, B, C, etc. To aid in interpretation, each element's genotype (diploid combinations of: T/N/G/1/2) has a letter subscript that denotes its order in the daisy-chain (e.g., ${}_{\mathrm{A}}\mathrm{TT}$). Any genotype with a different letter subscript will be at a different genomic locus. For drive elements that carry a gRNA, a superscript indicates which T allele locus they target (e.g., ${}_{\mathrm{A}}\mathrm{N}^{\mathrm{A}}$ or ${}_{\mathrm{C}}\mathrm{G}^{\mathrm{B}}$). Table 3.1 lists the genotype notation of the drives studied in this publication and the two-element split-drive.

Genotype	Name
${}_{\mathrm{A}}\mathrm{N}^{\mathrm{A}}\mathrm{T}$	Self-Perpetuating Drive
${}_{\mathrm{A}}\mathrm{G}^{\mathrm{A}}\mathrm{T}/{}_{\mathrm{B}}\mathrm{NT}$	Split-drive
${}_{\mathrm{A}}\mathrm{ET}/{}_{\mathrm{B}}\mathrm{N}^{\mathrm{A}}\mathrm{T}/{}_{\mathrm{C}}\mathrm{N}^{\mathrm{B}}\mathrm{T}$	Noble Daisy-Chain Drive
${}_{\mathrm{A}}\mathrm{ET}/{}_{\mathrm{B}}\mathrm{N}^{\mathrm{A}}\mathrm{T}/{}_{\mathrm{C}}\mathrm{G}^{\mathrm{B}}\mathrm{T}$	Direct Daisy-Chain Drive
${}_{\mathrm{A}}\mathrm{G}^{\mathrm{A}}\mathrm{T}/{}_{\mathrm{B}}\mathrm{NT}/{}_{\mathrm{C}}\mathrm{G}^{\mathrm{B}}\mathrm{T}$	Indirect Daisy-Chain Drive

Table 3.1: Genotype specification of drive designs. Drives are listed in their heterozygote state.

Haploid to Diploid stage

Each discrete generation, the current genotype frequencies can change due to the action of the drive and fitness effects (S3.1a Fig). For generations 2-100, this process starts by combining the haploid genotype frequencies of the previous generation from the male ($\underline{\mathbf{p}}^{t-1}$) and female ($\mathbf{p}^{t-1}$) individuals to form the frequency vector of the zygotes ($\mathbf{z}^t$) (equation 3.1). For generations 2-100, there is a single vector $\mathbf{z}$, since, in our model, sex effects manifest only at the embryonic stage. However, for the first generation, the embryonic frequency is calculated separately for males and females according to the release frequencies. As such, released individuals are reduced according to their fitness before their opportunity to mate.

The model tracks the deposition of drive components from one generation to another. For Cas9 deposition, we differentiate each allele received by the zygote by a sex-specific Cas9 deposition factor ($_0$ or $_1$). We do not differentiate whether the deposition comes from a nuclease element heterozygote, or homozygote parent. For example, a haploid T$\}_0$ allele has come from a non-Cas9 carrying diploid parent (e.g., TT), while a T$\}_1$ must have come from a nuclease carrying parent (e.g., TN). In the diploid genomes, these notations are combined, with the drive status of the male parent listed first (e.g., $\}_{0*}$) and then female parent (e.g., $\}_{*0}$). Equation 3.2 illustrates the combining of $\underline{\mathbf{p}}^{t-1}$ and $\mathbf{p}^{t-1}$ to form $\mathbf{z}^t$. Note, that many genotypes are functionally equivalent and will be summed when reporting the outcomes of the drive (e.g., $\{\text{T1}\}_{00} = \{\text{1T}\}_{00}$, $\{\underline{\mathbf{p}}_i/\mathbf{p}_j\}_{00} = \{\underline{\mathbf{p}}_j/\mathbf{p}_i\}_{00}$, $\{\underline{\mathbf{p}}_i/\mathbf{p}_j\}_{11} = \{\underline{\mathbf{p}}_j/\mathbf{p}_i\}_{11}$, $\{\underline{\mathbf{p}}_i/\mathbf{p}_j\}_{10} \neq \{\underline{\mathbf{p}}_j/\mathbf{p}_i\}_{01}$). Deposition of the gRNA is tracked in a similar way. The haploid genotypes are differentiated by the gRNAs that have been expressed in the diploid parent, and this is used in the calculation of deposition-based cutting in the diploid embryos.

$$\mathbf{z}^t = \underline{\mathbf{p}}^{t-1} \cdot \mathbf{p}^{t-1} \tag{3.1}$$

$$\begin{array}{c} \\ T\}_0 \\ N\}_1 \\ *\}_* \\ \underline{\mathbf{p}_h} \end{array} \begin{array}{c} \begin{array}{cccc} T\}_0 & N\}_1 & *\}_* & \mathbf{p}_h \end{array} \\ \begin{pmatrix} \{TT\}_{00} & \{TN\}_{01} & \{T*\}_{0*} & \{T\mathbf{p}_h\}_{0*} \\ \{NT\}_{10} & \{NN\}_{11} & \{N*\}_{1*} & \{N\mathbf{p}_h\}_{1*} \\ \{*T\}_{*0} & \{*N\}_{*1} & \{**\}_{**} & \{*\mathbf{p}_h\}_{**} \\ \{\underline{\mathbf{p}_h}T\}_{*0} & \{\underline{\mathbf{p}_h}N\}_{*1} & \{\underline{\mathbf{p}_h}*\}_{**} & \mathbf{z}_d \end{pmatrix} \end{array} = \underline{\mathbf{p}}^{t-1} \cdot \mathbf{p}^{t-1} = mat(\mathbf{z}^t) \tag{3.2}$$

Fitness costs, and genotype inter-conversion by deposition and expression

In the transition from the zygotic to the embryonic state, the model takes into account deposition and fitness effects. Through deposition, each individual zygotic genotype ($\mathbf{z}_i$) can give rise to multiple embryonic genotypes ($\mathbf{e}$). Genotype conversion is mediated by matrix $\mathbf{K}$, which converts genotypes according to the deposition cut and repair rates. Furthermore, based on the fitness parameters (Fig 3.1c), we can, for each

genotype, define a particular fitness cost (0-1) in the vector θ. The fitness cost vector has a female (θ) and male ($\underline{\theta}$) specific version because for some simulations sex-specific fitness costs are applied. The fitness of a particular genotype (θ_i) is determined by multiplying the fitness costs of individual alleles. The dominant fitness cost imposed by a particular allele is applied only once, even if two copies are present. Some simulations impose an additional recessive fitness cost on the A locus, which only occurs if two loss of function alleles are present at A (e.g., $_{\text{A}}$EE, $_{\text{A}}$E2, $_{\text{A}}$22). As the multiple genotypes produced by deposition are within the same mosaic organism, the fitness cost contribution of individual genotypes is summed and applied uniformly over all genotypes arising from the same zygotic genotype. The deposition conversion outcomes and fitness costs of any one genotype do not vary throughout a simulation, only the relative frequency of the input genotypes in the form of $\mathbf{z}$. As such, conversion matrix $\mathbf{D}$/ $\underline{\mathbf{D}}$ can be constructed that mediates both deposition conversion ($\mathbf{K}$), and fitness costs (θ_i and $\underline{\theta}_i$). Matrix $\mathbf{D}$ is the product of the row-wise multiplication of matrix $\mathbf{K}_{i,*}$ by θ and normalisation. When $\mathbf{z}$ is then subjected to conversion matrix $\mathbf{D}$ it produces the embryonic frequency vector $\mathbf{e}$ (equations 3.3). Vector $\mathbf{e}$ represents the genotype frequency at the end of the embryonic stage, after deposition and fitness costs have been applied. Vector $\mathbf{e}/\underline{\mathbf{e}}$ is subsequently multiplied by the expression-based gene conversion matrix $\mathbf{C}/\underline{\mathbf{C}}$ to produce the gamete frequency vector $\mathbf{g}/\underline{\mathbf{g}}$ (equations 3.4). The construction of matrices $\mathbf{C}$ and $\mathbf{K}$ is discussed after gamete production.

$$\mathbf{e}^t = \mathbf{z}^t \cdot \mathbf{D} \qquad \underline{\mathbf{e}}^t = \mathbf{z}^t \cdot \underline{\mathbf{D}} \tag{3.3}$$

$$\mathbf{g}^t = \mathbf{e}^t \cdot \mathbf{C} \qquad \underline{\mathbf{g}}^t = \underline{\mathbf{e}}^t \cdot \underline{\mathbf{C}} \tag{3.4}$$

Gamete production

The last step in each generation is the conversion of diploid genomes to haploid ones. This is done by multiplying the diploid genome frequency vector $\mathbf{g}$ by matrix $\mathbf{H}$ to form the haploid genome frequency vector $\mathbf{p}$ (equations 3.5). Matrix $\mathbf{H}$ simply splits each diploid genome into the possible haploid genomes it can produce, an example of which is shown in equation 3.6. An important step in this process is to keep track of nuclease and gRNA deposition. The presence of at least one N allele in the diploid germline genome causes the haploid genome to have an associated deposition marker 1 $(*\}_1)$. If no N allele is present in the diploid genome, the deposition marker is 0 $(*\}_0)$. Separately, any gRNA-expressing elements in the diploid parent cause each haploid genome to carry an indicator for deposition of that gRNA. Like Cas9 deposition, this occurs even if the gRNA-expressing element is not inherited. We assume that alleles at separate loci segregate fully independently, as is shown in equation 3.7 which shows $\mathbf{H}$ if two elements are considered.

$$\mathbf{p}^t = \mathbf{g}^t \cdot \mathbf{H} \qquad \underline{\mathbf{p}}^t = \underline{\mathbf{g}}^t \cdot \mathbf{H} \tag{3.5}$$

$$
\begin{array}{c}
 & \{T\}_0 & \{1\}_0 & \{2\}_0 & \{G\}_0 & \{T\}_1 & \{N\}_1 & \{1\}_1 & \{2\}_1 & \{G\}_1 \\
\{TT\} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\{T1\} & .5 & .5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\{T2\} & .5 & 0 & .5 & 0 & 0 & 0 & 0 & 0 & 0 \\
\{TG\} & .5 & 0 & 0 & .5 & 0 & 0 & 0 & 0 & 0 \\
\{TN\} & 0 & 0 & 0 & 0 & .5 & .5 & 0 & 0 & 0 \\
\{NN\} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
\{N1\} & 0 & 0 & 0 & 0 & 0 & .5 & .5 & 0 & 0 \\
\{N2\} & 0 & 0 & 0 & 0 & 0 & .5 & 0 & .5 & 0 \\
\{NG\} & 0 & 0 & 0 & 0 & 0 & .5 & 0 & 0 & .5 \\
\{11\} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\{12\} & 0 & .5 & .5 & 0 & 0 & 0 & 0 & 0 & 0 \\
\{1G\} & 0 & .5 & 0 & .5 & 0 & 0 & 0 & 0 & 0 \\
\{22\} & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
\{2G\} & 0 & 0 & .5 & .5 & 0 & 0 & 0 & 0 & 0 \\
\{GG\} & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0
\end{array}
= \mathbf{H} \tag{3.6}
$$

$$
\begin{array}{c}
 & \{{}_{\mathrm{A}}T/{}_{\mathrm{B}}T\}_0 & \{{}_{\mathrm{A}}T/{}_{\mathrm{B}}1\}_0 & \{{}_{\mathrm{A}}1/{}_{\mathrm{B}}T\}_0 & \{{}_{\mathrm{A}}1/{}_{\mathrm{B}}1\}_0 & \ldots \\
\{{}_{\mathrm{A}}TT/{}_{\mathrm{B}}TT\} & 1 & 0 & 0 & 0 & \ldots \\
\{{}_{\mathrm{A}}TT/{}_{\mathrm{B}}T1\} & .5 & .5 & 0 & 0 & \ldots \\
\{{}_{\mathrm{A}}T1/{}_{\mathrm{B}}T1\} & .25 & .25 & .25 & .25 & \ldots \\
\{{}_{\mathrm{A}}\ldots/{}_{\mathrm{B}}\ldots\} & \ldots & \ldots & \ldots & \ldots & \ldots
\end{array}
= \mathbf{H} \tag{3.7}
$$

Deposition and drive-induced conversion

To model drive activity within an individual, we approach gene editing as the conversion of genomes from one genotypic state to another genotypic state. The probabilities underlying this will depend on the probability that a DNA cut will occur (κ) and the HDR rate (α), together with the ratio between type-1 and type-2 resistance mutations (β). From this, DNA repair can result in three outcomes: interchromosomal-homology-directed repair (α), type-1 resistance mutations ($(1- \alpha) \cdot \beta = \mu$), and type-2 resistance mutations ($(1- \alpha) \cdot (1 - \beta) = \nu$). The distinction between type-1 and type-2 mutations is dependent on the target locus, and in our model only relevant for accounting for fitness effects. The probability of all repair outcomes sums to 1. For each simulation, the value of each of these parameters is fixed and the default values are listed in Fig 3.1c. Each of the previous parameters is defined separately for maternal deposition-based conversion, which is indicated by $_{01}$ (κ_{01}, α_{01}, and β_{01}).

In our model, there are two stages where the drive can induce a genotype conversion: embryonic and germline. For each genotype and its associated deposition state, we define a rate at which it converts to the other possible genotypes or remains the same

(the overall rate is always 1). These conversion probabilities are mediated by matrix **C** for the drive expression in the germline, and **K** for drive deposition in the embryo. For the majority of genotypes, the conversion rate maintaining the original genotype will be 1, and the conversion rate to all other genotypes will be 0. This is because only a select few genotypes can convert. The requirements for conversion are: first, the genotype must have a genomically expressed gRNA or deposited gRNA (in complex with Cas9). Second, that gRNA must target a locus that in that particular individual has one or more T alleles present. Lastly, there must be genomically expressed or deposited nuclease. If these conditions are met, we use the cut and repair probabilities to determine the conversion rate. If a T allele can be cut in the germline, $\kappa \cdot \alpha$ of the T alleles will convert to whatever allele is on the homologous chromosome. $\kappa \cdot \mu$ of the T alleles will convert to a type-1 resistance allele and $\kappa \cdot \nu$ of the T alleles will convert to a type-2 resistance allele. Finally, 1-κ of the T alleles will remain uncut. These conversion rates populate matrix **C**, as shown in equation 3.8. The exact same relationships hold for matrix **K** shown in equation 3.9, with the κ_{01}, α_{01}, μ_{01}, and ν_{01} parameters. The conversion matrix for genotypes with TT is more complex. For a TT genotype to be cut, the nuclease and gRNA must be expressed from a separate locus, or alternatively, the nuclease, or nuclease and gRNA can be provided by maternal deposition. For TT genotypes, we assume that $\alpha = 0$, or in a separate set of supplemental simulations alleles are cut one by one, resulting in the conversion probabilities shown in equation 3.10.

$$\begin{array}{c} \\ \{1T\} \\ \{2T\} \end{array} \begin{array}{c} \begin{array}{ccccc} \{1T\} & \{2T\} & \{12\}=\{21\} & \{11\} & \{22\} \end{array} \\ \begin{pmatrix} 1-\kappa & 0 & \kappa\cdot\nu & \kappa\cdot(\mu+\alpha) & 0 \\ 0 & 1-\kappa & \kappa\cdot\mu & 0 & \kappa\cdot(\nu+\alpha) \end{pmatrix} \end{array} = C \tag{3.8}$$

$$\begin{array}{c} \\ \{1T\}_{01} \\ \{2T\}_{01} \end{array} \begin{array}{c} \begin{array}{ccccc} \{1T\} & \{2T\} & \{12\}=\{21\} & \{11\} & \{22\} \end{array} \\ \begin{pmatrix} 1-\kappa_{01} & 0 & \kappa_{01}\cdot\nu_{01} & \kappa_{01}\cdot(\mu_{01}+\alpha_{01}) & 0 \\ 0 & 1-\kappa_{01} & \kappa_{01}\cdot\mu_{01} & 0 & \kappa_{01}\cdot(\nu_{01}+\alpha_{01}) \end{pmatrix} \end{array} = K \tag{3.9}$$

$$\begin{array}{c} \\ \{TT\} \\ \{TT\} \\ \{TT\} \end{array} \begin{array}{c} \begin{array}{cc} \{TT\} & \{1T\}=\{T1\} \end{array} \\ \begin{pmatrix} ((1-\kappa)+\kappa\cdot\alpha)^2 & (\kappa\cdot\mu)\cdot(1-\kappa)+((1-\kappa)+\kappa\cdot\alpha)\cdot(\kappa\cdot\mu) \end{pmatrix} \\ \begin{array}{cc} \{2T\}=\{T2\} & \{12\}=\{21\} \end{array} \\ \begin{pmatrix} (\kappa\cdot\nu)\cdot(1-\kappa)+((1-\kappa)+\kappa\cdot\alpha)\cdot(\kappa\cdot\nu) & (\kappa\cdot\mu)\cdot(\kappa\cdot\nu)\cdot 2 \end{pmatrix} \\ \begin{array}{cc} \{11\} & \{22\} \end{array} \\ \begin{pmatrix} (\kappa\cdot\mu)\cdot((\kappa\cdot\mu)+(\kappa\cdot\alpha)) & (\kappa\cdot\nu)\cdot((\kappa\cdot\nu)+(\kappa\cdot\alpha)) \end{pmatrix} \end{array} = C\{TT\} \tag{3.10}$$

Acknowledgements

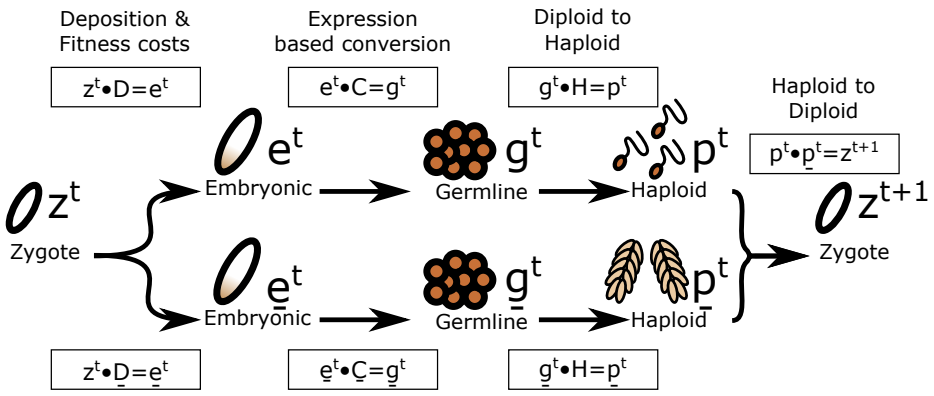
We thank Joshua X. D. Ang, Matthew P. Edgington, and Timothy Harvey-Samuel for their critical reading of the manuscript.

The views, opinions, and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the U.S. Government. The funders had no role in study design, data collection and analysis, the decision to publish, or preparation of the manuscript.

Author contributions

S.A.N.V. and M.B.B. conceptualised the study. S.A.N.V. performed all investigation, formal analysis, software development, visualisation, and writing of the original draft. M.B.B. provided direct supervision, with more general supervision from M.A.E.A., and L.A.. All authors contributed to review & editing of the manuscript draft.

3.6 Supporting information



(a)

Name	Symbol
Zygote Frequencies	z
Embryonic Frequencies	e
Germline Frequencies	g
Haploid Frequencies	p
Time in generations	t
Deposition and Fitness conversion	D
Expression conversion	C
Diploid to haploid conversion	H
Male specific variable	$\underline{*}$

(b)

Figure S3.1: **Overview of the genotype conversion process that repeats each generation. (a)** Illustration of the different stages of the model. **(b)** Names of the symbols used in a. Each genotype conversion step is expanded upon in the methods.

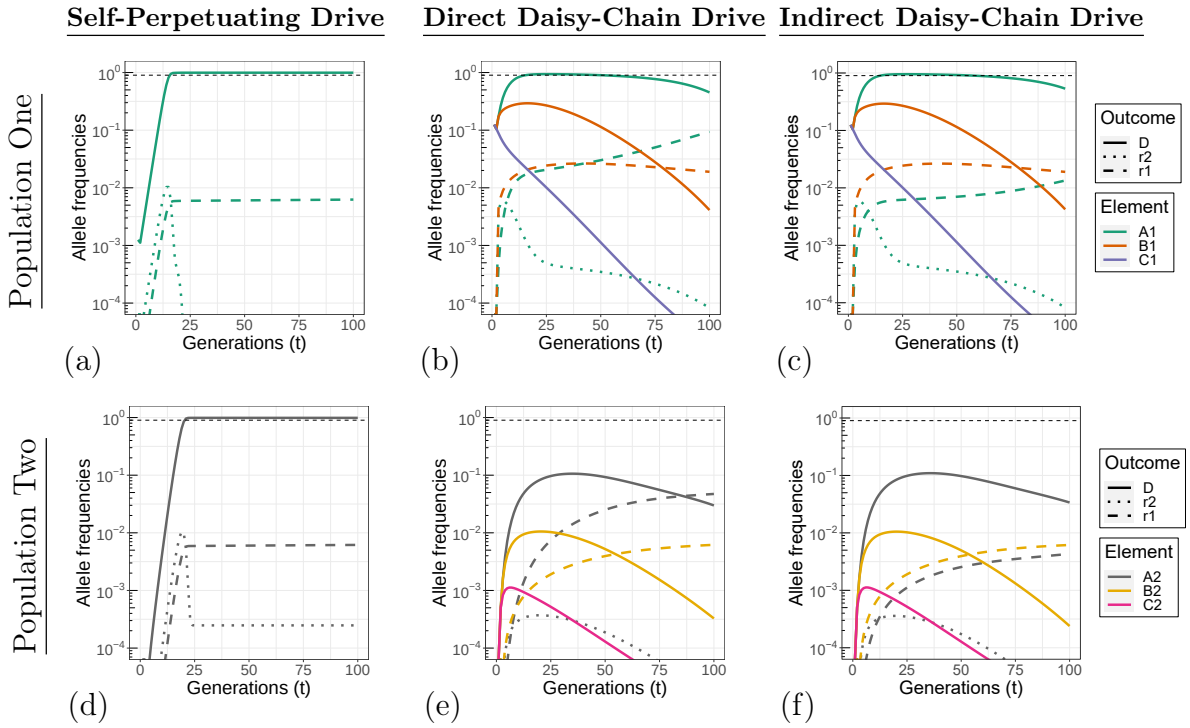


Figure S3.2: **Individual allele dynamics with all parameters at default. Row 1 (a-c).** Individual allele dynamics for population one. **Row 2 (d-f).** Individual allele dynamics for population two. **Column 1.** Self-Perpetuating Drive. **Column 2.** Direct Daisy-Chain Drive. **Column 3.** Indirect Daisy-Chain Drive. The thin dashed line indicates a frequency of 90%.

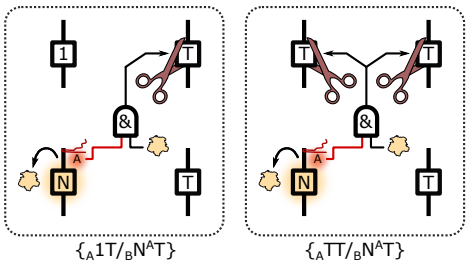


Figure S3.3: **Phantom cutting occurs when an upstream element attempts to bias the inheritance of a downstream drive element that is not present.** With cut-rates less than 100%, downstream T alleles can be inherited with a drive element that can allow the phantom cutting of a homozygous TT genotype.

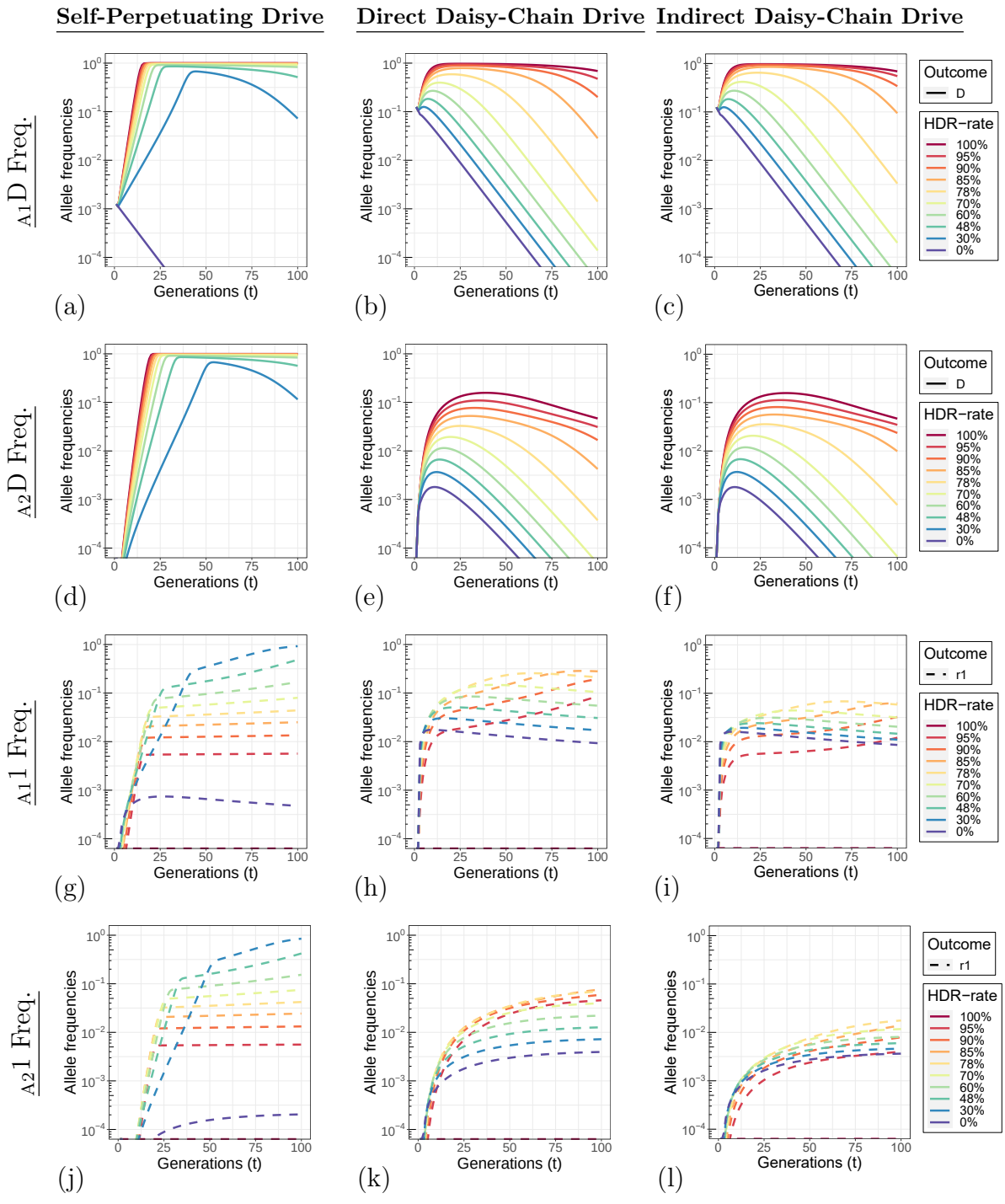


Figure S3.4: **Germline HDR rate parameter sweep. Row 1 (a-c).** Allele dynamics of the A locus drive element in population one. **Row 2 (d-f).** Allele dynamics of the A locus drive element in population two. **Row 3 (g-i).** Allele dynamics of the A locus type-1 resistance mutations in population one. **Row 4 (j-l).** Allele dynamics of the A locus type-1 resistance mutations in population two. **Column 1.** Self-Perpetuating Drive. **Column 2.** Direct Daisy-Chain Drive. **Column 3.** Indirect Daisy-Chain Drive.

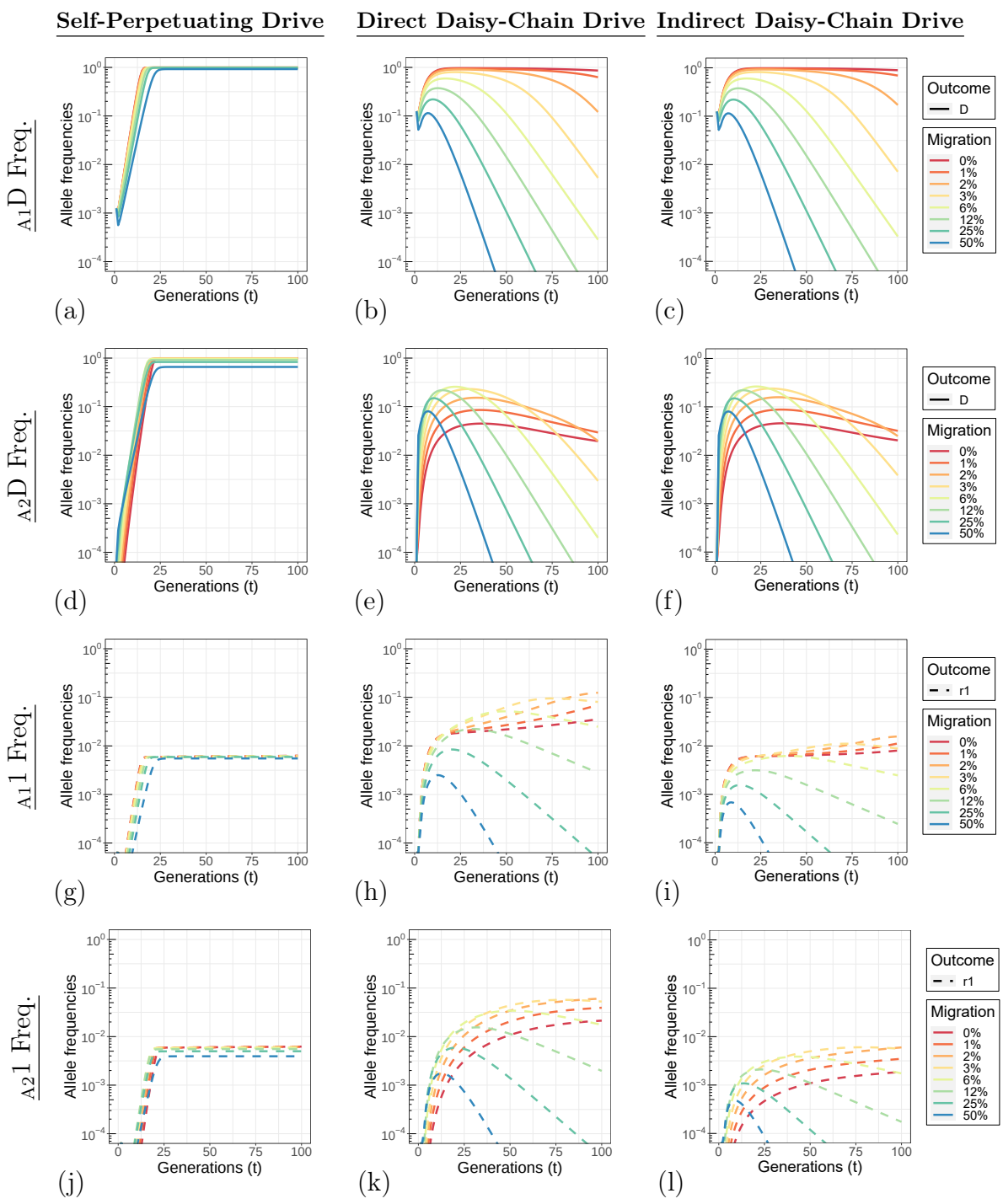


Figure S3.5: **Migration rate parameter sweep. Row 1 (a-c).** Allele dynamics of the A locus drive element in population one. **Row 2 (d-f).** Allele dynamics of the A locus drive element in population two. **Row 3 (g-i).** Allele dynamics of the A locus type-1 resistance mutations in population one. **Row 4 (j-l).** Allele dynamics of the A locus type-1 resistance mutations in population two. **Column 1.** Self-Perpetuating Drive. **Column 2.** Direct Daisy-Chain Drive. **Column 3.** Indirect Daisy-Chain Drive. Migration rates are 0.5^n, with n from 1 to 8.

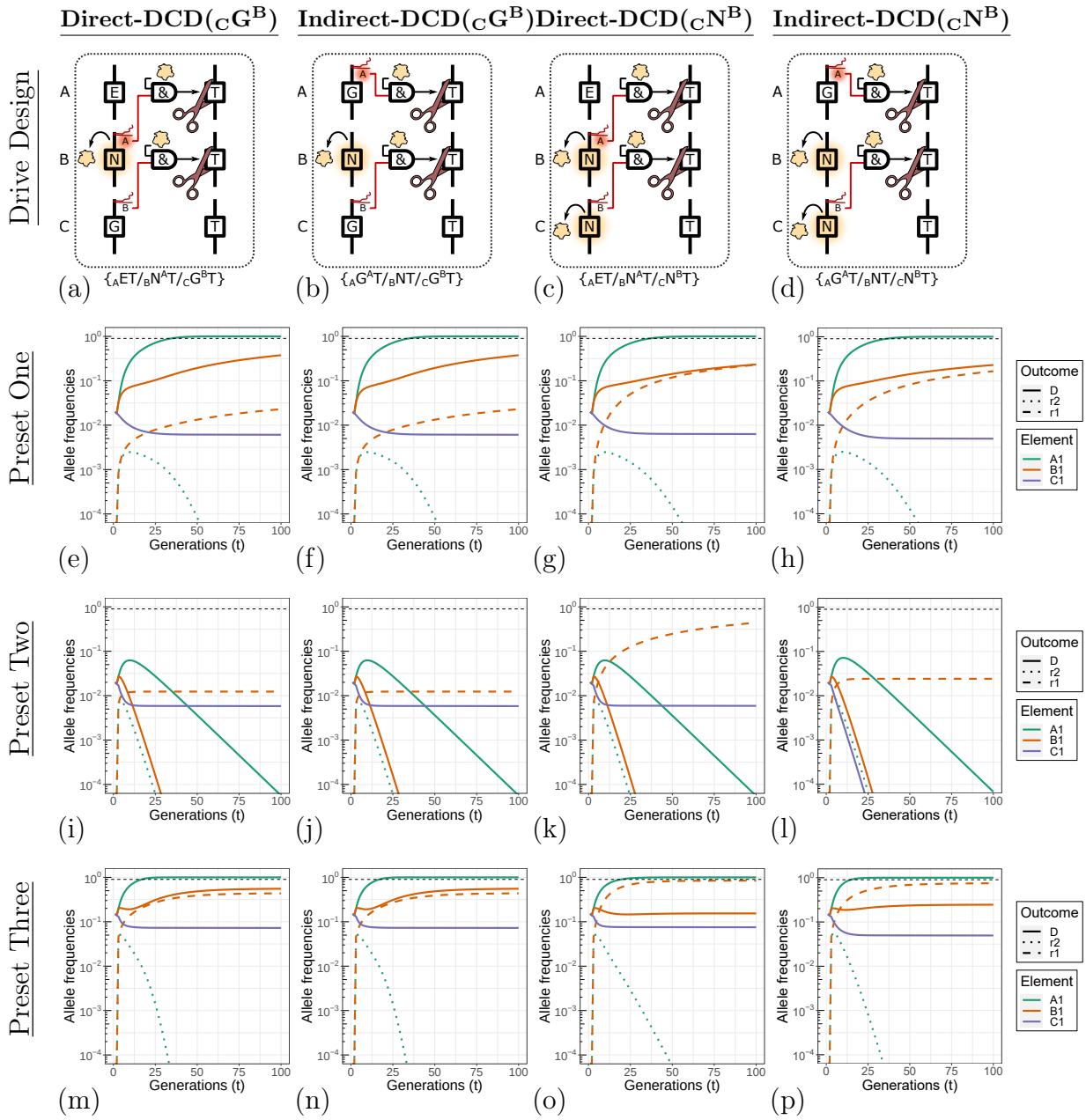


Figure S3.6: **Simulation of different drive designs using previously published parameters. Row 1 (a-d).** Illustration of the daisy-chain design. The remaining rows are simulations of each daisy-chain drive under different release and HDR rates used by Noble et al. 'Fig.2B'. **Row 2 (e-h).** Preset 1 is a simulation with a release frequency of 2% and HDR rate of 95%. **Row 3 (i-l).** Preset 2 is a simulation with a release frequency of 2% and HDR rate of 60%. **Row 4 (m-p).** Preset 3 is a simulation with a release frequency of 15% and an HDR rate of 60%. For all simulations, the allele fitness is: $_A$D = 92%, $_A$2 = 0%, $_B$D and $_C$D = 99.99%, $_B$2 = 100%. Cut-rate = 100%. These simulations were of a single population. **Column 1.** Direct Daisy-Chain Drive. **Column 2.** Indirect Daisy-Chain Drive. **Column 3.** Direct Daisy-Chain Drive with Cas9 also expressed from the C drive element. **Column 4.** Indirect Daisy-Chain Drive with Cas9 also expressed from the C drive element. The thin dashed line indicates a frequency of 90%.

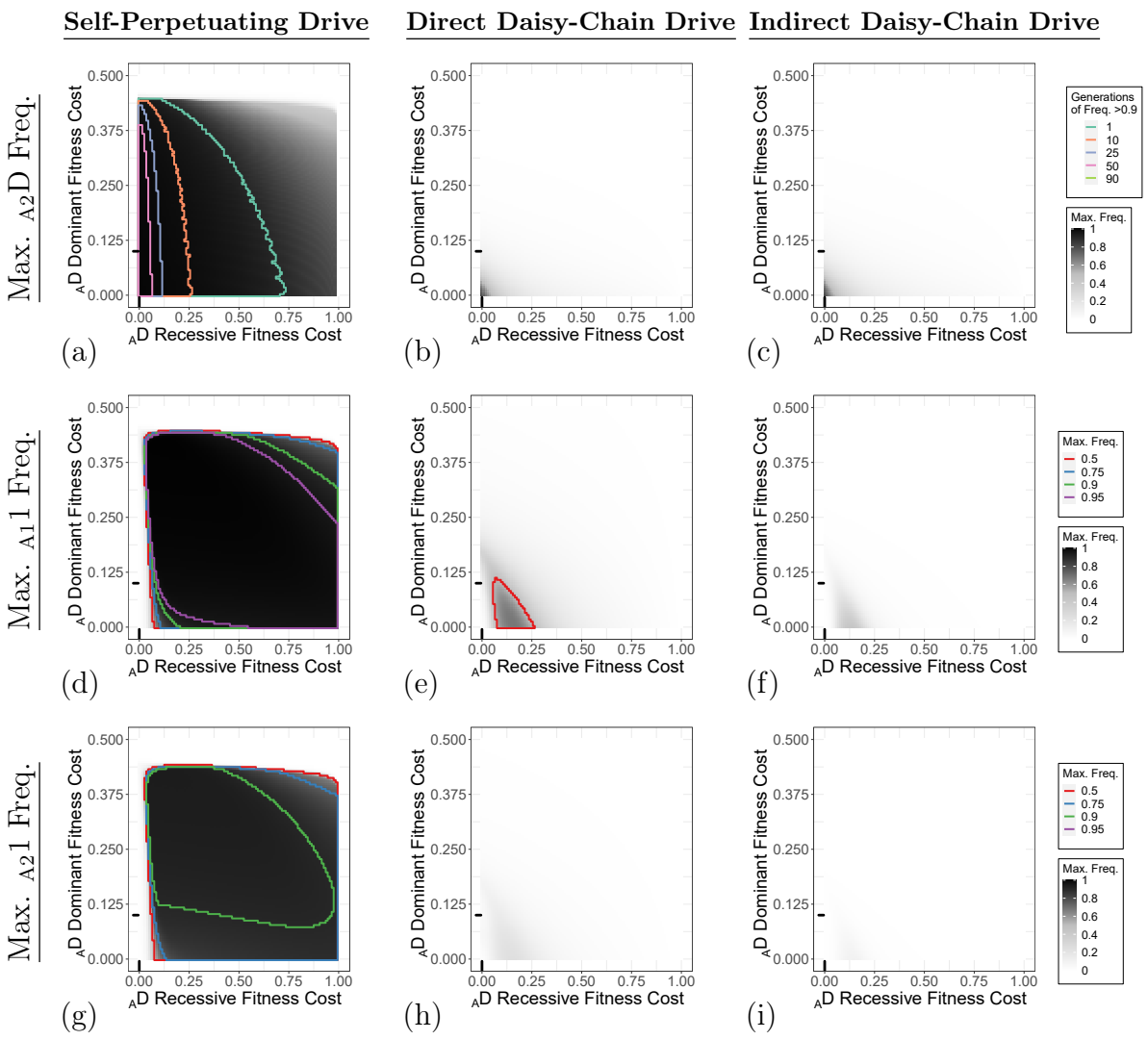


Figure S3.7: **Simultaneous changes to the recessive and dominant fitness cost of the $_{A}$D allele. Row 1 (a-c).** Maximum $_{A}$D allele frequency in population two. **Row 2 (d-f).** Maximum $_{A}$1 allele frequency in population one. **Row 3 (g-i).** Maximum $_{A}$1 allele frequency in population two. The maximum frequency of $_{A}$D in population one and the difference between population one and two is shown in Fig 2.

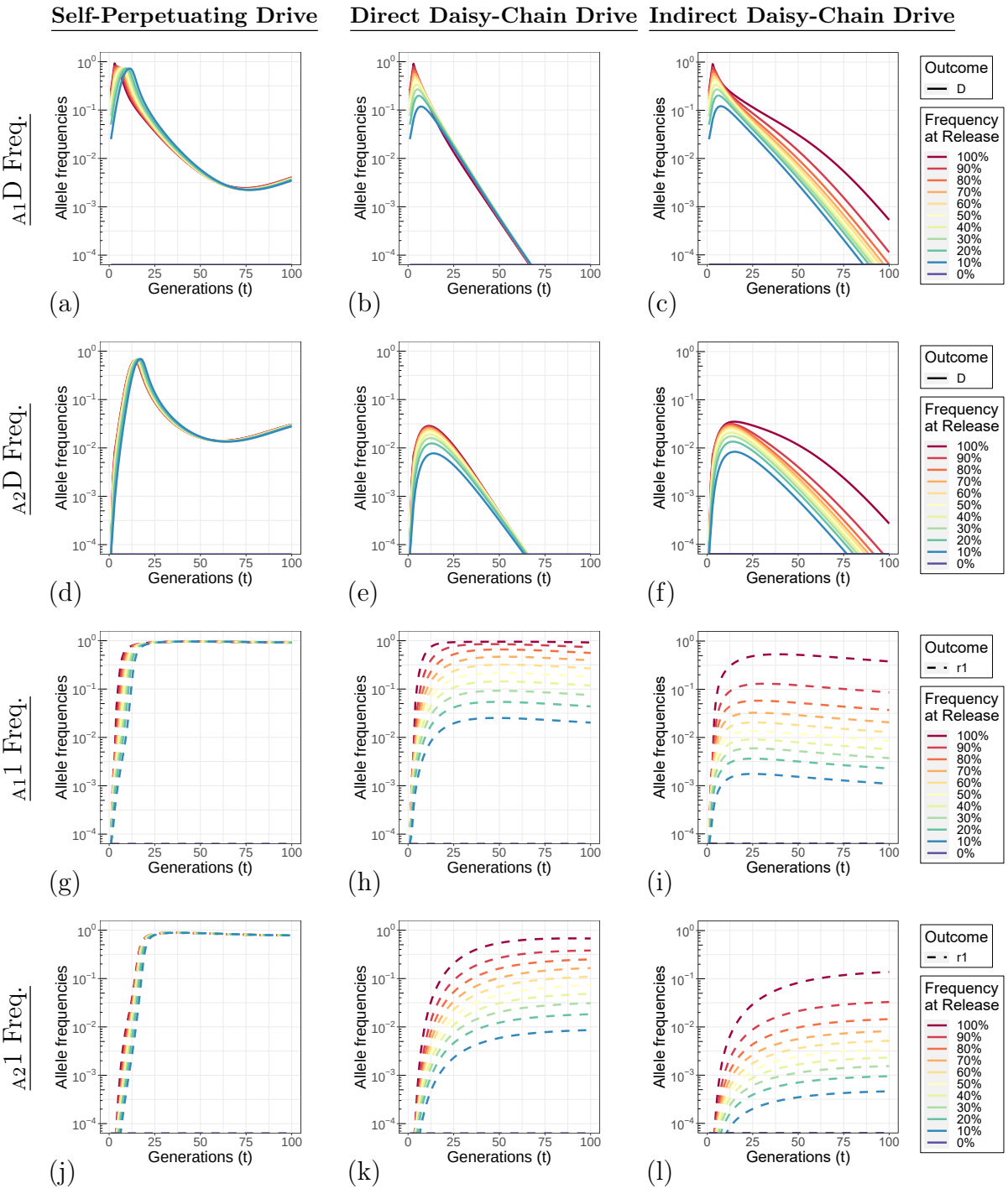


Figure S3.8: **Male release frequency parameter sweep with a recessive lethal A drive element. Row 1 (a-c).** Allele dynamics of the A locus drive element in population one. **Row 2 (d-f).** Allele dynamics of the A locus drive element in population two. **Row 3 (g-i).** Allele dynamics of the A locus type-1 resistance mutations in population one. **Row 4 (j-l).** Allele dynamics of the A locus type-1 resistance mutations in population two. **Column 1.** Self-Perpetuating Drive. **Column 2.** Direct Daisy-Chain Drive. **Column 3.** Indirect Daisy-Chain Drive. A heterozygous release frequency of 100% means that all males in population one are heterozygous drive carriers at the start of the simulation. This equates to a drive allele frequency of 50% among males and a frequency of 25% among all alleles when females are included (the y-axis value shown).

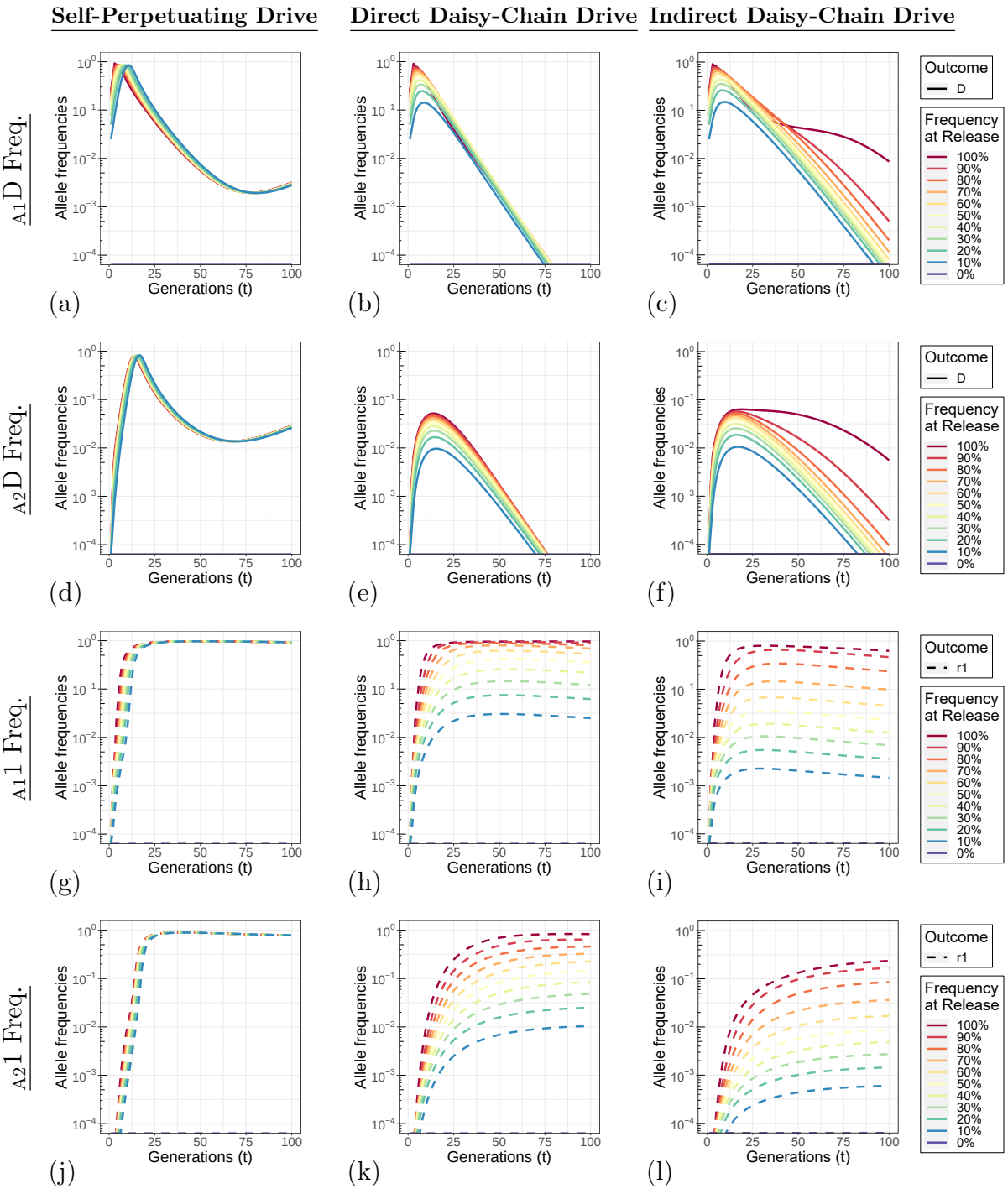


Figure S3.9: **Male release frequency parameter sweep with a female-specific recessive lethal A drive element. Row 1 (a-c).** Allele dynamics of the A locus drive element in population one. **Row 2 (d-f).** Allele dynamics of the A locus drive element in population two. **Row 3 (g-i).** Allele dynamics of the A locus type-1 resistance mutations in population one. **Row 4 (j-l).** Allele dynamics of the A locus type-1 resistance mutations in population two. **Column 1.** Self-Perpetuating Drive. **Column 2.** Direct Daisy-Chain Drive. **Column 3.** Indirect Daisy-Chain Drive. A heterozygous release frequency of 100% means that all males in population one are heterozygous drive carriers at the start of the simulation. This equates to a drive allele frequency of 50% among males and a frequency of 25% among all alleles when females are included (the y-axis value shown).

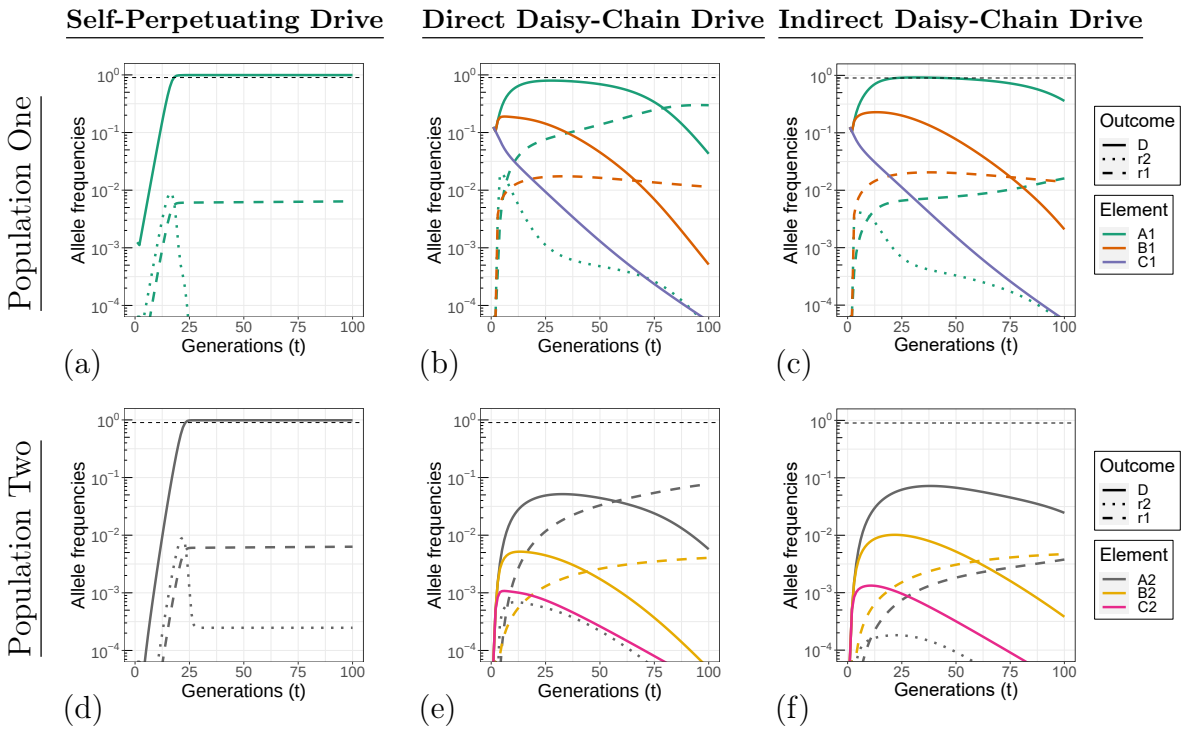


Figure S3.10: **Individual allele dynamics with a germline cut-rate of 85%. Row 1 (a-c).** Individual allele dynamics for population one. **Row 2 (d-f).** Individual allele dynamics for population two. **Column 1.** Self-Perpetuating Drive. **Column 2.** Direct Daisy-Chain Drive. **Column 3.** Indirect Daisy-Chain Drive. The thin dashed line indicates a frequency of 90%.

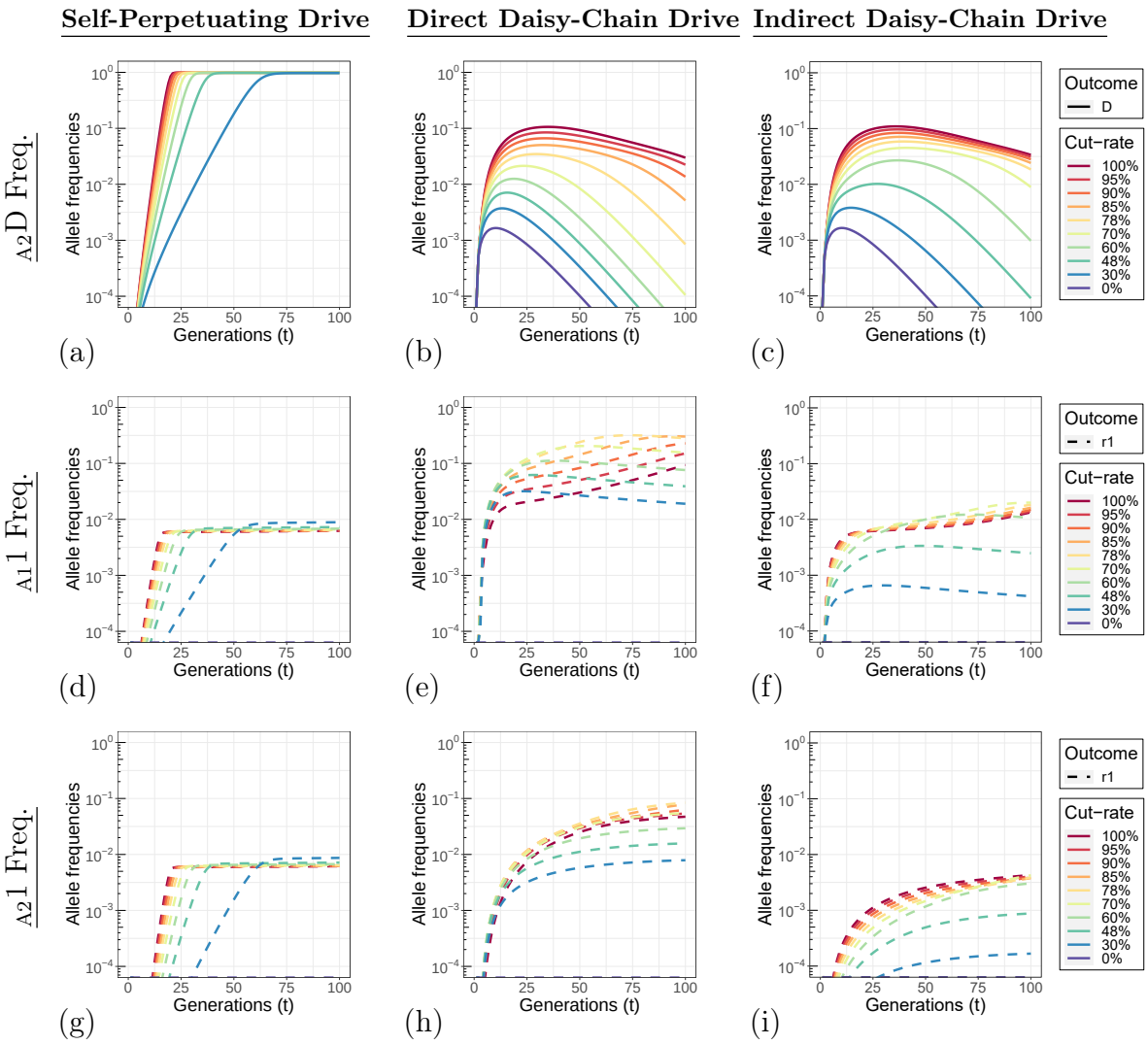


Figure S3.11: **Germline cut-rate parameter sweep with no HDR repair of cut TT genotypes. Row 1 (a-c).** Allele dynamics of the drive element at the A locus in population two. **Row 2 (d-f).** Allele dynamics of type-1 resistance mutations at the A locus in population one. **Row 3 (g-i).** Allele dynamics of type-1 resistance mutations at the A locus in population two. Allele dynamics of the drive element at the A locus in population one are shown in Fig 3d-3f. **Column 1.** Self-Perpetuating Drive. **Column 2.** Direct Daisy-Chain Drive. **Column 3.** Indirect Daisy-Chain Drive.

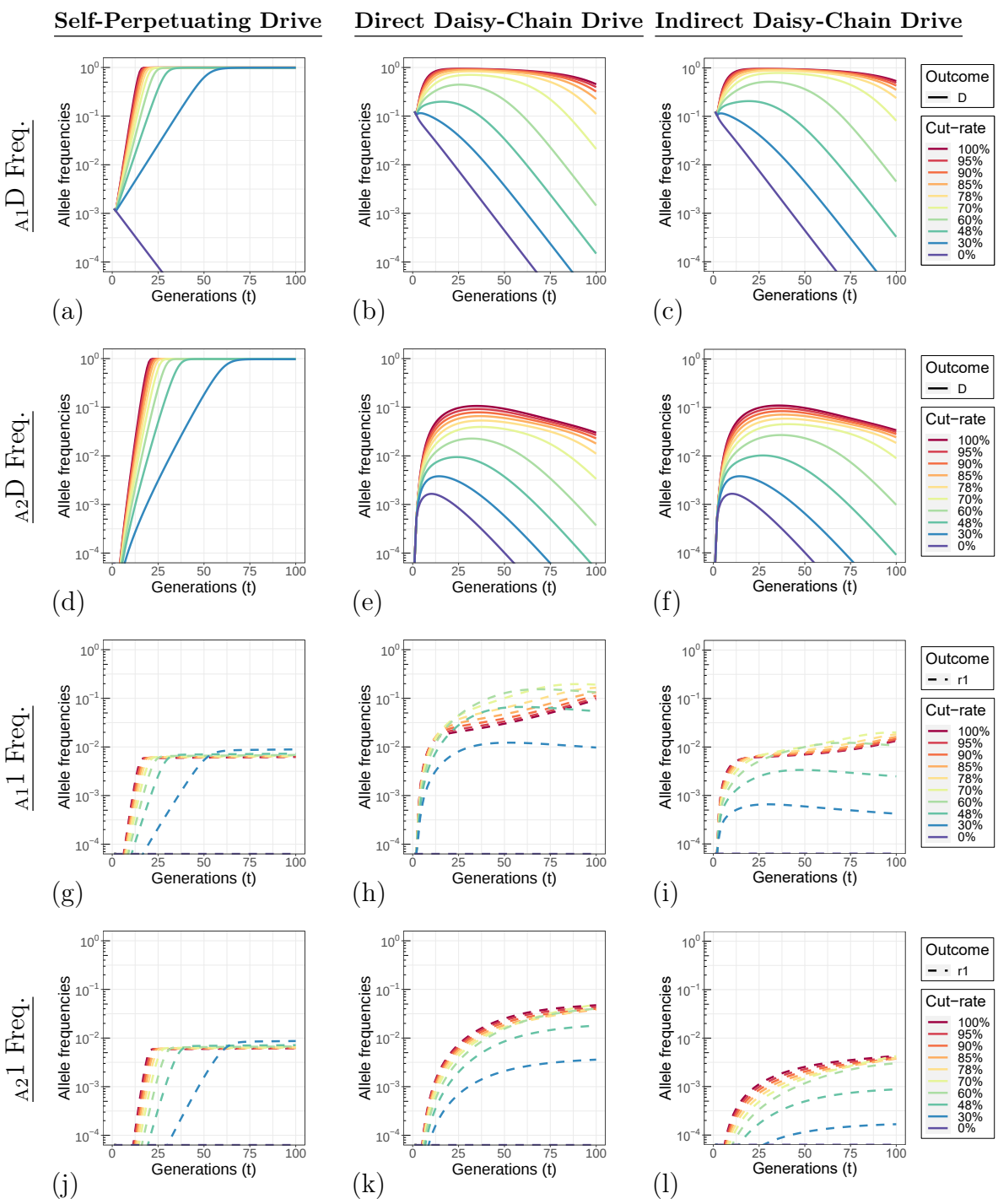


Figure S3.12: **Germline cut-rate parameter sweep with sequential cutting of TT genotypes. Row 1 (a-c).** Allele dynamics of the A locus drive element in population one. **Row 2 (d-f).** Allele dynamics of the A locus drive element in population two. **Row 3 (g-i).** Allele dynamics of the A locus type-1 resistance mutations in population one. **Row 4 (j-l).** Allele dynamics of the A locus type-1 resistance mutations in population two. **Column 1.** Self-Perpetuating Drive. **Column 2.** Direct Daisy-Chain Drive. **Column 3.** Indirect Daisy-Chain Drive.

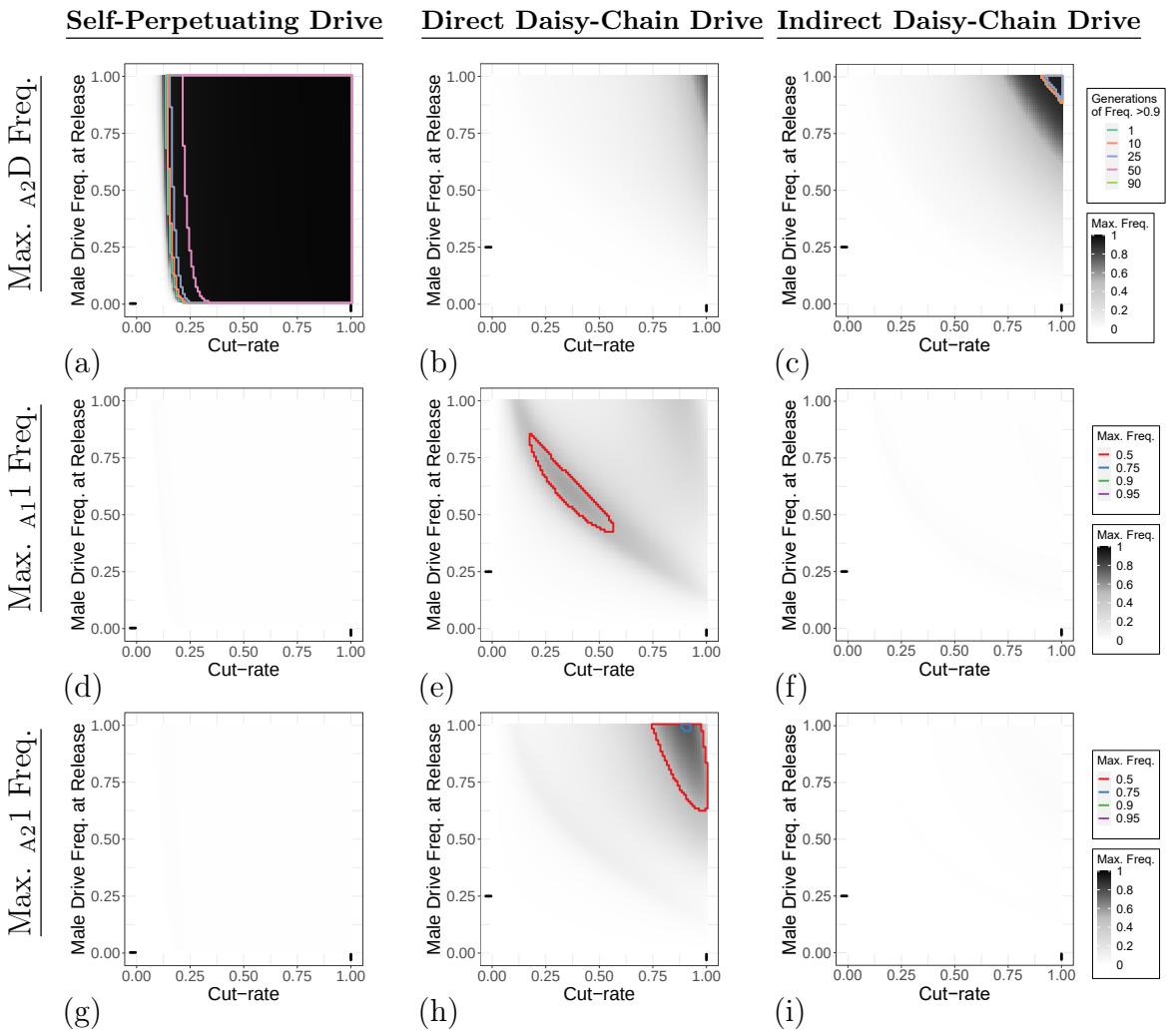


Figure S3.13: **Simultaneous changes to the germline cut-rate and male drive carrier frequency at release. Row 1 (a-c).** Maximum $_{A}$D allele frequency in population two. **Row 2 (d-f).** Maximum $_{A}$1 allele frequency in population one. **Row 3 (g-i).** Maximum $_{A}$1 allele frequency in population two. The maximum frequency of $_{A}$D in population one and the difference between population one and two is shown in Fig 3.

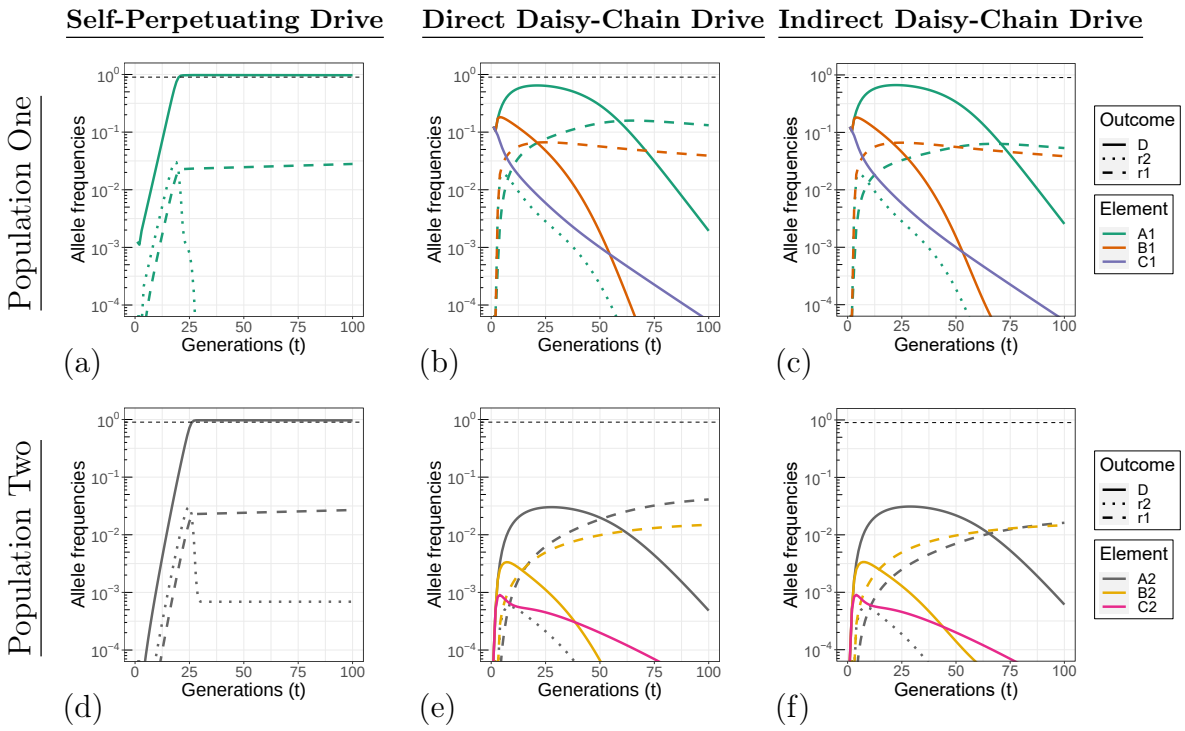


Figure S3.14: **Individual allele dynamics with a maternal deposition cut-rate of 25%. Row 1 (a-c).** Individual allele dynamics for population one. **Row 2 (d-f).** Individual allele dynamics for population two. **Column 1.** Self-Perpetuating Drive. **Column 2.** Direct Daisy-Chain Drive. **Column 3.** Indirect Daisy-Chain Drive. HDR does not occur with deposition-mediated cutting, and all cuts result in resistance mutations following the 1:9 ratio of type-1 to type-2. The thin dashed line indicates a frequency of 90%. The dynamics of the direct and indirect DCD ${}_{B1}$D elements may be more similar than initially expected from the genetics. However, when the DCD B elements are together with a drive element at A, there is no difference in deposition between the direct (${}_{A}$E/${}_{B}$N^{A}) and indirect (${}_{A}$G^{A}/${}_{B}$N) DCDs. Only when the B element is isolated from the A drive element does deposition affect the two designs differently. This only occurs when the B element segregates away with a type-1 resistance allele (100% expression-based cut-rate, and type-2 resistance alleles are lethal). Although the resistance allele rate is substantially higher with deposition, isolation of ${}_{B}$D from ${}_{A}$D is still a rare event. Moreover, due to the high ${}_{A}$D frequency, isolated B elements are rapidly reacquainted with a A drive element. Isolated ${}_{B}$D elements are much less likely to be reacquainted with a ${}_{A}$D element in population two, which is reflected in a more pronounced difference in ${}_{B}$D allele dynamics between the indirect and direct DCDs.

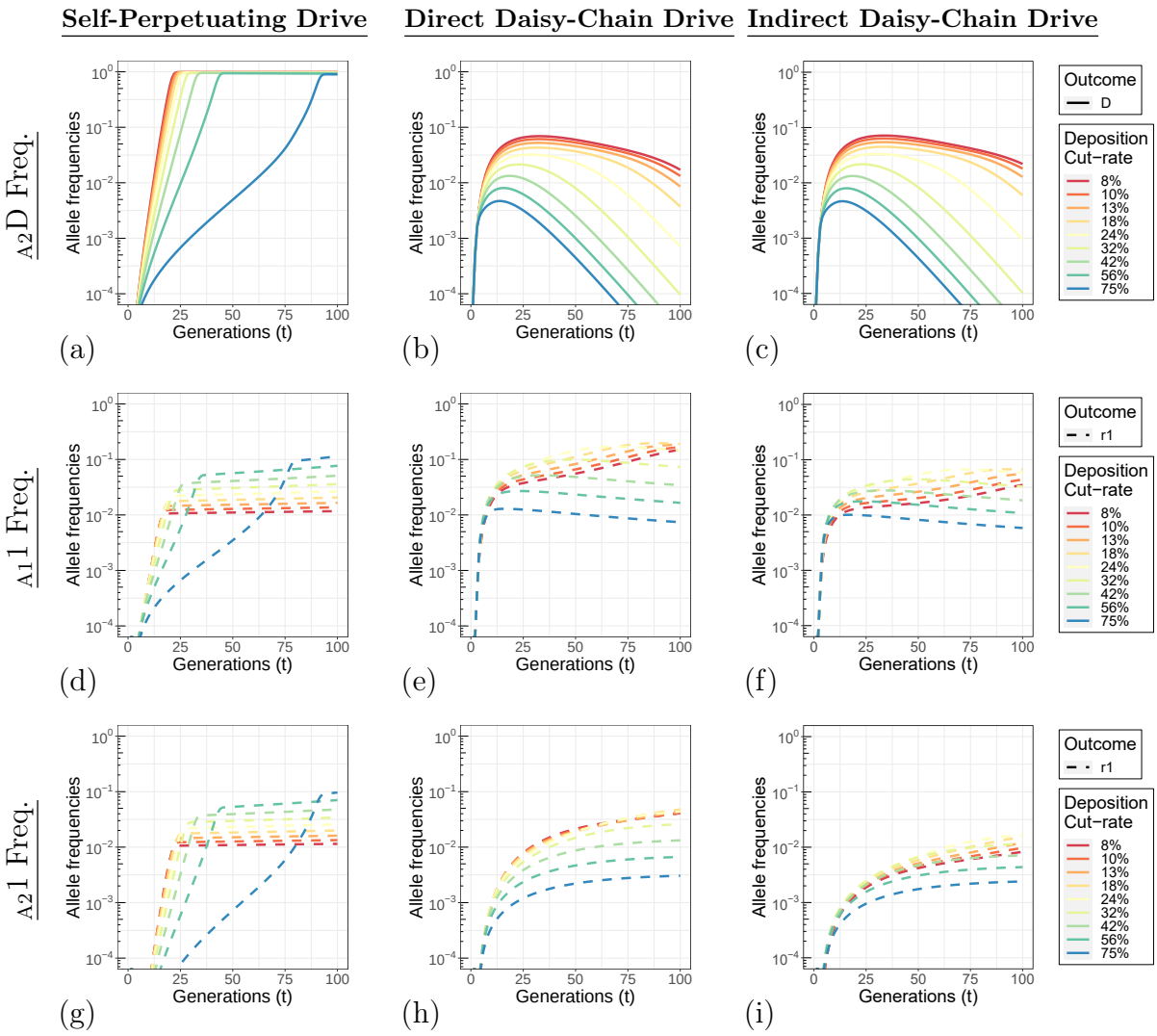


Figure S3.15: **Maternal deposition cut-rate parameter sweep. Row 1 (a-c).** Allele dynamics of the drive element at the A locus in population two. **Row 2 (d-f).** Allele dynamics of type-1 resistance mutations at the A locus in population one. **Row 3 (g-i).** Allele dynamics of type-1 resistance mutations at the A locus in population two. Allele dynamics of the drive element at the A locus in population one are shown in Fig 4d-4f. **Column 1.** Self-Perpetuating Drive. **Column 2.** Direct Daisy-Chain Drive. **Column 3.** Indirect Daisy-Chain Drive. HDR does not occur with deposition-mediated cutting, and all cuts result in resistance mutations following the 1:9 ratio of type-1 to type-2.

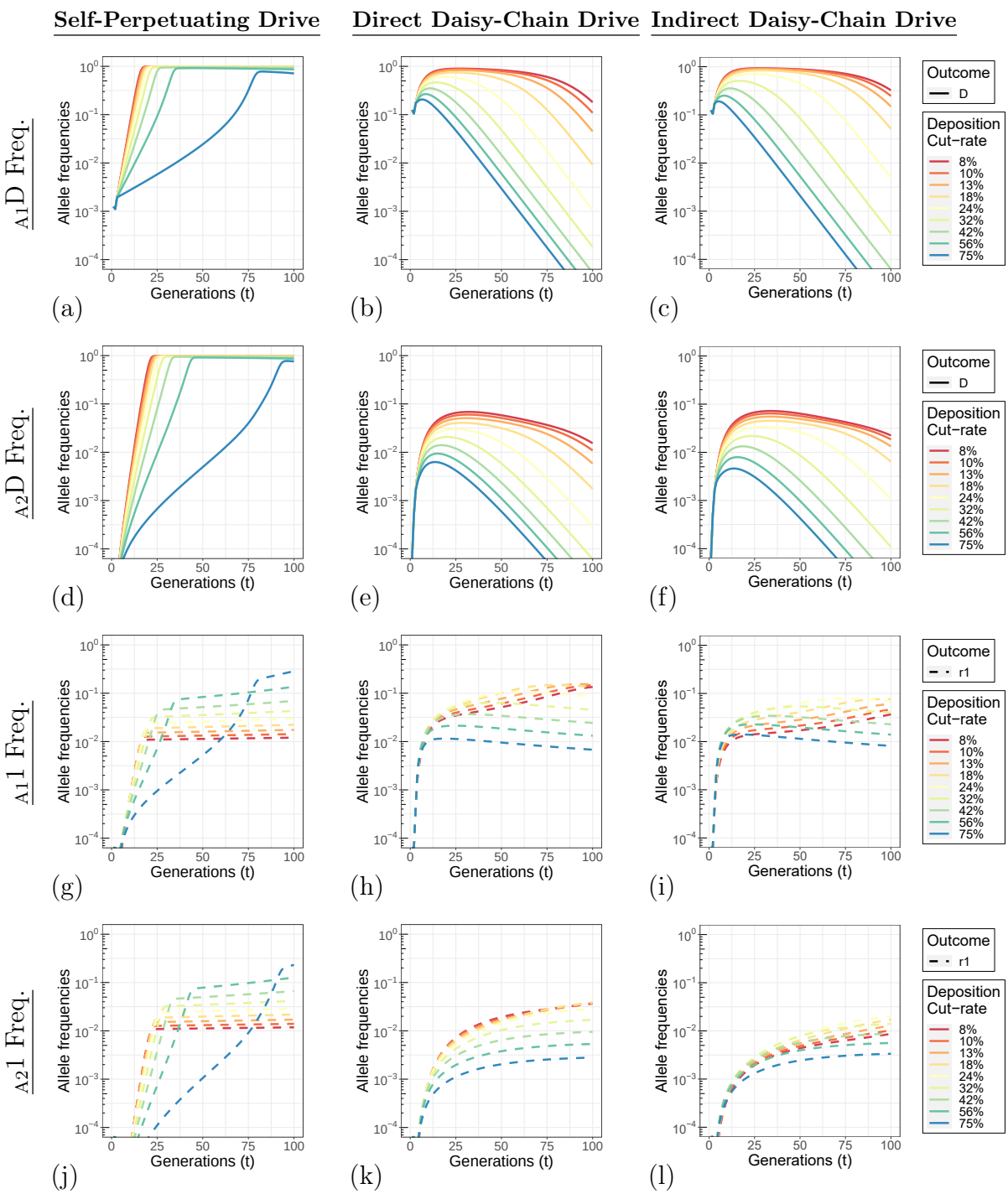


Figure S3.16: **Maternal deposition cut-rate parameter sweep with no gRNA deposition.** In these simulations only the Cas9 protein is deposited. **Row 1 (a-c).** Allele dynamics of the A locus drive element in population one. **Row 2 (d-f).** Allele dynamics of the A locus drive element in population two. **Row 3 (g-i).** Allele dynamics of the A locus type-1 resistance mutations in population one. **Row 4 (j-l).** Allele dynamics of the A locus type-1 resistance mutations in population two. **Column 1.** Self-Perpetuating Drive. **Column 2.** Direct Daisy-Chain Drive. **Column 3.** Indirect Daisy-Chain Drive. HDR does not occur with deposition-mediated cutting, and all cuts result in resistance mutations following the 1:9 ratio of type-1 to type-2.

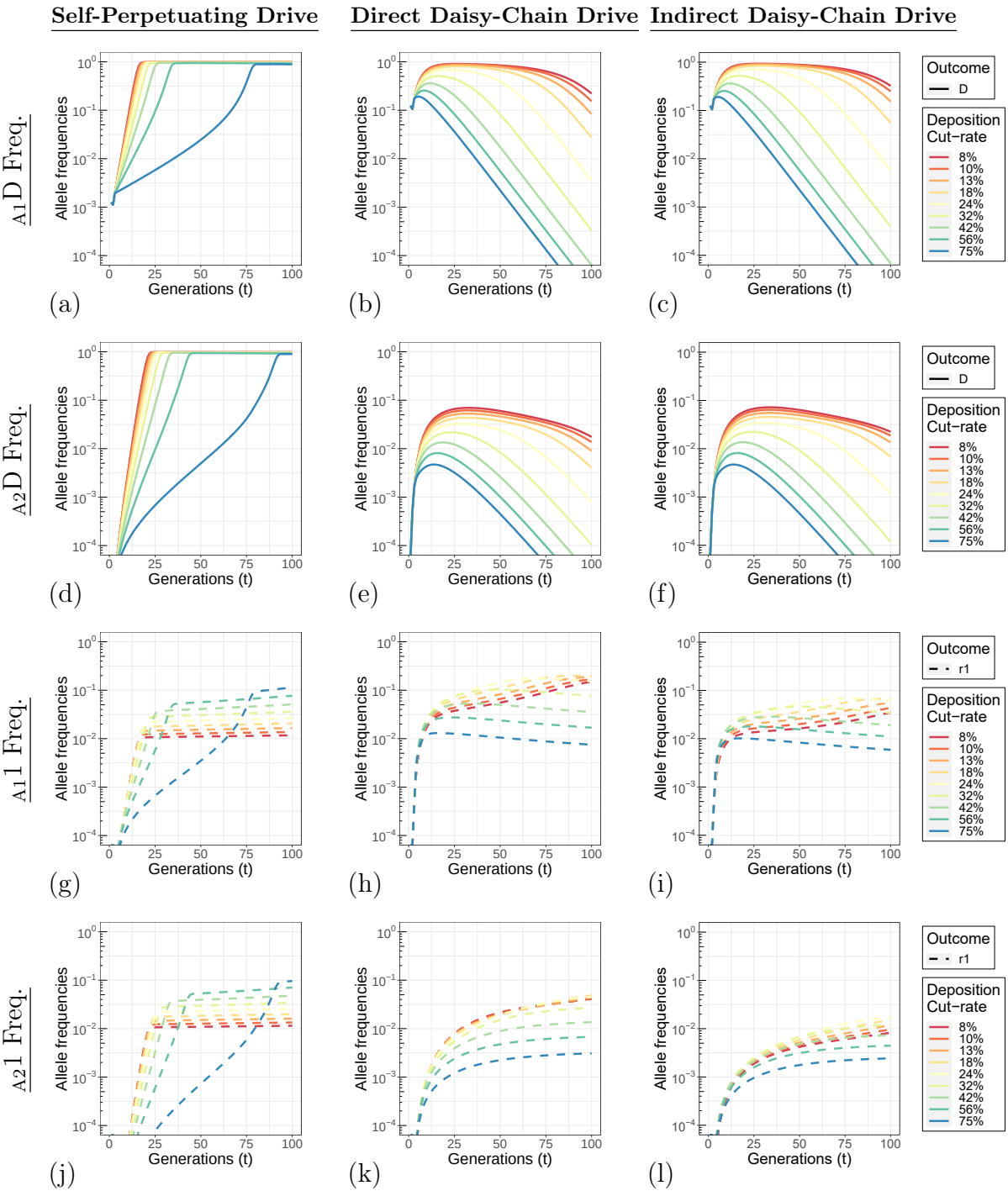


Figure S3.17: **Maternal deposition cut-rate parameter sweep with shadow drive.** In these simulations, when the deposited Cas9 protein is paired with an expressed gRNA, DNA repair uses the germline HDR rate. If Cas9 and a gRNA are deposited simultaneously, this takes precedence, and HDR repair is not possible. In practice, shadow drive occurs only when the mother carries the Cas9 and the father provides the gRNA gene. Note that this scenario cannot occur with the self-perpetuating drive as the Cas9 and gRNA genes are always linked. **Row 1 (a-c).** Allele dynamics of the A locus drive element in population one. **Row 2 (d-f).** Allele dynamics of the A locus drive element in population two. **Row 3 (g-i).** Allele dynamics of the A locus type-1 resistance mutations in population one. **Row 4 (j-l).** Allele dynamics of the A locus type-1 resistance mutations in population two. **Column 1.** Self-Perpetuating Drive. **Column 2.** Direct Daisy-Chain Drive. **Column 3.** Indirect Daisy-Chain Drive.

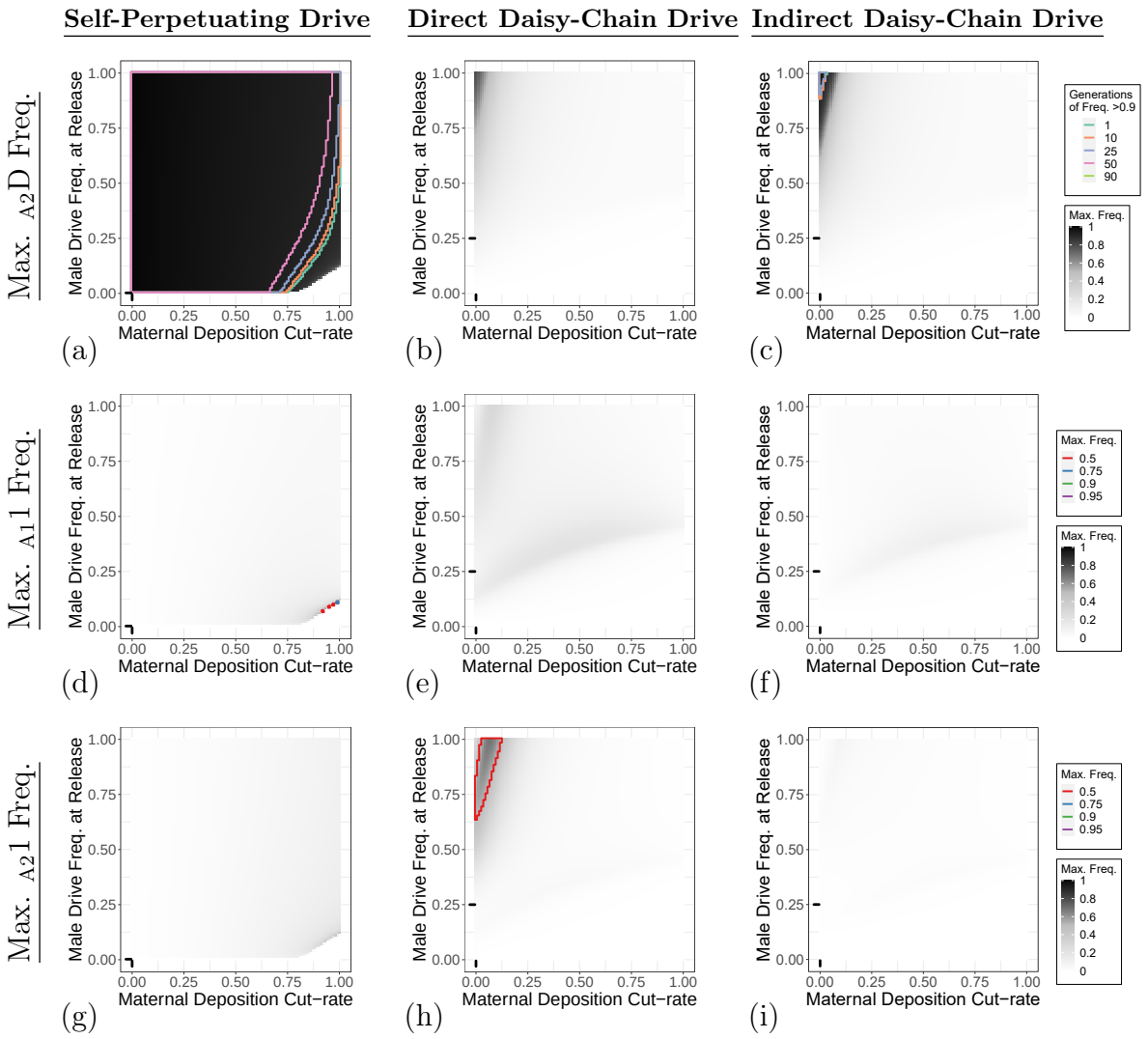


Figure S3.18: **Simultaneous changes to the maternal deposition cut-rate and male drive carrier frequency at release. Row 1 (a-c).** Maximum $_{A}$D allele frequency in population two. **Row 2 (d-f).** Maximum $_{A}$1 allele frequency in population one. **Row 3 (g-i).** Maximum $_{A}$1 allele frequency in population two. The maximum frequency of $_{A}$D in population one and the difference between population one and two is shown in Fig 4. HDR does not occur with deposition-mediated cutting, and all cuts result in resistance mutations following the 1:9 ratio of type-1 to type-2.

Chapter 4

Sex and gene drive: application of a large-scale homing gene drive database and analysis pipeline

Sebald A. N. Verkuijl[1,2], Philip T. Leftwich[3], Edward Ivimey-Cook[4], Michael B. Bonsall[1]

1: Mathematical Ecology Research Group, Department of Zoology, University of Oxford, 11a Mansfield Road, Oxford, OX13SZ, U.K
2: Arthropod Genetics, The Pirbright Institute, Ash Road, Pirbright GU24 0NF, U.K.
3: School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, U.K.
4: Institute of Biodiversity, One Health and Veterinary Medicine, University of Glasgow, Glasgow, G61 1QH, U.K.

In the previous chapter, we performed a computational analysis of different gene drive designs. Through parameter sweeps, we evaluated gene drive outcomes under a range of values for a selection of parameters. This provides important insight into the challenges facing specific gene drive designs and what parameters may specifically be important to measure and optimise. However, we were unable to draw specific conclusions about what this implies for the gene drives that have been implemented so far. Furthermore, we found that the ways in which deposition can present can be exceedingly complex. By varying one or two parameters at a time, we have not covered the specific combination of values of different parameters that occur with the *in vivo* gene drives. We have also not considered the complexity of measuring each parameter or whether there may be alternative measures that may be more readily derived from experimental work. In this chapter, we develop a database and an analysis pipeline to evaluate the experimental performance of published homing gene drives. This will allow us to place our computation results in context and adapt them to the specific strengths and weaknesses of the experimental results.

4.1 Abstract

Synthetic gene drives have the potential to control harmful pests by amplifying genetic modifications through successive generations. Extensive research has gone into implementing homing endonuclease gene drive based on programmable nucleases such as CRISPR-Cas9. Understanding the factors that influence the function of homing gene drives is crucial to be able to predict their behaviour in wild populations through computational modelling. In addition, making experimental data more accessible and systematically analysing the experimental, biological, and transgene design factors will aid their further development. Here, we describe the development of a homing endonuclease meta-analysis pipeline applied to 238 crosses and 327,706 individually scored insects (a subset of the publications returned by our literature search). We applied this analysis framework to investigate the effect of sex on two results of the gene drive process. We found a small but significant increase in gene drive inheritance rates from female drive carriers compared to males (parental germline), but no significant difference in somatic phenotype rates between equivalent males and females (progeny somatic tissue). Nuclease activity in somatic tissues was substantially higher in the progeny of female nuclease carriers (in almost all cases with the gRNA) compared to males, indicative of maternal deposition. However, no significant effect on gene drive propagation rates where found when the nuclease-carrying grandparent was female (and in many cases, the gRNA gene was carried by the father). We developed a companion interactive tool available from: `https://osf.io/k9fbd/`. This tool has been designed to allow users to identify the highest quality data for a specific research question and investigate the large and growing body of gene drive literature.

4.2 Introduction

Gene drive is a process through which the inheritance of a particular genetic region can be biased. This region can comprise multiple genes or even a whole chromosome. Through gene drive, the biased region will be passed onto a larger proportion of offspring than would be expected by the Mendelian rules of inheritance. Synthetic gene drive-based technologies are predicted to be able to spread a particular trait throughout a wild population or achieve localised suppression of a wild pest population more cost-effectively[64]. These technologies can potentially be transformative in the areas of disease vector control, agriculture, and conservation.

Homing gene drives are a type of gene drive that functions by copying a specific allele from one chromosome to the homologous chromosome. Cells in which this process occurs will go from being heterozygous to being homozygous for the "drive" allele. If these cells are part of the germline lineage, the allele will be passed to a higher proportion of the progeny.

In the field of gene drive research, this copying process is known as “homing” and is mediated by homology-directed repair (HDR). In a broader context, the DNA processes that underlie homing, or similar outcomes, can be referred to as "interchromosomal

HDR"[75], "loss of heterozygosity"[184], "allelic exchange"[185], "gene conversion"[186], and others. Each of these terms has nuanced and context-dependent uses and indicates a broad, and somewhat parallel, study of a shared process. The DNA repair response associated with homing occurs naturally and is associated with meiotic recombination and endogenous endonucleases such as Spo11[187;188].

In the context of synthetic gene drives, homing is induced by the deliberate induction of DNA damage at the site allelic to the drive allele. This DNA damage then activates repair pathways that mediate gene conversion or, as is often the case, other unintended repair outcomes. With the advent of the CRISPR-Cas9 programmable endonuclease gene editing tool[65], development of homing drives has boomed[189]. However, a substantial amount of fundamental homing drive research was first performed with fixed-sequence meganucleases such as I-SceI[68;78;85;190] and early programmable gene editing technologies[87].

Cas9 is a two-component nuclease composed of a programmable single guide RNA (sgRNA) and the Cas9 protein. The first reports of CRISPR-Cas9-based homing drives in *Drosohopila melanogaster*[88], *Saccharomyces cerevisiae*[60], *Anopheles stephensi*[53], and *Anopheles gambiae*[47] reported extremely high inheritance bias rates. This rapid success suggested that efficient Cas9-based homing drives were easy to engineer in a variety of species. However, subsequent research has indicated that inheritance bias rates in yeast and *Anopheles* mosquitoes were outliers in their immediate high efficiency. More recent implementation in *D. melanogaster* and other species have been less efficient before substantial optimisation efforts[1].

Much of the research in the field of homing gene drive has focused on understanding what factors contribute to their performance and implementing different optimisation strategies. A complication of this approach is that many changes are made simultaneously to the gene drive transgene designs in different implementations, and many design features are often not made explicit. Additionally, there are differences in the methodological, analysis, and reporting approaches that make it difficult to compare the results between publications without a detailed investigation.

An important factor in the variability of inheritance bias levels of the homing gene has been attributed to the developmental timing and the cell type in which DNA repair occurs[90]. Technical developments have attempted to limit nuclease activity to when gene conversion is favoured, supposedly in meiotic cells[49;50;64;75;81;93;94;97;106–110]. This is approached by testing a panel of different transgene designs with different germline regulatory elements to express the nuclease protein gene[3;5;49;92;93;95;96;99;118;129]. Other gene drive design features are often varied between implementations, but have received limited systematic investigation. These include the number of gRNAs[81;110;152;191], the gRNA promoter[95], the target gene[47;96;103], the target site within the same gene[90], and the transgene insertion site[5;50].

In addition to experimental work, a substantial part of gene drive research takes the form of computational modelling[192]. In a minority of cases, this is directly informed by experimentally measured parameters (e.g., HDR rate, germline resistance

rate, deposition rate, etc.)[84;135;193]. Modelling can inform what parameters are most important for drive performance, and thereby direct genetic design optimisation. Moreover, computational modelling can be used to prioritise experimental work to measure drive performance to discard non-viable candidates as early and efficiently as possible. Most importantly, modelling is used to predict the ultimate results of a gene drive release based on data collected in controlled experiments. However, currently, modelling is generally done in parallel to experimental work using crude summary statistics of drive performance and does not capture the richness, nor the shortcomings, of commonly measured experimental data. A parameter based on 100s of observations may be treated the same as a parameter that was estimated based on fewer than 10 individuals. A more direct connection between the measured data and computational models would aid both in directing gene drive development and the ultimate representativeness of modelling efforts.

Here, we report a database structure and analysis pipeline for large-scale analysis of homing endonuclease gene drive performance and design characteristics. We have developed a companion web tool that allows users to perform custom analyses and review in-depth the data of individual crosses. We applied meta-analysis tools to a subset of available data to investigate the effect of sex on gene drive outcomes.

4.3 Results

4.3.1 Literature search

We first performed a literature search to produce a list of publications for subsequent data extraction (Fig 4.1). The goal was to produce a list of publications for which we could not, without a detailed examination, exclude that they report relevant data. The data we aimed to include was that of the controlled breeding (crosses) of male and female individuals that is informative of the homing process in regard to a homing gene drive. Individuals of each sex that are crossed should have a defined genotype with respect to the relevant nuclease target locus and nuclease transgenes. This does not mean that the experiment must necessarily have involved a gene drive in a strict sense, instead that the underlying process would be directly analogous to that of a homing endonuclease gene drive (e.g., split drives).

We used a non-systematic database of publications involving experimental data of homing gene drives to validate the thoroughness of our systematic search process. This prior database of 65 papers (Fig 4.1a) was accumulated over multiple years through a combination of literature alerts, personal communication, and social media. The systematic literature search described below aims to expand upon this list and correct for potential biases.

Relevant publication records were collected by querying PubMed with the search terms listed in Fig 4.1b. Duplicate records were removed as were records already present in the prior database (Fig 4.1c-g). The remaining records were manually investigated for relevance and when records were excluded, we noted the reasons and degree of

investigation performed. The number of records and the reason for exclusion are summarised in Fig 4.1h-i. Finally, records from the prior database and newly identified publications were compiled into a single database listed in Table S4.1 (Fig 4.1j-l).

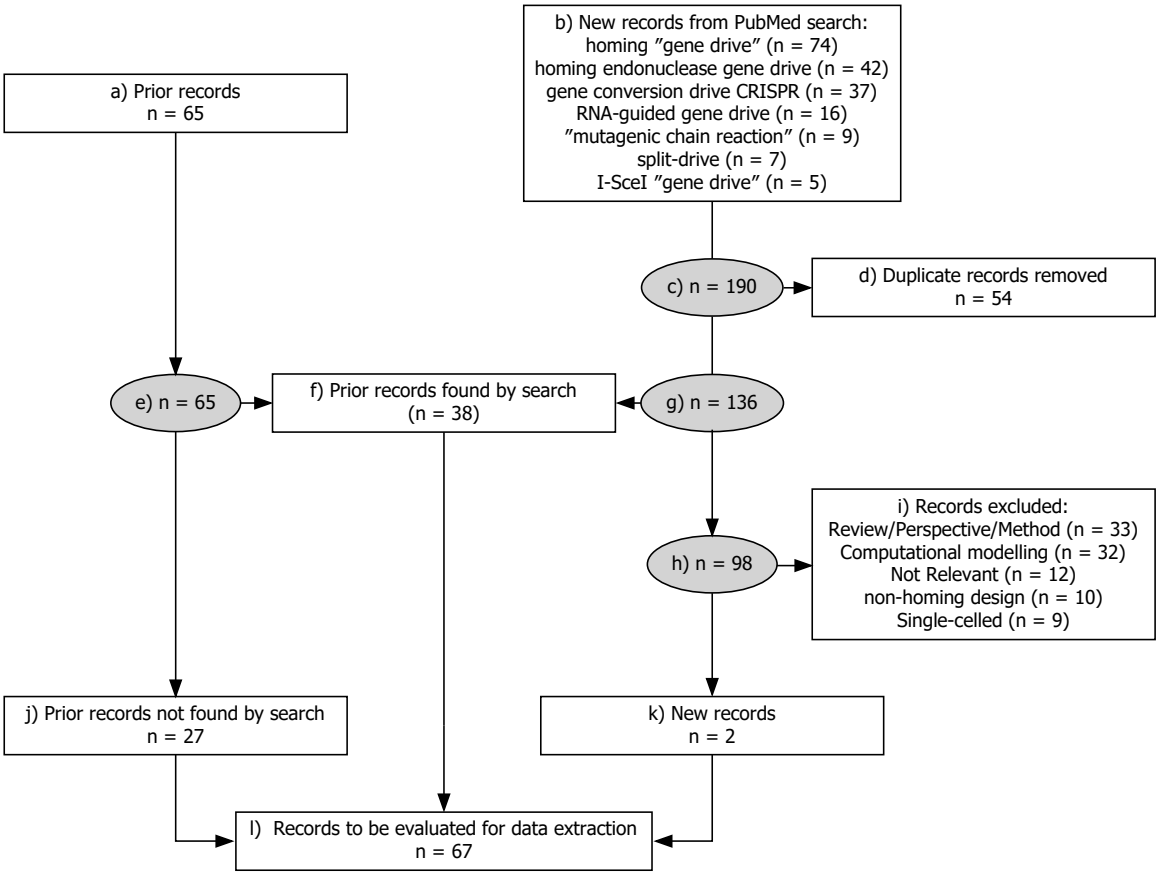


Figure 4.1: **Literature search pipeline and validation through a prior database of relevant publications.** (a) 65 records were present in a non-systematic database of publications accumulated over multiple years from various sources. This database comprises seven species, and more than half of the publications have been reported in *D melanogaster*. (b) Search terms used in PubMed literature database, and the number of results obtained from a search in January 2023. (c) 190 records were returned by the search. (d) 54 records were returned by more than one search term and removed as duplicates. (e) The 65 records in the prior database were already unique and were not processed further. (f) 38 of the records identified by the literature search were already present in the prior database. (g) 136 unique records were returned by the search, including those already present in the prior database. (h) 98 unique records from the search did not overlap with the prior database. (i) An overview of the records returned by the search that were excluded after individual evaluation and the reason why. (j) 27 records from the prior database were not identified by the literature search. (k) 2 records were identified by the literature search that were not excluded before the data extraction process. (l) 67 records were considered sufficiently relevant to merit a detailed analysis and data extraction.

Our search only identified two new publications (Fig 4.1k) that were not already in the prior database, in both cases non-homing gene drive studies that may nevertheless contain qualifying data[61;194]. However, more than 40% (27/65) of the publications in the prior database were not identified by the literature search at all. This strongly indicates that the search terms need to be expanded for a comprehensive overview of

the gene drive literature. In addition, while our focus for this meta-analysis is the recent field of engineering synthetic homing gene drives, the DNA repair process underlying homing has also been the focus of substantial fundamental research. There may be experimental work in fundamental DNA repair research that is directly analogous to gene drive crosses that should be included[101;167;195–200]. It should be noted that these publications are particularly difficult and time consuming to evaluate due to parallel terminology and differing reporting standards. Table S4.2 lists an expanded set of search terms that can be used to more thoroughly identify the relevant literature. However, the goal of this chapter is to develop the database structure and analysis pipeline for which the prior database and the narrower search process provided enough candidate publications.

A total of 67 publications passed our initial general assessment of relevance. To combine these data, specific requirements are placed on the quality and format of underlying data, and a subset of publications that were included in the literature search will be found to contain no data that meets these criteria. We have split the literature and data extraction processes to maintain the reproducibility of the former.

We found data extraction a more subjective task than the publication search, as few quantifiable reference points exist. For example, while some data inclusion criteria, such as defining the transgene, were clear; identifying the number of specified crosses for inclusion within a publication was highly subjective, indicating a need for standardised data representation.

For this thesis chapter, we made a non-random selection of the publications from our literature search to develop a database and analysis framework on. We specifically selected publications with diverse, and non-standard methods to ensure that they could be accommodated in addition to the more common experimental methods. As such, the selection of publications should not be taken as fully reflective of the general body of homing gene drive literature.

4.3.2 Data inclusion criteria and database structure

To analyse the performance of homing endonuclease gene drives, we focused on data derived from the scoring of progeny generated from defined crosses. This is a commonly reported measure, relatively standardised, and can be used to estimate many parameters of the inheritance biasing process in relative isolation (e.g., in absence of mating competition). We selected publications from the prior publication list and compiled their data into a single database through an iterative process, frequently returning to previously processed publications based on later improvements to the data structure. Through this process, we developed a set of inclusion criteria for the data:

- **Heterozygous**: At least two alleles are described for the same locus, one susceptible to induced DNA damage and one resistant.
- **Defined transgene**: DNA damage is expected to be generated at the suscep-

tible allele through the product of the integrated transgene(s). This excludes exogenously supplied nuclease components (e.g., micro-injection).

- **Heterozygous cross**: A genetic cross is performed that may be expected to result in inheritance bias of a defined allele through homing.
- **Defined cross**: Individuals with a defined genotype for the target allele and transgenes are crossed. This generally excludes multigenerational cage trials.
- **Germline assessed**: The result of the DNA repair process is evaluated by scoring the progeny of the individuals in whom it is expected to occur.
- **Countable**: Individual progeny counts are reported, and an inference can be made about the status of the allele of interest.

Data that matched these criteria were compiled into a single database. In some cases, additional crosses are reported using the same transgenes that do not match the above criteria. They may nevertheless be informative as they provide a reference, or baselines, for certain parameters. As such, if a cross is identified that meets the above criteria, additional crosses with the same transgenes may be included. The database is structured around a list of "N" F2 progeny with the most descriptive possible grouping that can be made based on over 90 metadata fields (Fig 4.2a and Supplemental methods 4.7). To increase the reproducibility of our database, we have included a brief justification for our interpretation or a source for the chosen values for each factor in each cross. In addition, when we cannot report on a particular metric, we have included an additional brief justification. Individual cross-data can be most readily evaluated on the cross-over page of the interactive web tool.

An important part of the data is a description of the different genotypes involved in a cross. We focus on the core nuclease components, ignoring other genes such as fluorescent proteins. Examples of simplified genotype descriptions are given for a two-locus split drive (Fig 4.2b), and two self-perpetuating drives (Fig 4.2c-d). We record the genotypes for three generations: the F2s that are scored and counted, the F1 generation in whose germline the gene conversion process is expected to occur (D1) and their mate (T1), and the F0s that could affect the outcomes by the deposition of nuclease components (D0.Male, D0.Female, T0.Male, and T0.Female). These descriptions are simply a shorthand for the generations relative to the scored F2s, and the same individuals may be the F2 in one cross and the F1 in another. An example of a common cross is given in Fig 4.2e.

In general, the most important observation made of a single cross is the number of individuals with different F2 zygotic genotypes that emerge. This distribution of F2 zygotic genotypes will reflect the action of the drive in the "D1" parent. A complication is that the F2 genotype may change through development through the action of the nuclease components. A common example of this is that a mutated target allele may have been inherited in a mutated form from the D1 parent's germline or become mutated in the F2 due to deposition or expression of the nuclease components. To standardise the reported outcomes, we report the inferred (ambiguous) zygotic

		Cross specification				Outcomes
	Publication metadata	Experimental Factors	Targeting factors	Nuclease factors	Parent and grandparent factors	F2 Factors
Description:	General information about the publication independent of the content.	Information about general experimental conditions.	Characteristics of the gRNA and target sites organised by target site.	Characteristics of the nuclease expressing transgene.	A description of the genotype and other factors of the drive carrying parent and grandparents.	Characteristics of the scored progeny. Includes the parent of origin and the number of scored individuals.
Example:	DOI	Developmentalstage scored	gRNA target sequence	Nuclease promoter&5'UTR	Drive parent sex	Progeny sex
Consistency:	Identical for all data from a single publication.	Usually consistent for all crosses within a publication.	Often identical over all crosses in a publication	Often a defining alteration that differentiates otherwise identical crosses.	Often identical between tests of different genetic designs in the same publication.	Differs within a cross, but the same grouping of progeny will occur in different crosses.

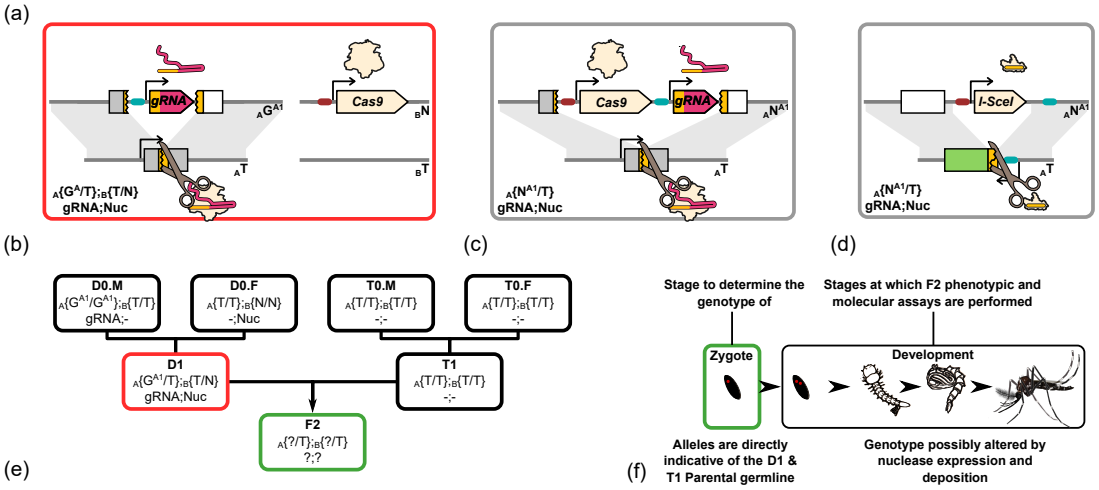


Figure 4.2: **Database structure and genotype notation.** (a) An overview of the different categories of data that are recorded for each cross and how they generally vary. Each individual factor that falls within the categories shown here is discussed in detail in the supplemental methods section 4.7. (b) An illustration of a homing split-drive genetic design. The gRNA and Cas9 genes are located at separate loci (A and B). To the right of each allele illustration is the simplified genotype notation used in the database and described in Fig S4.1 and Verkuijl et al.[2]. (c) A CRISPR-Cas9 self-perpetuating homing gene drive design. (d) An I-SceI meganuclease self-perpetuating homing gene drive design. The targeting sequence of I-SceI is fixed and targets a separately inserted transgene; in this case, a fluorescent protein gene is modified to carry the I-SceI target sequence. For simplicity of notation, we treat the single-gene nucleases (i.e., I-SceI, I-OnuI, ZFNs, and TALENs) as a combined nuclease:gRNA element. (e) An example of the three-generation family tree we record for each cross. In the database, we organise the tree such that the "D1" individual carries the allele that may be subject to inheritance bias. The database is structured around estimating the distribution of alleles passed along by the D1 individual. The D1 individual shown in this example is the same genotype as shown in b. (f) An illustration of the F2 generation, which is scored and used to calculate a range of metrics. We record for each cross the F2 genotype at the zygotic stage which is composed of the alleles inherited from the D1 and T1 parent's germline before any changes occur in the genotype during F2 development. Phenotypic and molecular assays are performed at various stages of F2 development, allowing for nuclease-mediated changes in genotypes, and the death of specific genotypes. As such, commonly performed assays often leave some degree of ambiguity in the exact zygotic F2 genotype which is reflected by our notation system.

genotype of F2s before changes may have occurred during development (Fig 4.2f). We then separately record nuclease-mediated phenotypes in the F2 progeny that may indicate expression- and deposition-based nuclease activity in the F2 progeny.

In addition to target and nuclease alleles, we record other unintended DNA repair outcomes (Fig 4.3a-b). We categorise mutations by their functional consequences for the target gene, with type-1 (1) mutations that change the nuclease target sequence

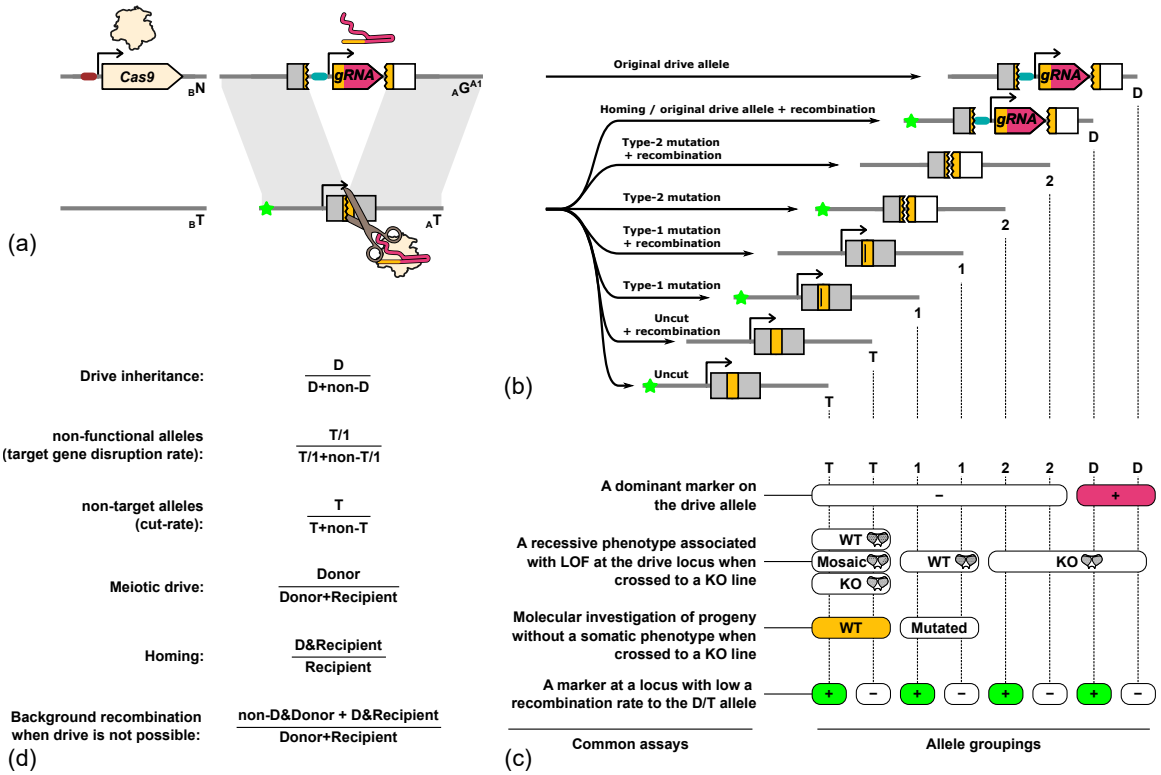


Figure 4.3: **DNA repair outcomes and accounting for ambiguity in common assays and metrics.** (a) An illustration of a split drive element being cut by the Cas9:gRNA complex. The donor chromosome carries the drive allele, which may be biased by gene conversion. The targeted allele on the recipient chromosome carries a linked marker (green star) that is far enough away to be unlikely to be affected by DNA repair through gene conversion. (b) Different alleles that may result from DNA repair and recombination. Mutations are categorised by whether they disrupt the target gene (2), or change the sequence, but leave the target gene functional (1). "D" is used as a catch-all symbol for the drive allele. (c) Common phenotypic and molecular assays can provide information on the zygotic genotype. In most cases, a combination of assays is performed that can inform the identity of the allele that was inherited from the D1 parent. WT = wild-type. KO = knock-out. LOF = loss of function. (d) Examples of the >10 gene drive metrics that we include in our analysis pipeline. These metrics can be useful for extrapolating drive performance and are informed by commonly performed experimental assays.

but leave the gene functionally intact, and type-2 (2) mutations that change the target sequence and disrupt its function[90]. Note that these are useful but overly simplified categorisations, as both the degree of function of the target gene and the ability to be cut can be affected in a non-binary way by sequence changes.

Fig 4.3c gives an example of how different commonly performed assays can be used to identify the D1 allele that was inherited by the F2 progeny. In most cases, the identity of the inherited allele cannot be fully resolved for all progeny. Based on the inferred D1 allele, we can derive a number of metrics that are reflective of the drive's performance (Fig 4.3d). We originally intended to record only a small set of factors in addition to the inferred D1 allele, which would then be used to calculate many independent metrics such as those shown in Fig 4.3d. However, in the process of developing the analysis framework, we found that this frequently led to conflicts

between metrics, where we had to omit recording data for one metric because it was not justified to use the same data to calculate other metrics. To reduce the rigidity of our data requirement, we moved to evaluating and recording most metrics individually at the data ingestion stage. In supplemental methods section 4.7 we have discussed some of the complexities arising from the diversity of data and reporting standards and describe how we handle those with the data structure we developed.

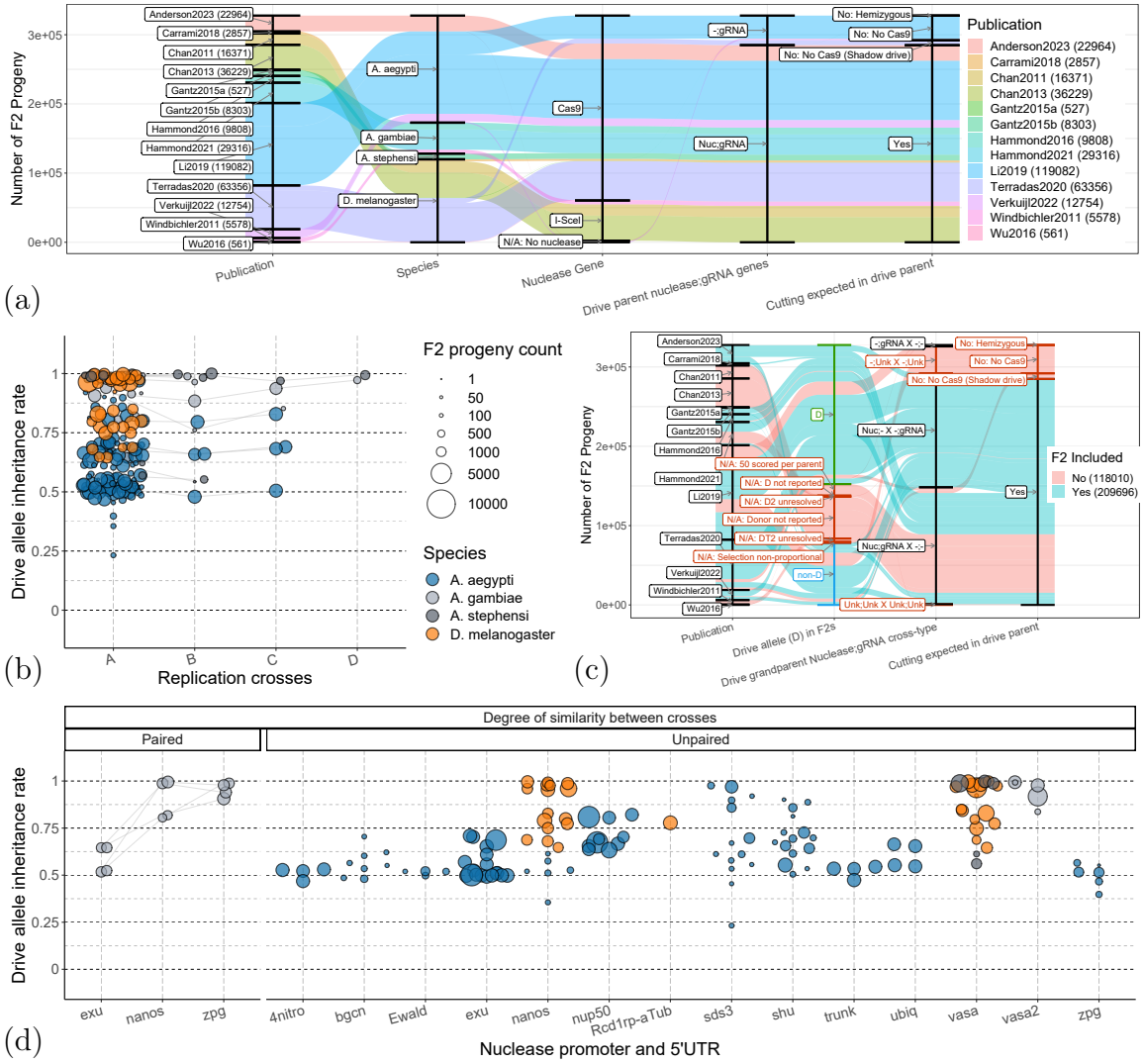


Figure 4.4: **Database scope and gene drive inheritance replicability and rate by nuclease promoter & 5′UTR.** (a) Sankey diagram of all F2 progeny scored in the database. Each column in the diagram shows the values that have been recorded in the dataset for the factors labelled on the X-axis. The size of each value's bracket is proportional to the number of F2 progeny associated with that value. Streams originating at the publication column indicate the specific combination of values reported for each publication. (b) Drive allele inheritance rate ($D/Total$), equivalent crosses have been placed at separate X-axis positions. The lines connect crosses that are considered replications as they do not differ by any of the criteria listed in the supplemental methods section 4.5. In all other figures, the data of replication crosses are merged. (c) Sankey diagram indicating the sub-set of recorded F2 progeny for which gene drive inheritance rates were calculated. In the data selection Sankey diagrams, the green labels and columns indicate values considered positive for the metric being measured. The blue labels and columns indicate values that are negative but are still valid and necessary to calculate an overall rate. Red labels and columns indicate values that cause those F2 counts to be excluded, and black labels and columns indicate neutral values that do not affect the metric calculation. If one or more values cause data to be excluded, the streams between columns representing those progeny are coloured red. (d) Drive inheritance by nuclease promoter & 5′UTR. Lines are drawn between paired crosses that are considered equivalent, apart from the respective X-axis factor. This generally represents the highest-quality comparative data available.

4.3.3 Gene drive inheritance rates vary greatly but are consistent with the same conditions and design

Fig 4.4a gives an overview of the 13 individual publications for which we extracted data and the total number of scored F2s we included from each. These are for Cas9 in *A. agypti*: Anderson et al. 2023[5], Li et al. 2020[95], and Verkuijl et al. 2022[3]. For Cas9 in *A. gambiae*: Hammond et al. 2016[47], Hammond et al. 2021[129] and with I-SceI Windbichler et al. 2011[85]. For Cas9 in *A. stephensi*: Gantz et al.(b) 2015[53]. For Cas9 in *D. melanogaster*: Gantz et al.(a) 2015[88], Wu et al. 2016[201], Carrami et al. 2018[99], Terradas et al. 2021[50] and with I-SceI these are Chan et al. 2011[68] and Chan et al. 2013[78].

We recorded many factors for each cross, and the differences between the conditions and gene drive designs reported in the various publications are generally extensive, making it difficult to judge outcomes in aggregate. However, it is very common that within each study, comparative crosses are performed where a single condition or design element is varied to evaluate its effect. In the coming figures, we have emphasised these "paired" crosses where the only substantial difference between the crosses is the factor being contrasted. This generally represents the highest-quality data available and, in most cases, an intentional experiment to investigate that factor.

From all the factors we recorded, we made a selection sufficient to differentiate each comparative cross in the data set. The factors that were chosen for this are listed together with a justification in the supplemental methods section 4.5. Fig 4.4b shows all the crosses in the database for which drive inheritance can be calculated. Crosses that are described separately in the source publication, but are equivalent by our criteria, are highlighted as paired crosses (linked by grey lines).

The crosses recorded as replications are separate generations of the same transgenic line crossed in the same manner. This figure highlights that there is a large spread in gene drive inheritance rate, but rates are stable when crosses are replicated. It further indicates that the criteria we use are sufficiently detailed and specific to differentiate the large body of crosses captured in the database. In the web tool, we have provided the option to modify this selection to be more or less stringent. Note that the data for individual factors can be very sparse, not reported for almost all crosses except for a few cases where it is the only factor that varies between certain crosses.

Fig 4.4c gives an overview of the data for which the drive inheritance rate is calculated. Red values indicate the F2 progeny for which the drive inheritance rate could not be calculated with a brief justification of why (e.g., all progeny were counted, but only up to 50 progeny from each individual cross were scored for fluorescence[129]). Additionally, it highlights data that were excluded because the drive-carrying parent did not carry a Cas9 transgene or inheritance bias was otherwise impossible. This selection of 209,696 included and 118,010 excluded F2s applies to all figures that report the drive inheritance metric. Notably, no non-Cas9 publication contained valid drive inheritance data according to our criteria because the number of progeny that inherited the donor chromosome was not reported[68;78], or the drive allele could not be

differentiated from resistance mutations[68;78;85]. These data can still be investigated in the web tool using less common metrics (e.g. the ratio of drive to type-2 resistance alleles on the recipient chromosome).

Fig 4.4d shows the rate of inheritance of the drive by the promoter and 5′UTR of the nuclease gene. Our pairing criteria result in only a few cases where the nuclease promoter and 5′UTR are considered the only factor that varies between a set of crosses. Three criteria worth noting here: For crosses to be paired, they must have been reported in the same publication (Fig S4.2a). In addition, based on the variability that has been observed between separate transposon-mediated integrations of the same nuclease transgenes[172] (Fig S4.2b), we only pair crosses where the transgenes regulatory elements are varied at the same genomic site. Fig S4.2b indicates that in some cases, the performance between different insertions of the same nuclease transgene is highly similar. This may be in part due to separate transposon meditated integrations being more variable than separate site-specific HDR or recombinase-mediated cassette exchange (RMCE) integrations. Lastly, for pairing crosses, the target gene, gRNA promoter, and even the specific gRNA sequence must be identical (Fig S4.2c).

4.3.4 There are many metrics of gene drive performance

In addition to the drive inheritance rate, we have recorded over 10 other metrics associated with the gene drive crosses that can be explored individually in the associated web tool. An important set of metrics can be derived from a marker that is recombinatorially linked to the gene drive locus (illustrated by the green star in Fig 4.3).

A recombinatorially linked marker can differentiate the donor-drive allele from newly generated drive alleles on the recipient chromosome. This can improve the power of detecting drive activity by removing the deviations from ∼50% Mendelian inheritance rate of the drive allele independent of any inheritance bias. The marker can also give insight into the underlying mechanism of inheritance bias[1]. Fig S4.3 shows the crosses for which we can, through recombinatorially linked markers, report homing rates and donor-to-recipient chromosome inheritance rates. Frequently, the marker is coincidental and not otherwise analysed in the source publication, such as in publications where in a sub-set of crosses the secondary transgene in a split-drive can function as a chromosome marker.

In the process of developing this meta-analysis, we noted that the sex-determining locus in *A. aegypti* could function as a marker for the *white* targeting drive reported in Li et al.[95]. Surprisingly, the observed inheritance bias was not associated with homing (Fig S4.3c), but instead with an increased inheritance of the donor chromosome (Fig S4.3e). This observation led to a set of experiments described in the next thesis chapter and published in Verkuijl et al.[3].

An important motivation for this project was to lower the barrier to using experimental data in modelling efforts. In addition, to drive inheritance rates, key modelling metrics

are the overall cut-rate (Fig S4.5a), functional disruption rate (Fig S4.5c), and type-1 to type-2 repair rate (Fig S4.5e). Although these metrics are very commonly used in mathematical models, including the one described in the previous chapter[2], experimental data are almost completely absent from our dataset. The lack of data is exacerbated by variable methods and reporting standards that make it difficult to include the few measurements that have been performed (e.g., those in Terradas et al.[50]) in a common database. We expand on this in supplementary method 4.7.

There is another class of metrics that function primarily as controls, but can be relevant for specialised analyses. To aid with interpreting the marker-based metrics, we can estimate in some cases the marker background recombination rate in crosses where inheritance bias is not expected to occur (Fig S4.3a). Another (control) metric is the rate at which the unbiased nuclease transgene is scored in F2s with a split-drive system. Deviations from Mendelian rates may indicate fitness costs associated with the nuclease insertion or nuclease expression[105] (Fig S4.4a). Care must be taken that this metric is only considered for appropriate F1 genotypes (Fig S4.4b). Lastly, the sex of F2 progeny is a commonly recorded factor that is usually not otherwise investigated or analysed (Fig S4.4c). There are cases where the F2 sex deviates from 50% for specific reasons. These include a sex-conversion drive or the previously described cases where the sex-determining locus is linked to a drive element functioning through meiotic drive. We next set out to investigate the effect of individual factors on gene drive outcomes.

4.3.5 Differences in gene drive inheritance rates cannot be attributed to chance alone

As described previously, many factors can vary between different crosses, and we need to take this into account when comparing different values of a single factor. To do this, we analysed the data through meta-analytic multivariate models. The number of drive-inheriting progeny per cross was used to calculate the log odds and corresponding sampling variance for each cross. Table S4.3 lists the attributes of the intercept-only model (Intercept: 86% CI: 73%-93%). Calculation of I^2 index indicates that over 99% of the variability observed in the gene drive inheritance rate between crosses cannot be attributed to chance alone (Table S4.3d). Unaccounted variability is almost equally distributed to between-cluster variability $I^2 = 47\%$ (between studies) and within cluster variability $I^2 = 53\%$ (within studies).

We made a selection of explanatory variables for the drive inheritance rate that were reported for all included crosses and were expected to be the most biologically relevant. Notably, no explanatory factors that reflect the target gene could be included as, in the current data set, many gene drive variables are highly correlated within the distinct designs. We take this lack of independence (particularly in distinct design attributes) into account by limiting our analysis to the effect of sex, which is commonly independently investigated for different designs. Table S4.4 lists the attributes of the meta-analytic multivariate model with additional moderators: species, nuclease

promoter and 5′UTR, drive-carrying parent sex, nuclease carrying grandparent sex, and nuclease carrying grandparent gRNA status.

4.3.6 Inheritance of the drive allele is affected by the drive-carrying parent's sex.

The most common comparative assay performed for synthetic gene drives is to perform the same set of crosses while varying the sex of the parent carrying the gene drive transgenes (Fig 4.5a). In many cases, these are male or female drive-carrying siblings that are separately crossed to a wild-type line. To isolate the effect of parental sex, we calculated the estimated marginal means, which represent the drive inheritance for each parental sex while averaging over the values of all other moderators. This indicated a slightly higher drive inheritance rate when the drive-carrying parent was female (92%, std.error = 0.05) compared to male (88%, std.error = 0.08). A Tukey's post hoc pairwise comparisons indicated that this was a statistically significant difference (Female/Male odds ratio = 1.61, p = 0.01, Table S4.5).

An effect of the drive-carrying sex may also manifest in the drive parent through deposition from the grandparent generation. Deposition can manifest when nuclease components expressed in the grandparent persist in the gametes and affect the parental generation's gene drive propagation rate. Fig 4.5c and Fig 4.5b show the drive inheritance rate separated by the sex of the grandparent(s) carrying the gRNA and nuclease transgenes. In Fig 4.5c we compare the crosses where the sex of the gRNA and nuclease-carrying grandparents are exactly mirrored. Calculation of the estimated marginal mean did not indicate a significant difference in gene drive inheritance due to nuclease carrying grandparent sex (Female/Male odds ratio = 0.75, p = 0.21, Table S4.6). There are a number of hypotheses that are consistent with this result. It may suggest that the absence of deposition or sex-specific deposition occurs but does not have a detectable effect on gene drive propagation rates. Alternatively, deposition occurs, but at equivalent rates from grandparents of either sex. Contrasting crosses in which deposition occurs from grandparents of the same sex but in a different form provides additional context for these results.

The effect of deposition may vary with the form in which it occurs, specifically if the nuclease and gRNA are deposited simultaneously from the same grandparent or from different grandparents. In Fig 4.5b, we compare crosses with the same grandparent carrying both the nuclease transgene and gRNA (Nuc;gRNA X -;-) or crosses where the gRNA gene is contributed by the other grandparent (Nuc;- X -;gRNA). Comparing the estimated marginal mean did not find a significant difference in gene drive inheritance with different forms of deposition from the grandparents ((Nuc;- X -;gRNA) / (Nuc;gRNA X -;-) odds ratio = 0.66, p = 0.19, Table S4.7). It should be noted that comparisons of deposition form is prone to bias from methodological and design factors. It is impossible to perform a deposition form comparison for single-element self-perpetuating gene drive because the gRNA and nuclease gene are genetically linked. In addition, different research groups may adhere to different conventions on how

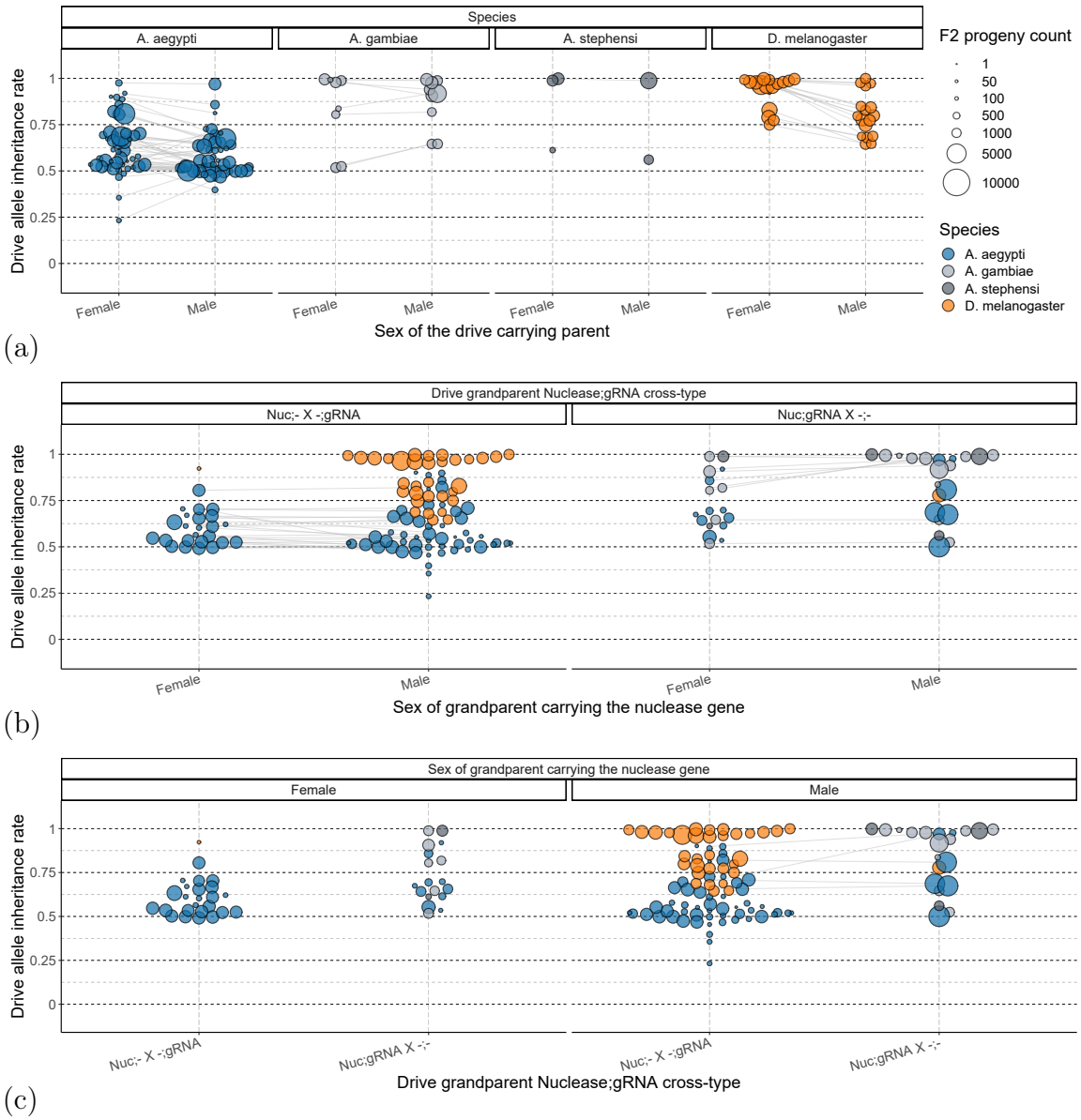


Figure 4.5: **Gene drive inheritance by drive-carrying parent and grandparent sex.** (a) Drive inheritance by sex of the parent in which inheritance bias is expected to occur. Data are separated by species. (b) Drive inheritance by the sex of the grandparent (D0) which carries the nuclease protein gene. Comparisons are only shown between the same deposition forms (nuclease and gRNA present in the same grandparent, or separate grandparents). Nuc;- X -;gRNA indicates a cross in which a nuclease-carrying individual, male or female, was crossed to a gRNA-carrying individual. Nuc;gRNA X -;- indicates a cross where an individual, male or female, carrying both the nuclease and gRNA transgene was crossed to an individual carrying neither. (c) Drive inheritance by grandparent deposition forms. Comparisons are shown between nuclease deposition from the same grandparent sex, with or without the gRNA gene in addition to the nuclease. Lines are drawn between paired crosses that are considered equivalent, apart from the respective X-axis factor. This generally represents the highest-quality comparative data available.

grandparent crosses are performed, and it is not generally specifically investigated or varied within a publication. As such, controlling for confounders is important for such

an analysis, and the paucity of paired crosses in our data suggests that investigating the interaction between gRNA and nuclease deposition is understudied.

A feature of deposition that has been noted in multiple publications is shadow drive[50;96;103;115;118], which occurs when deposited nuclease components mediate inheritance bias in the germline of the deposition recipient (as opposed to only generating resistance mutations). We are able to accommodate these data in our database and specifically isolate crosses where inheritance bias would only be expected to occur through shadow drive (Fig S4.7). Of interest is a paper that we included in our database (Terradas et al.[50]) that has indicated that, in addition to shadow drive being based on maternal deposition, shadow drive is more pronounced in females that receive maternal deposition compared to males who receive maternal deposition (Fig S4.7c). This interpretation of shadow drive depending on the sex of the individual receiving the deposition has been restated in follow-up computational modelling[116]. This result has important implications and, in our opinion, has not received the attention it deserves. Sex differences in inheritance biasing rates that occur with exclusively deposited nuclease would provide important evidence for the potential mechanism of the sex differences with expressed nuclease components we report here. The expansion of our dataset may indicate whether this is a more general phenomenon.

The collective effect of (maternal) deposition may be less pronounced from our analysis than expected from a general reading of the gene drive literature[1]. An important contributing factor to this is likely that individuals that have been affected by deposition can in some cases be identified by an enhanced somatic phenotype (e.g. fully knock-out eye phenotype eyes instead of mosaic-eye phenotype in a gene drive targeting the *white* gene). Performing crosses only with individuals with matched phenotypes may obscure the effect of deposition, as individuals that are most effected by it are not crossed. Put differently, the effect of deposition is obscured because the somatic phenotype rate among crossed individuals sometimes does not reflect the overall phenotype rate among siblings of that genotype. For target genes with a reduced fecundity phenotype, there may be an unavoidable selection for individuals that have experienced a lower degree of deposition as these are the only individuals that mature.

When reported, we have recorded nuclease-induced somatic phenotypes in drive-carrying parents. An example of this is Gantz et al., where individuals with similar somatic phenotypes have relatively similar drive inheritance bias rates irrespective of the possibility of receiving maternal deposition from their grandparents[53]. However, the same study reported large differences in inheritance bias rates between individuals with different somatic phenotypes (Fig 4.5b). Key is that maternal deposition had a large effect on the relative proportion of individuals with the different somatic phenotypes, which in turn reflects a substantial effect on gene drive inheritance rates. The non-proportional crossing of progeny with particular phenotypes will need to be accounted for to fully investigate the effect of deposition on drive inheritance rates. However, in some cases, the phenotype of the drive-carrying parent is not disclosed, making it impossible to correct for. In the next section, we will investigate the effect of sex on somatic phenotype rates.

4.3.7 Progeny's somatic phenotypes are greatly affected by maternal deposition.

In the context of gene drive, somatic phenotypes are visible changes in the phenotype that are generated by the action of the drive nuclease. In most cases, the somatic phenotype is generated by a haplosufficient endogenous gene into which the gene drive is inserted and which the drive also targets the wild-type version. In individuals heterozygous for the knock-out gene drive allele, embryonic or somatic disruption of the recipient chromosome's target gene will result in a somatic phenotype.

In contrast to scoring fluorescent markers, scoring somatic phenotypes is likely much more affected by the exact methodology followed (e.g., developmental stage of screening) and the subjective interpretation of the individual who performs the scoring. Furthermore, somatic phenotypes generated by commonly used target genes (e.g., *white*, *kmo*, *yellow*) can affect outcomes more than differences between the particular fluorescent protein gene and its regulatory elements for scoring transgene inheritance. Lastly, once scored, the analysis of somatic phenotypes is substantially more complex than that of the inheritance of a drive element.

F2s with a somatic phenotype can be generated by a combination of gene drive alleles and type-2 resistance mutations. This disruption of the wild-type target gene can occur in the parental germline, through deposition into the F2 embryo, through the autonomous expression of nuclease components in the F2, or through a combination of the three. Comparing the overall rate of the somatic phenotype between crosses will be heavily confounded by the proportion of F2 progeny that inherited the gene drive allele and other nuclease genes. This can bias the comparison due to a larger fraction of F2 progeny expressing nuclease components and due to the drive allele in most cases acting as a loss-of-function allele. As such, somatic phenotypes can only meaningfully be analysed by a much more detailed stratification of the progeny within each cross to only compare equivalent groups of progeny (e.g., only those that inherited the gene drive allele).

The factors we consider when illustrating the paired data with respect to the somatic phenotype in this manuscript and the web tool are detailed in the supplemental methods section 4.5. These factors cover the F2 nuclease and gRNA carrying status (i.e., can they express the nuclease genes), the number of target alleles that need to be disrupted for a phenotype to occur (between 0 and 2), possible nuclease and gRNA deposition from the F1 parents, if mutation may have been inherited from the germline of the F1 parent, and the F2 sex.

We stratify the F2 progeny by their genotype that is not informed by phenotype data. We ignore the phenotype data to avoid the problem that the somatic phenotype data may have been used to select progeny for a molecular assay that then is used to infer the genotype. This leads to circular reasoning, where all the progeny of a certain genotype have a certain phenotype because only the progeny with that phenotype were genotyped. If a genotype determination can be made independent of the somatic phenotype (e.g., a fluorescently marked drive allele), we may use

that factor to subdivide the progeny. This leads to a complex system of parallel bookkeeping. Where we used all available evidence to determine the genotype of a particular allele, but only use non-phenotype data to determine the genotype absent cutting for the phenotype analysis.

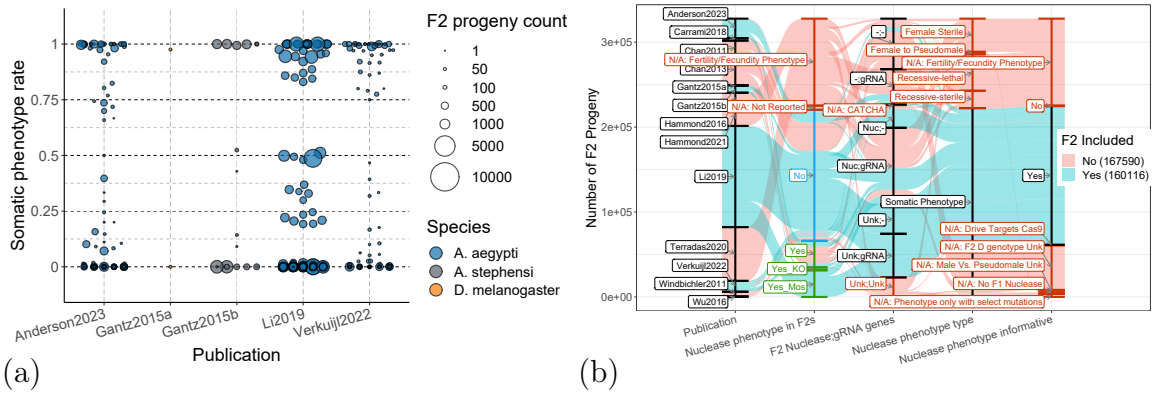


Figure 4.6: **Analysis of nuclease-induced phenotype requires splitting F2 populations from the same cross into distinct sub-groups.** (a) Nuclease-induced phenotype of F2 progeny by publication. F2 progeny from the same cross are separated into additional groups. (b) Sankey diagram indicating the subset of recorded F2 progeny for which the somatic phenotype rate was calculated. In the data selection Sankey diagrams, the green labels and columns indicate values considered positive for the metric being measured. The blue labels and columns indicate values that are negative but are still valid and necessary to calculate an overall rate. Red labels and columns indicate values that cause those F2 counts to be excluded, and black labels and columns indicate neutral values that do not affect the metric calculation. If one or more values cause data to be excluded, the streams between columns representing those progeny are coloured red.

Fig 4.6a gives an overview of the F2 somatic phenotype rate data by publication. Notable is that somatic phenotypes often present in extremes, either fully penetrant or fully absent. Fig 4.6b shows how those data were selected and crosses for which data were excluded because the nuclease target did not generate a scorable somatic phenotype.

Data were analysed using a meta-analytic multivariate model. The number of progeny described as having a somatic phenotype (either a knockout or a mosaic) was used to calculate the log odds and corresponding sampling variance. For each cross, progeny were grouped by the inferred F2 zygotic genotype that can be derived independently from the somatic phenotype (e.g., if they inherited the nuclease transgene or not). Table S4.8 lists the attributes of the intercept-only model (Intercept: 10% CI: 4%-22%). Calculation of the I^2 index indicates that 99% of the variability seen in the rates of somatic phenotypes cannot be attributed to chance alone (Table S4.8d). Unaccounted variability is distributed to variability between studies (I^2 = 3%), between crosses (I^2 = 18%) and between F2 zygotic genotypes (I^2 = 78%).

Explanatory variables were selected to include in the multivariate model. However, again no explanatory factors that reflect the target gene are included as in the current data set many gene drive moderators are highly correlated. Table S4.9 lists the attributes of the meta-analytic multivariate model with moderators: nuclease

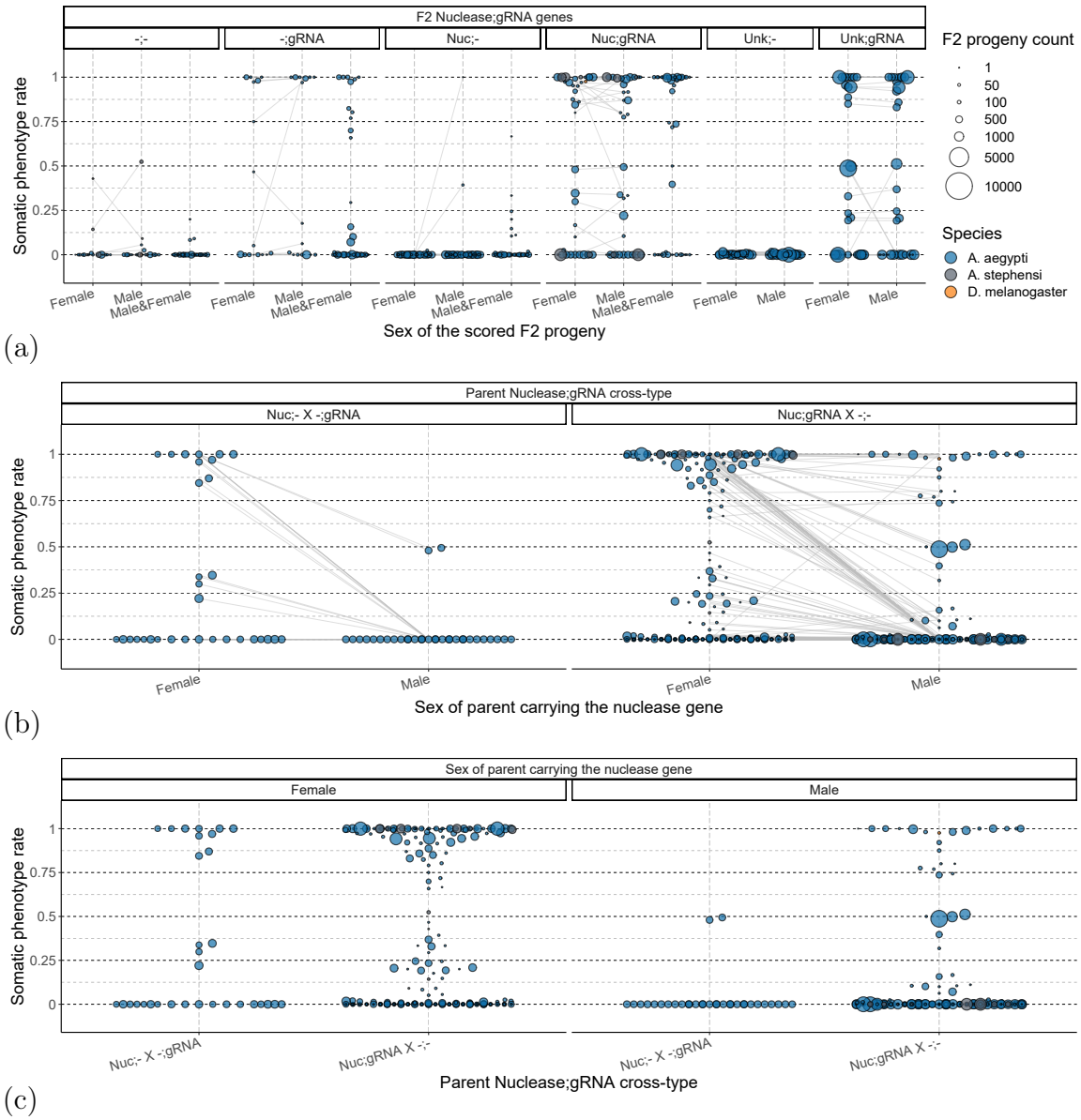


Figure 4.7: **Nuclease-induced somatic phenotype by individual and parental sex.** (a) Somatic phenotype rate by the sex of the F2 progeny. (b) Somatic phenotype rate by the sex of the nuclease-carrying parent. Comparisons are only shown between the same deposition forms (nuclease and gRNA present in the same parent, or separate parents). Nuc;- X -;gRNA indicates a cross in which a nuclease-carrying individual, male or female, was crossed to a gRNA-carrying individual. Nuc;gRNA X -;- indicates a cross where an individual, male or female, carrying both the nuclease and gRNA transgene was crossed to an individual carrying neither. (c) Somatic phenotype rate by parent deposition forms. Comparisons are shown between nuclease deposition from the same parent sex, with or without the gRNA gene in addition to the nuclease. Lines are drawn between paired crosses that are considered equivalent, apart from the respective X-axis factor. This generally represents the highest-quality comparative data available.

promoter, nuclease carrying parent's sex, number of alleles that must be disrupted in F2 for a phenotype to occur, F2 nuclease and gRNA transgene carrying status, the possibility of inheriting resistance alleles from parent's germline, and F2 sex.

In Fig 4.7a we compare the somatic phenotype rate between the sex of the F2 progeny. Despite the differences in the gene drive inheritance rate between male and female parents, there is no significant difference in the rate of somatic phenotype between F2s of different sexes (Female/Male odds ratio = 1.04, p = 0.99, Table S4.10). Fig S4.8a shows the somatic phenotype rate by what nuclease transgenes the F2 of each sex carry.

We performed the same analyses of deposition previously described for drive inheritance. However, now it is the F1 generation that is being evaluated for potential deposition. Fig 4.7b and Fig 4.7c show the F2 somatic phenotype rate separated by the sex of the F1 parent carrying the gRNA and nuclease transgenes. In Fig 4.7b we compare the crosses where the sex of the gRNA and nuclease carrying F1 parents are exactly mirrored.

The estimated marginal means (Table S4.11a) showed a higher F2 somatic phenotype rate when the nuclease carrying F1 parent was female (100%, std error = 0.01) compared to male (96%, std error = 0.17). When compared by Tukey's post hoc pairwise comparisons, this was a significant difference (Female/Male odds ratio = 26.48, p = <0.01, Table S4.11b).

In Fig 4.7c we compare crosses with the same F1 sex carrying the nuclease transgene, or the nuclease transgene and gRNA gene. There was no significant difference (Female/Male odds ratio = 0.24, p = 0.66, Table S4.12). In our current data set, we have no examples of paired crosses where the effect of nuclease deposition with or without the gRNA transgene is compared for the F2 somatic phenotype rate. As was the case for the drive inheritance rate, investigating the interplay between gRNA and nuclease deposition is understudied.

4.4 Discussion

Here we have described the development of a pipeline for a meta-analysis of the performance of homing endonucleases. We used a prior list of known relevant publications to measure the reach of a systematic literature search and proposed refinements to expand coverage. With a subset, but still substantial body of data, we have developed a database structure and analysis pipeline. This comprised 13 publications that include 238 crosses and data from 327,706 individual scored F2s. We have made this available as a web tool and have shown the utility of the work by analysing the effect of sex of gene drive carriers.

First, we evaluated the effect of sex on the nuclease process that occurs in that individual. We find a small but significantly higher inheritance rate from female drive carriers compared to males. However, we did not find significant differences in the somatic phenotype rates between male and female F2s. When looking at deposition, we find that the effects on drive inheritance and somatic phenotypes are reversed. We found no significant differences in the drive inheritance rate between crosses with the opportunity for maternal or paternal nuclease deposition. However, we find that the

somatic phenotype rate is substantially increased when there is an opportunity for maternal deposition. We find no effect of the deposition form (Nuclease complex, or nuclease and gRNA separately) on the rates of drive inheritance or somatic phenotype.

The higher gene drive inheritance rate from females we find may in part be due to maternal deposition. Data from Terradas et al. comprises gene drive elements that rescue the function of a deleterious target gene[50]. Deposition can be expected to decrease the viability of F2 progeny that did not inherit the rescue drive element. If maternal deposition was greater than paternal deposition, this could result in a relative increase in drive-carrying progeny from female drive carriers. In contrast, sex differences should only manifest if there is at least some level of inheritance biasing activity. As such, the inclusion of poorly performing drives may have muted the underlying difference in inheritance biasing ability between the sexes. Furthermore, in some cases, a nuclease promoter is specifically chosen to function only in one of the sexes, and the data are not reported for both sexes, or they are reported but in a format that cannot be included in the database (e.g., Windbichler et al.[85], although these data were excluded for other reasons stated earlier).

We have compiled a large data set with substantial statistical power to draw conclusions about the functioning of the gene drive. We deliberately chose papers with diverse methods and transgene designs. This has allowed us to design our pipeline to accommodate a diversity of data that includes different nuclease genes, multiplexing, linked markers, different numbers of loci, and different metrics. However, a substantial drawback is that the data set we have compiled is not a systematic and unbiased representation of a specific subset of the gene drive literature. For a balanced analysis, all papers that passed our inclusion criteria will need to be added, although this represents a large amount of work. Alternatively, an additional non-biased method can be used to narrow the scope of data collection, such as focussing exclusively on *Aedes agypti* which is already relatively well represented in our data set. We hope that follow-up work will validate the broad approach we have taken and that contributions by other researchers will help this project reflect the full scope of homing gene drive research.

Recently, there has been a shift in the power and accessibility of generative artificial intelligence tools[202]. Large language models may soon become common tools in scientific research, specifically in summarising prior research. It will likely become a much easier task to extract and reformat data with these tools which will prove timely in aiding with the expansion of our dataset. This will still be a useful task, as the pipeline we have developed will be able to perform analyses and draw conclusions from the data that a simple summary of those publications using language models would not be capable of.

We have analysed both drive inheritance rates and somatic phenotype rates through multivariate meta-analytic models. A limitation of our statistical analysis is that we currently only consider the effect of different factors in isolation. There are clear biological reasons to anticipate that certain variables will interact in their effect on gene drive outcomes. An example is the expectation that nuclease deposition into

F2 progeny and gRNA expression by F2 progeny will be interdependent for their effects on gene drive outcomes (both gRNA and nuclease are required for a cut to occur). This is particularly the case for somatic phenotype rates, where the relatively binary outcomes we report are not simply due to the sum of individual factors but likely a complex interplay of different factors. For drive inheritance rates, the data is considerably more homogeneous, with in all cases a relatively equivalent heterozygote drive parent being evaluated. We have chosen to keep our models relatively simple to increase their interpretability, especially with this analysis on a reduced dataset.

Frequently, studies by different research groups will differ in both experimental methods as well as molecular designs making it impossible to disentangle their individual contribution. This is particularly important in regard to split- and single-element homing drives. Split drives with intermediate inheritance biasing efficiency have been reported in *A. aegypti* and *D. melanogaster* while highly efficient single-element drive designs have been reported in the *Anopheles* mosquitoes. The negative effects of deposition on drive inheritance may only be apparent in otherwise efficient drives, which are under-represented in our current data set. Furthermore, split drives are frequently studied with crosses in which the different grandparents each contribute one of the Cas9 and gRNA transgenes. This may further reduce our ability to detect the effect of deposition on drive inheritance rates. An expanded data set may reduce the co-linearity of the dataset and enable additional analyses of individual gene drive design factors and subgroups.

Our results suggest that, on the whole, deposition has a large effect on what tissues are cut (i.e., somatic tissues), but less so on the outcome of cutting (i.e., inheritance bias vs. unintended repair outcomes). However, as noted in the results, the effect of deposition on inheritance bias may be obscured by the (undisclosed) selection of individuals for crosses that are equally affected by deposition. We think it is important to highlight this possibility because this selection for phenotypically similar individuals may well be done intentionally in an effort to maximise the correspondence of comparative crosses. However, if not disclosed and not consciously considered, this can reduce the reproducibility and representativeness of the results that are produced. This interpretation of our results indicates that accounting for the variability of deposition between equivalent siblings may be very important for anticipating its effects.

An important application of the work presented here is in aiding computational models of gene drives. The flexible data framework we have developed can allow computational modelling efforts to anticipate the general format of experimental data and develop analytic pipelines in parallel to experimental work. In addition, the depth of detail we have captured allows computational models to account for the experimental and parent-by-parent variability, which is an underexplored factor in gene drive modelling. We further anticipate that this project will be useful for modellers to supplement limited experimental data on the specific gene drive design they are studying with finding from the most closely related gene drives.

One of the key limitations of our approach is that we only considered a relatively narrow set of cross-requirements and data in our analysis. Many of the studies we

reviewed included additional data on gene drive performance, such as cage trial data, which we did not incorporate. Care should be taken with regard to this, as the fact that we cannot calculate a particular parameter relevant for modelling does not mean that parameter was not estimated with additional experiments that were not compatible with our database. In particular, molecular assays such as sequencing were often excluded because we could not attribute those results to a specific cross or fully place them in the context of siblings that were not reported. Although our work cannot capture the complete breadth of work presented in each individual research project, we nonetheless expect that presenting this work in a machine-readable format will aid in computational modelling efforts.

A trend that has emerged in the field is the testing of many candidate transgene designs to find high-performing designs. We expect our work to be particularly useful for developing approaches to prioritise what and how single-cross experiments are performed on candidate gene drives. The combination of parameters that can be estimated from a single experiment is generally not considered when using modelling data to prioritise experiments. Neither is that often outcomes that are experimentally measured will be the result of multiple individual parameters that would need additional experiments and assumptions to deconvolute. The importance of measuring variability and non-linear sensitivities is also not clear. For example, the estimation of certain parameters with high precision may only be relevant if they are above or below a certain threshold. In addition, for many gene drive experiments, parameters can be estimated as the average over many individuals, or laboriously measured per individual. Although the individual measure generally provides more information about the same parameters, it requires much more work, limiting the scope of the number of individuals that can be measured. Our database provides a framework for planning out experiments and developing modelling efforts.

Another important utility of our tools is to assist molecular biologists in the design of new gene drive systems. We have collected many design factors and made it possible to individually evaluate them in the web tool, and identify experiments where these specific design criteria have been varied (paired crosses). We have also provided the option to expand or narrow the pairing criteria. If the specific experiment comparing a design factor is not available, loosening the criteria can point researchers to the next best available evidence and the caveats that come with that. Where possible, we have also recorded the DNA sequence of the drive element(s) and the unmodified target site. We expect that this will be a valuable resource for further analysis to identify additional design criteria. With only these sequences, simple bioinformatic analysis can identify gene drive element size, homology arm alignment, need for homology arm resection, internal homology, and gRNA efficiency parameters.

We have only scratched the surface of the possible analyses that can be performed on these data to improve our understanding of homing gene drives. We hope that by making this set of tools available, the colleagues whose data comprise it can also help guide its development. Because so many different analyses can be run on a single cross, a lot of utility comes from developing a flexible data format that directly feeds

into an analytical pipeline to run many analyses simultaneously. This can quickly highlight anomalous results that may warrant further investigation. An example of this is that the analysis of recombinatorially linked markers unexpectedly revealed the lack of homing in the data of Li et al.[95]. In the next chapter, we analyse experimental work that was performed to investigate this phenomenon further.

4.5 Methods

Inclusion and exclusion criteria

The publication search was performed as described in Fig 4.1. For publications selected for data extraction, we read the publication thoroughly and, in most cases, relied on supplemental data files with unprocessed cross-data. Where relevant, other parts of the article were consulted for factors not listed in the cross-data. With the data for each cross, we have included a brief quote, reference, or deduction that was used to inform the values we recorded for that factor. Data were not included if only summary statistics were reported instead of individual progeny count. In some cases, this means excluding data because we cannot exactly determine the denominator of a reported frequency. An example of this is the crosses reported in Fig.S1.D. of Carrami et al.[99]. The crosses reported in Fig.S3.B of this same paper are included as individual progeny counts are reported.

Database verification

Currently, all data were compiled and verified by one person, and we propose that at least one additional person verify the values recorded for each publication. Part of the web tool has been dedicated to providing an interface for this. Part of the data validation process may include sending the compiled data set to the authors of the original publication and giving them the opportunity to check the data.

Database structure

Each row of the database lists a particular grouping of individuals from a published cross. In most cases, these are the "F2" progeny from a cross where one of the "F1" parents was a carrier of the gene drive being investigated. The progeny are grouped together when they cannot be further distinguished by any of the factors that we and the original publication record. As such, the "F2" progeny from a single parent will be represented over multiple rows of the database. Almost all factors will be identical for these progeny, except one or a few. Examples of these differences will be the presence of a marker indicative of drive inheritance, or the particular NHEJ mutations they were recorded as having. Factors regarding drive design will be common among all F2 individuals in a cross and often within a publication. We have, in rare cases, partitioned a publication's data set described as a single cross if additional factors were recorded that can be used to sub-set it (e.g., data of different D1 somatic phenotypes

are combined, while we consider these separate crosses). In section 4.7, we describe the recorded factors of the database with an additional description and examples.

Factors deemed essential in the comparative analysis

We made a selection of factors to use as criteria for the pairing of crosses. These factors are only used to group (draw lines between) outcomes for data visualisation. A large number of metadata fields are ignored because they effectively describe the same thing as another factor (e.g. the publication name and publication doi). Other factors we combined because, in almost all cases, a change to one factor is paired with a change to the other (e.g. F0 maternal and paternal gRNA/Nuclease status). Another irreducible set of factors is removed because of judgment calls on their biological relevance. For example, we match crosses based on the F0 gRNA/Nuclease status, but not the exact genotype. Of course, the same subjective decisions were made when deciding what factors to record in the database and by the original authors on what data to report. We have not made paired comparisons between publications to reduce the risk of undisclosed differences for which we cannot control.

The pairing criteria are as follows: Publication, Species, Nuclease Insertion Variant, D1 parent somatic phenotype, D1 parent sex, T1 parent cutting expected, stage screened, Nuclease gene, Nuclease Promoter and 5′UTR, D1 housing temperature, D0 nuclease carrying sex, D0 cross form, target gene, target gene fitness type, target gene sufficiency, gRNA sequence, gRNA target site, gRNA promoter.

For the somatic phenotype metric, we further subdivide the data by: The F2 zygotic genotype, the number of alleles required to be knocked out in the F2 for a phenotype to occur, the F2 nuclease and gRNA genotype, the F1 nuclease carrying sex, if a nuclease component might be biased in either parent, if a resistance mutation may have been inherited from the parent's germline, and the F2 sex. Note that the sex of the drive-carrying parent is not considered for the pairing of somatic phenotype data (it is instead replaced by the sex of the nuclease-carrying parent).

Many of these factors currently do not have an effect because none of the crosses in the data set varies for this factor. For example, there are no cases where we anticipate nuclease activity to have occurred in the non-drive-carrying parent. However, we expect this to be biologically important if it did, and have such included it as a criterion in anticipation of future data where this might be the case.

Meta-analytic multilevel model for drive inheritance

Multivariate model analysis was performed in R (v4.2.2)[203] using the metafor package (v4.0-0)[204]. The I^2 index[205] was calculated using the orchaRd package (v2.0)[206]. I^2 represents the amount of variation not attributable to the sampling error. Estimated marginal means (least-squares means) were calculated using the emmeans package (v1.8.5)[207]. The marginal means of the variable of interest reflect the contribution of a particular factor while averaging over all other levels of the other factors. Differences

between factor levels were assessed with the Tukey's post hoc pairwise comparison test using the pairs() function in the emmeans package.

For drive inheritance models, the number of drive-inheriting progeny per cross was used to calculate the log odds and the corresponding sampling variance for each cross (escalc() in the metafor package). The source publication was included as a random effect. Data were included as shown in Fig 4.4c.

For the nuclease-induced somatic effect models, the number of progeny with either a knock-out or mosaic phenotype was used to calculate the log odds and corresponding sampling variance (escalc() in the metafor package). Progeny were grouped by the cross and the F2 zygotic genotype absent cutting (Fig 4.2f). Grouping by F2 zygotic genotype absent cutting results in the nuclease phenotype rate being calculated separately for F2 progeny with or without the nuclease and gRNA transgenes. The source publication and cross were included as a random effect. Data were included as shown in Fig 4.6b.

Code and Data Availability

The web tool and analysis scripts were written in R (v4.2.2) and are available at: `https://osf.io/k9fbd/`. A file listing all publications returned by the literature and their manual classification is also provided on the external host.

Acknowledgements

Biotechnology and Biological Sciences Research Council (BBSRC) supported S.A.N.V. (BB/M011224/1) and M.B.B. (BB/H01814X/1, BB/L00948X/1, BB/V008110/1). The views, opinions, and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the funders. The funders had no role in study design, data collection and analysis, the decision to publish, or the preparation of the manuscript.

Author contributions

S.A.N.V. conceptualised the study. S.A.N.V. performed all investigations, formal analysis, software development, visualisation, and writing. M.B.B. provided direct supervision. P.T.L and E.I.C provided guidance on the statistical analysis approach and provided comments on the manuscript draft.

4.6 Supplemental Data

PMID	First author	Citation	Returned by search	Prior database
16299390	Preston CR	Genetics. 2006 Feb;172(2):1055-68. doi: 10.1534/genetics.105.050138. Epub 2005 Nov 19.	No	Yes
21203963	Chen Y	Protein Cell. 2010 May;1(5):478-90. doi: 10.1007/s13238-010-0058-2. Epub 2010 Jun 4.	No	Yes
21368273	Chan YS	Genetics. 2011 May;188(1):33-44. doi: 10.1534/genetics.111.127506. Epub 2011 Mar 2.	Yes	Yes
21508956	Windbichler N	Nature. 2011 May 12;473(7346):212-5. doi: 10.1038/nature09937. Epub 2011 Apr 20.	Yes	Yes
23349805	Chan YS	PLoS One. 2013;8(1):e54130. doi: 10.1371/journal.pone.0054130. Epub 2013 Jan 18.	Yes	Yes
24040217	Chan YS	PLoS One. 2013 Sep 10;8(9):e74254. doi: 10.1371/journal.pone.0074254. eCollection 2013.	Yes	Yes
24803674	Simoni A	Nucleic Acids Res. 2014 Jun;42(11):7461-72. doi: 10.1093/nar/gku387. Epub 2014 May 6.	No	Yes
25908821	Gantz VM	Science. 2015 Apr 24;348(6233):442-4. doi: 10.1126/science.aaa5945. Epub 2015 Mar 19.	Yes	Yes
26598698	Gantz VM	Proc Natl Acad Sci U S A. 2015 Dec 8;112(49):E6736-43. doi: 10.1073/pnas.1521077112. Epub 2015 Nov 23.	Yes	Yes
26641531	Hammond A	Nat Biotechnol. 2016 Jan;34(1):78-83. doi: 10.1038/nbt.3439. Epub 2015 Dec 7.	No	Yes
26849513	Wu B	Nat Biotechnol. 2016 Feb;34(2):137-8. doi: 10.1038/nbt.3444.	No	Yes
27334272	Lin CC	Genetics. 2016 Aug;203(4):1613-28. doi: 10.1534/genetics.116.191783. Epub 2016 Jun 22.	No	Yes
27638686	Lin CC	G3 (Bethesda). 2016 Nov 8;6(11):3685-3691. doi: 10.1534/g3.116.034884.	Yes	No
28727785	Champer J	PLoS Genet. 2017 Jul 20;13(7):e1006796. doi: 10.1371/journal.pgen.1006796. eCollection 2017 Jul.	Yes	Yes
28976972	Hammond AM	PLoS Genet. 2017 Oct 4;13(10):e1007039. doi: 10.1371/journal.pgen.1007039. eCollection 2017 Oct.	Yes	Yes
29274230	Xu XS	Elife. 2017 Dec 23;6:e30281. doi: 10.7554/eLife.30281.	No	Yes
29735716	Champer J	Proc Natl Acad Sci U S A. 2018 May 22;115(21):5522-5527. doi: 10.1073/pnas.1720354115. Epub 2018 May 7.	Yes	Yes
29844184	Carrami EM	Proc Natl Acad Sci U S A. 2018 Jun 12;115(24):6189-6194. doi: 10.1073/pnas.1713825115. Epub 2018 May 29.	Yes	Yes
30224454	Oberhofer G	Proc Natl Acad Sci U S A. 2018 Oct 2;115(40):E9343-E9352. doi: 10.1073/pnas.1805278115. Epub 2018 Sep 17.	Yes	Yes
30247490	Kyrou K	Nat Biotechnol. 2018 Dec;36(11):1062-1066. doi: 10.1038/nbt.4245. Epub 2018 Sep 24.	No	Yes
30666960	Champer J	Elife. 2019 Jan 22;8:e41439. doi: 10.7554/eLife.41439.	Yes	Yes
30675057	Grunwald HA	Nature. 2019 Feb;566(7742):105-109. doi: 10.1038/s41586-019-0875-2. Epub 2019 Jan 23.	Yes	Yes
30918006	Champer J	Genetics. 2019 May;212(1):333-341. doi: 10.1534/genetics.119.302037. Epub 2019 Mar 27.	Yes	Yes
30967548	Guichard A	Nat Commun. 2019 Apr 9;10(1):1640. doi: 10.1038/s41467-019-09694-w.	No	Yes
31856182	Pham TB	PLoS Genet. 2019 Dec 19;15(12):e1008440. doi: 10.1371/journal.pgen.1008440. eCollection 2019 Dec.	No	Yes
31882406	Kandul NP	G3 (Bethesda). 2020 Feb 6;10(2):827-837. doi: 10.1534/g3.119.400985.	Yes	Yes
31953404	Lopez Del Amo V	Nat Commun. 2020 Jan 17;11(1):352. doi: 10.1038/s41467-019-13977-7.	No	Yes
31960794	Li M	Elife. 2020 Jan 21;9:e51701. doi: 10.7554/eLife.51701.	No	Yes
32109227	Champer J	Nat Commun. 2020 Feb 27;11(1):1082. doi: 10.1038/s41467-020-14960-3.	Yes	No
32181354	Champer SE	Sci Adv. 2020 Mar 4;6(10):eaaz0525. doi: 10.1126/sciadv.aaz0525. eCollection 2020 Mar.	Yes	Yes
32393821	Simoni A	Nat Biotechnol. 2020 Sep;38(9):1054-1060. doi: 10.1038/s41587-020-0508-1. Epub 2020 May 11.	No	Yes
32610142	Lopez Del Amo V	Cell Rep. 2020 Jun 30;31(13):107841. doi: 10.1016/j.celrep.2020.107841.	No	Yes
32786353	Chae D	ACS Synth Biol. 2020 Sep 18;9(9):2362-2377. doi: 10.1021/acssynbio.0c00117. Epub 2020 Aug 24.	Yes	Yes
32839345	Carballar-Lejarazu R	Proc Natl Acad Sci U S A. 2020 Sep 15;117(37):22805-22814. doi: 10.1073/pnas.2010214117. Epub 2020 Aug 24.	No	Yes
32929034	Champer J	Proc Natl Acad Sci U S A. 2020 Sep 29;117(39):24377-24383. doi: 10.1073/pnas.2004373117. Epub 2020 Sep 14.	Yes	Yes
32949493	Xu XS	Mol Cell. 2020 Oct 15;80(2):246-262.e4. doi: 10.1016/j.molcel.2020.09.003. Epub 2020 Sep 18.	No	Yes
33095043	Pfitzner C	CRISPR J. 2020 Oct;3(5):388-397. doi: 10.1089/crispr.2020.0050.	Yes	Yes
33144570	Adolfi A	Nat Commun. 2020 Nov 3;11(1):5553. doi: 10.1038/s41467-020-19426-0.	No	Yes
33513149	Hammond A	PLoS Genet. 2021 Jan 29;17(1):e1009321. doi: 10.1371/journal.pgen.1009321. eCollection 2021 Jan.	Yes	Yes
33666174	Kandul NP	Elife. 2021 Mar 5;10:e65939. doi: 10.7554/eLife.65939.	Yes	Yes
33674604	Terradas G	Nat Commun. 2021 Mar 5;12(1):1480. doi: 10.1038/s41467-021-21771-7.	Yes	Yes
33845943	Hoermann A	Elife. 2021 Apr 13;10:e58791. doi: 10.7554/eLife.58791.	No	Yes

PMID	First author	Citation	Returned by search	Prior database
33976171	Li Z	Nat Commun. 2021 May 11;12(1):2625. doi: 10.1038/s41467-021-22927-1.	Yes	Yes
34017003	Feng X	Nat Commun. 2021 May 20;12(1):2960. doi: 10.1038/s41467-021-23239-0.	No	Yes
34050017	Garrood WT	Proc Natl Acad Sci U S A. 2021 Jun 1;118(22):e2004838117. doi: 10.1073/pnas.2004838117. Epub 2021 Apr 30.	Yes	Yes
34158505	Zhang T	Nat Commun. 2021 Jun 22;12(1):3854. doi: 10.1038/s41467-021-24195-5.	Yes	Yes
34172748	Taxiarchi C	Nat Commun. 2021 Jun 25;12(1):3977. doi: 10.1038/s41467-021-24214-5.	No	Yes
34321476	Hammond A	Nat Commun. 2021 Jul 28;12(1):4589. doi: 10.1038/s41467-021-24790-6.	No	Yes
34610011	Fuchs S	PLoS Genet. 2021 Oct 5;17(10):e1009740. doi: 10.1371/journal.pgen.1009740. eCollection 2021 Oct.	Yes	Yes
34791161	Terradas G	G3 (Bethesda). 2022 Jan 4;12(1):jkab369. doi: 10.1093/g3journal/jkab369.	Yes	Yes
34893590	Gamez S	Nat Commun. 2021 Dec 10;12(1):7202. doi: 10.1038/s41467-021-27333-1.	No	Yes
34941868	Weitzel AJ	PLoS Biol. 2021 Dec 23;19(12):e3001478. doi: 10.1371/journal.pbio.3001478. eCollection 2021 Dec.	Yes	Yes
35022402	Kaduskar B	Nat Commun. 2022 Jan 12;13(1):291. doi: 10.1038/s41467-021-27654-1.	No	Yes
35285719	Xu X	CRISPR J. 2022 Apr;5(2):224-236. doi: 10.1089/crispr.2021.0129. Epub 2022 Mar 14.	Yes	Yes
35389492	Carballar-Lejarazu R	Genetics. 2022 May 31;221(2):iyac055. doi: 10.1093/genetics/iyac055.	No	Yes
35394026	Yang E	G3 (Bethesda). 2022 May 30;12(6):jkac081. doi: 10.1093/g3journal/jkac081.	Yes	Yes
35534475	Bishop AL	Nat Commun. 2022 May 9;13(1):2595. doi: 10.1038/s41467-022-29868-3.	Yes	Yes
35606745	Metzloff M	BMC Biol. 2022 May 24;20(1):119. doi: 10.1186/s12915-022-01292-5.	Yes	Yes
35613590	Lopez Del Amo V	Cell Rep. 2022 May 24;39(8):110843. doi: 10.1016/j.celrep.2022.110843.	Yes	Yes
35646834	Nash A	Front Bioeng Biotechnol. 2022 May 12;10:857460. doi: 10.3389/fbioe.2022.857460. eCollection 2022.	No	Yes
35653396	Ellis DA	PLoS Genet. 2022 Jun 2;18(6):e1010244. doi: 10.1371/journal.pgen.1010244. eCollection 2022 Jun.	Yes	Yes
35671288	Castle AR	PLoS One. 2022 Jun 7;17(6):e0269342. doi: 10.1371/journal.pone.0269342. eCollection 2022.	Yes	Yes
35845988	Asad M	Front Physiol. 2022 Jun 29;13:938621. doi: 10.3389/fphys.2022.938621. eCollection 2022.	Yes	Yes
36129981	Hoermann A	Sci Adv. 2022 Sep 23;8(38):eabo1733. doi: 10.1126/sciadv.abo1733. Epub 2022 Sep 21.	No	Yes
36135925	Langmuller AM	Elife. 2022 Sep 22;11:e71809. doi: 10.7554/eLife.71809.	Yes	Yes
36250791	Reid W	G3 (Bethesda). 2022 Dec 1;12(12):jkac280. doi: 10.1093/g3journal/jkac280.	No	Yes
36414618	Verkuijl SAN	Nat Commun. 2022 Nov 21;13(1):7145. doi: 10.1038/s41467-022-34739-y.	Yes	Yes

Table S4.1: Publications included in the prior database and returned by the systematic literature search. Two new publications that were not already in the prior database were identified[61;194].

Search Term	PubMed Results
"Gene drive"	584
"gene conversion" (nuclease | endonuclease | I-SceI | Cas9 | CRISPR)	502
Allele-Specific (nuclease | endonuclease | I-SceI | Cas9 | CRISPR)	482
"loss of heterozygosity" (nuclease | endonuclease | I-SceI | Cas9 | CRISPR)	217
allelic exchange (nuclease | endonuclease | I-SceI | Cas9 | CRISPR)	150
allelic HDR (nuclease | endonuclease | I-SceI | Cas9 | CRISPR)	110
homing drive (nuclease | endonuclease | I-SceI | Cas9 | CRISPR)	94
noncrossovers recombination (nuclease | endonuclease | Cas9 | I-SceI | CRISPR)	65
recombination between homologous chromosome arms (nuclease | endonuclease | Cas9 | I-SceI | CRISPR)	64
intragenic recombination (nuclease | endonuclease | I-SceI | Cas9 | CRISPR)	59
interchromosomal recombination (nuclease | endonuclease | I-SceI | Cas9 | CRISPR)	52
interhomolog recombination (nuclease | endonuclease | I-SceI | Cas9 | CRISPR)	34
non-crossover recombination (nuclease | endonuclease | I-SceI | Cas9 | CRISPR)	21
Super-Mendelian Inheritance	19
Inter-Homolog Recombination (nuclease | endonuclease | Cas9 | I-SceI | CRISPR)	8
Copy-neutral (nuclease | endonuclease | I-SceI | Cas9 | CRISPR)	3
interchromosomal HDR (nuclease | endonuclease | I-SceI | Cas9 | CRISPR)	1

Table S4.2: **Expanded search terms with a focus on the underlying DNA repair processes.** The "|" symbol is used as an OR operator. Search performed in January 2023.

Formula	Random effects
~ 1	~ 1 | Publication/obsID

(a)

term	type	estimate	std.error	statistic	p.value
overall	summary	1.78	0.41	4.35	0.00

(b)

pred	ci.lb	ci.ub	pi.lb	pi.ub
0.86	0.73	0.93	0.18	0.99

(c)

I2_Total	I2_Publication	I2_Publication/obsID
99.82	46.80	53.01

(d)

Table S4.3: **Model attributes of the intercept-only gene drive inheritance multivariate model.** (a) Model moderator specification and random effects. Analysis was performed using the rma.mv() function from the metafor package. (b) Model results on the log-odds scale. This component displays the coefficients and their corresponding standard errors, which provide information on the magnitude and direction of the effects of the model predictors on the outcome variable. (c) Results of the model back-transformed from log-odds to drive-inheritance rate. The backtransformation provides a more interpretable representation of the model results. (d) Model I^2 index[205]. Shown is the total amount of variance not attributable to sampling error and this variance subdivided over the model's random effects.

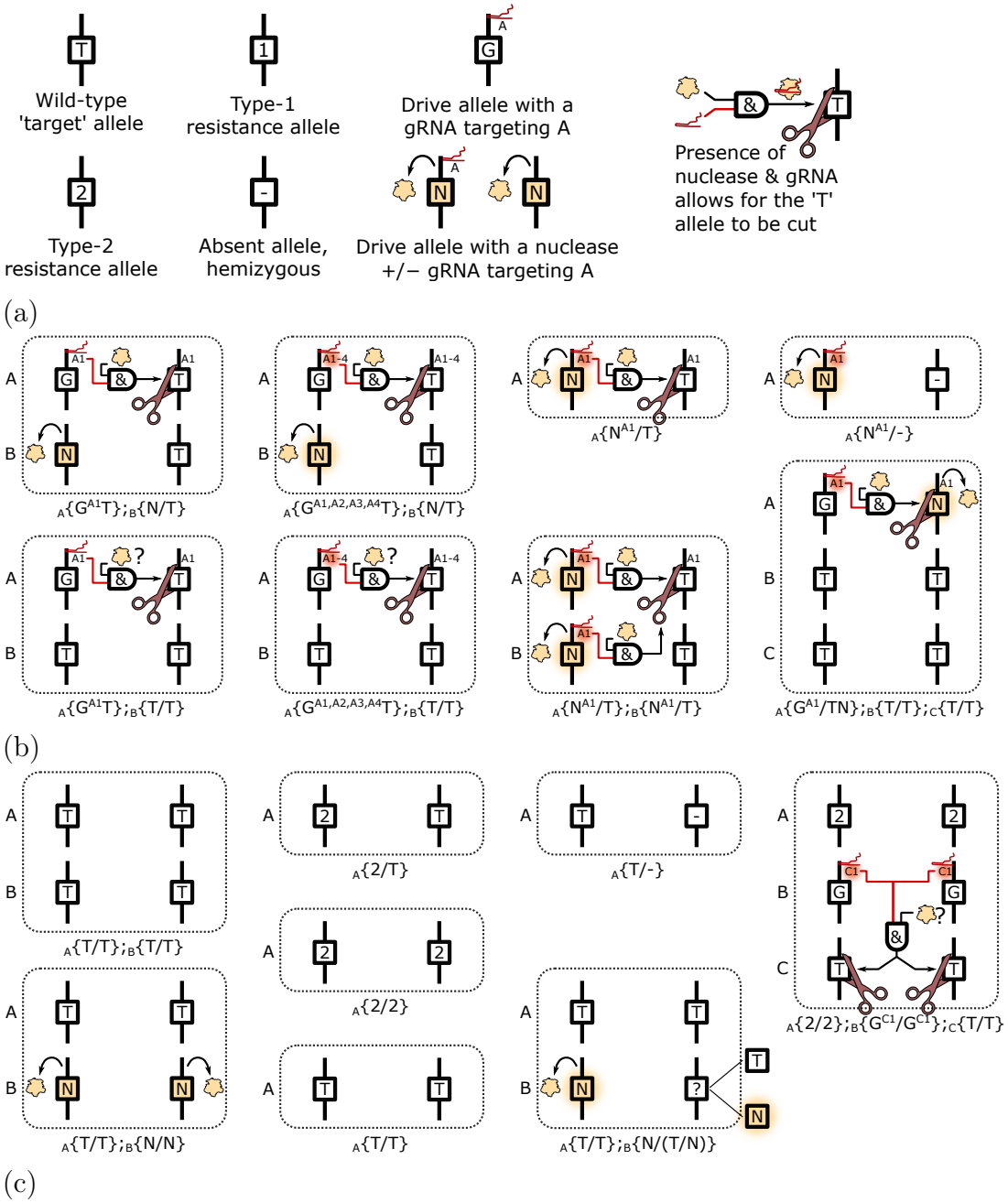


Figure S4.1: **Allele and genotype notation system.** We have expanded the genotype notation system described in Verkuijl et al. [2]. The allele that is being biased is described as "D", and in all cases shown here, this is a "G" or "N" allele. (a) Genotype descriptions are composed of multiple alleles. These alleles describe the nuclease components (nuclease/gRNA) and the targeting components. (b) Illustrations of the different genotypes of the drive parent (D1) genotypes found in the data set. (c) Illustrations of the genotypes to which the drive parent is crossed (T1). F2 genotypes cover a much larger range of genotypes described using the same system.

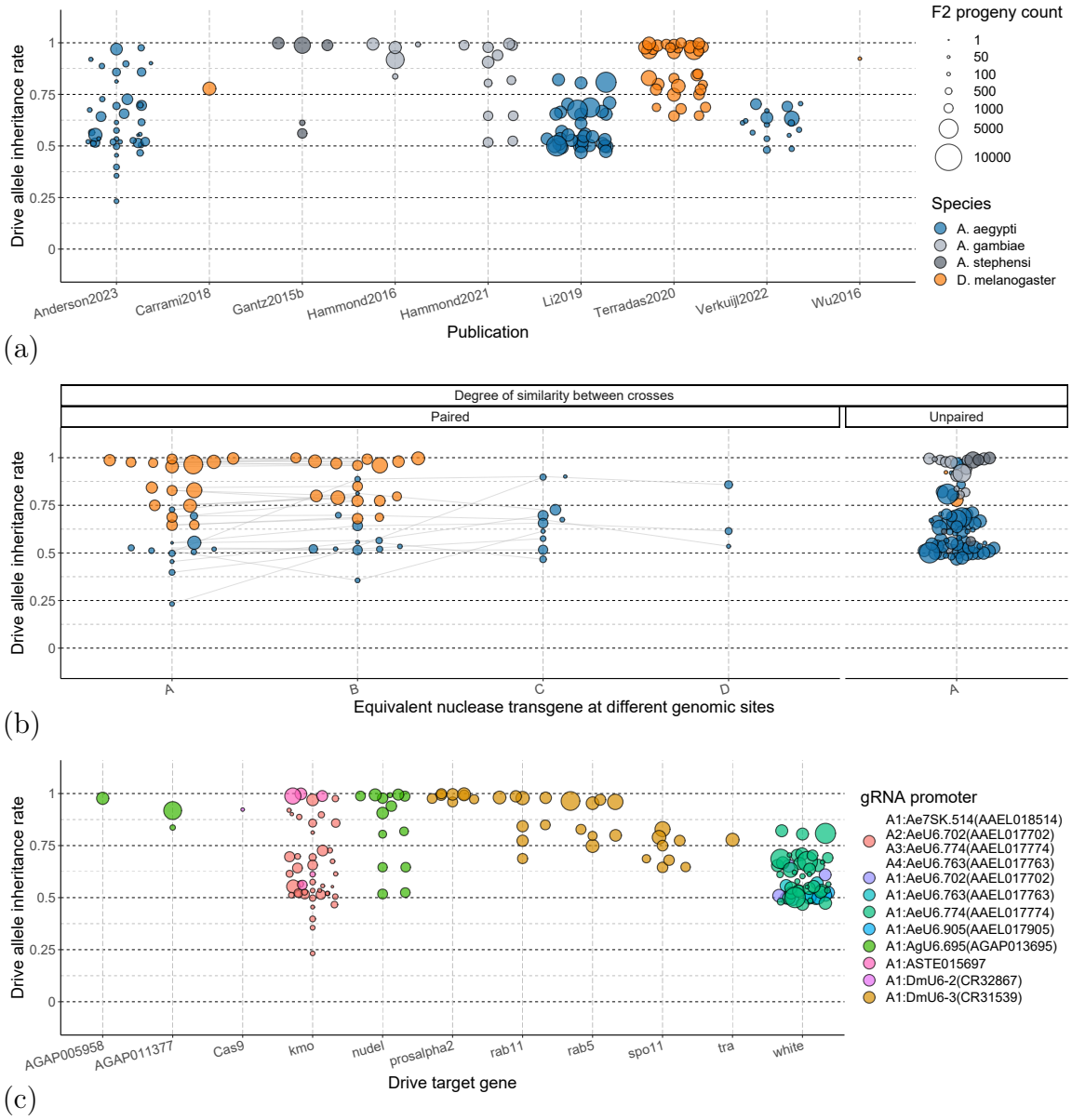


Figure S4.2: **Drive inheritance by publication, nuclease insertion site, and target gene.** (a) Drive inheritance rate by nuclease insertion variant. (b) Drive inheritance by publication. (c) Drive inheritance by the gRNA promoter. Lines are drawn between paired crosses that are considered equivalent, apart from the respective X-axis factor. This generally represents the highest-quality comparative data available. In the data selection Sankey diagrams, the green labels and columns indicate values considered positive for the metric being measured. The blue labels and columns indicate values that are negative but are still valid and necessary to calculate an overall rate. Red labels and columns indicate values that cause those F2 counts to be excluded, and black labels and columns indicate neutral values that do not affect the metric calculation. If one or more values cause data to be excluded, the streams between columns representing those progeny are coloured red.

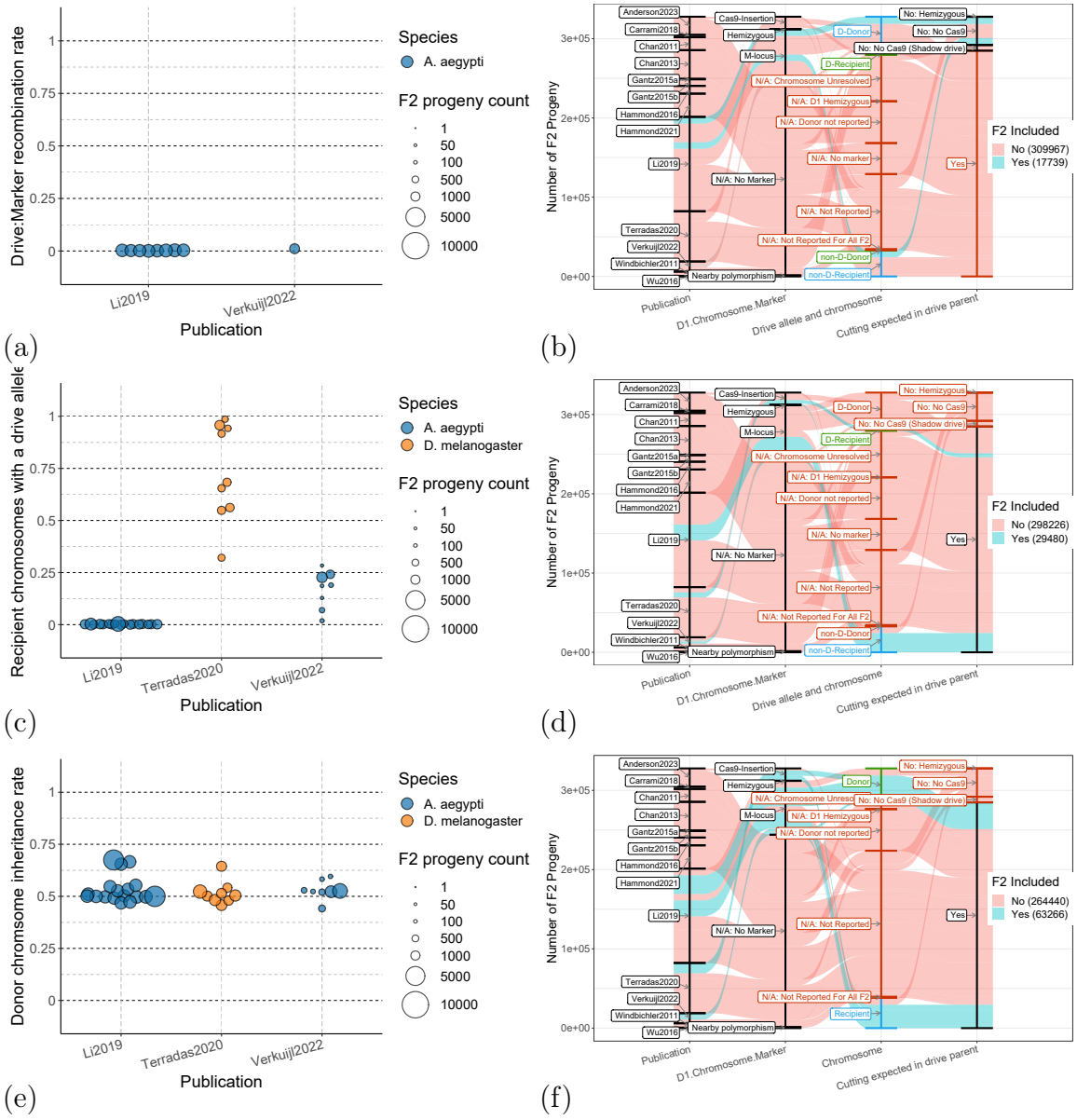


Figure S4.3: **Additional metrics can be calculated from a separate but recombinatorially linked marker.** (a) Recombination of the marker with the drive allele in the absence of drive. (nonD_Donor + D_Recipient/Total). (b) Sankey diagram indicating the subset of recorded F2 progeny for which the background recombination rate was calculated. This excluded all crosses where cutting could be expected in the drive-carrying parent. (c) Homing rate is calculated by the fraction of drive alleles on the recipient chromosome (D_Recipient/Recipient). (d) Sankey diagram indicating the subset of recorded F2 progeny for which the homing rate was calculated. (e) Meiotic drive rate is calculated as the inheritance rate of the donor chromosome. (Donor/Total). (f) Sankey diagram indicating the subset of recorded F2 progeny for which the meiotic drive rate was calculated. In the data selection Sankey diagrams, the green labels and columns indicate values considered positive for the metric being measured. The blue labels and columns indicate values that are negative but are still valid and necessary to calculate an overall rate. Red labels and columns indicate values that cause those F2 counts to be excluded, and black labels and columns indicate neutral values that do not affect the metric calculation. If one or more values cause data to be excluded, the streams between columns representing those progeny are coloured red.

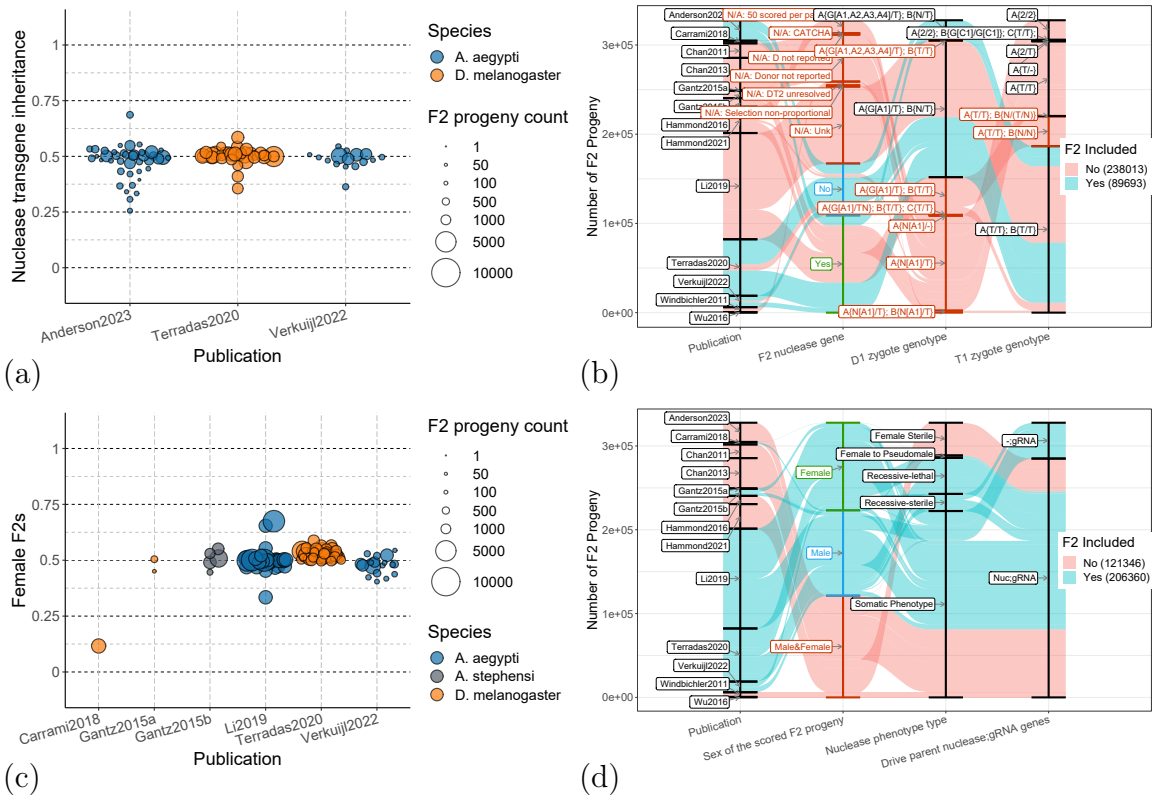


Figure S4.4: **Split-drive nuclease inheritance and F2 sex ratio.** (a) Nuclease inheritance from split-drive heterozygotes. (Nuclease carrying F2s / Total). (b) Sankey diagram indicating the subset of recorded F2 progeny for which gene drive inheritance rates were calculated. These were further reduced to only split drive genotypes (see Fig S4.1). (c) F2 sex ratio (F2 female / Total). Carrami et al. 2018 [99] targeted a gene with a sex-conversion phenotype. Li et al. 2020 [95] targeted a gene near the male sex-determining locus with a drive that functioned through meiotic drive [1]. (d) Sankey diagram indicating the subset of recorded F2 progeny for which the F2 sex-ratio could be calculated. In the data selection Sankey diagrams, the green labels and columns indicate values considered positive for the metric being measured. The blue labels and columns indicate values that are negative but are still valid and necessary to calculate an overall rate. Red labels and columns indicate values that cause those F2 counts to be excluded, and black labels and columns indicate neutral values that do not affect the metric calculation. If one or more values cause data to be excluded, the streams between columns representing those progeny are coloured red.

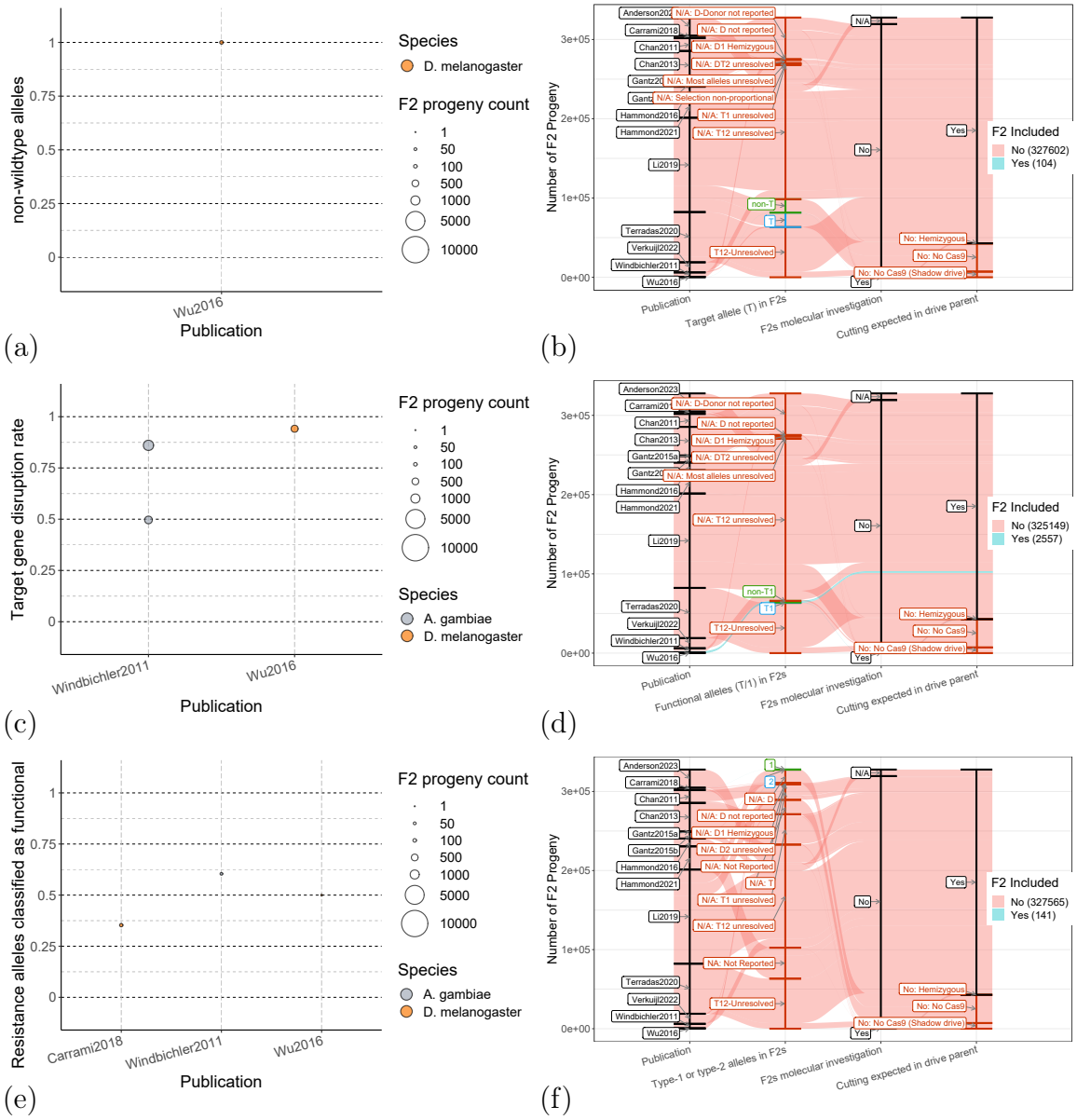
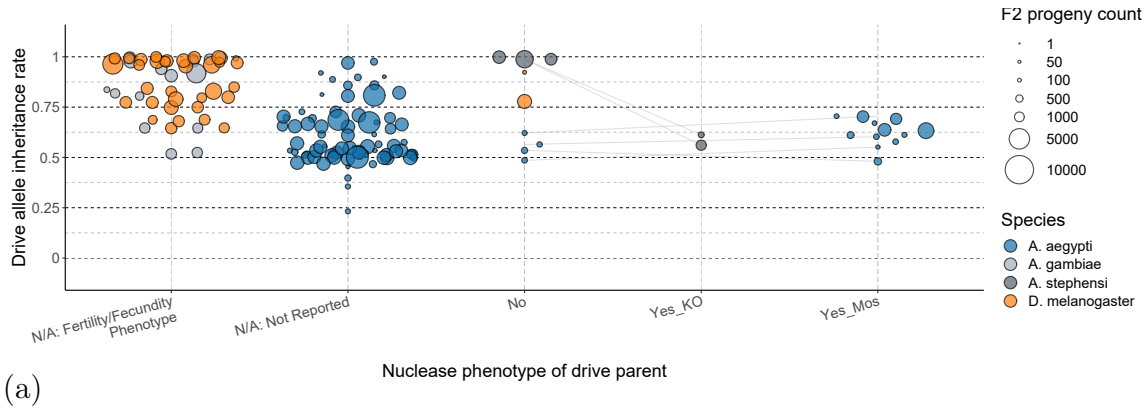


Figure S4.5: **Commonly used metrics in modelling are rarely measured experimentally** (a) The fraction of alleles that are non-wildtype (non-T / Total). This is sometimes described as the cut-rate, although it includes the non-wildtype donor alleles that were never cut (e.g., donor drive alleles). It may also include target alleles that were cut and subsequently perfectly repaired. (b) Sankey diagram indicating the subset of recorded F2 progeny selected for a. Almost all resolved T and non-T alleles are from crosses where the nuclease is absent, and these are excluded because they are not informative. (c) Fraction of non-functional alleles (non-T or non-1 / Total). This is more commonly reported as it can be derived from phenotypic screening. (d) Sankey diagram indicating the subset of recorded F2 progeny selected for c. (e) Fraction of functional resistance alleles among all resistance alleles (1 / 1 + 2). (f) Sankey diagram indicating the sub-set of recorded F2 progeny selected for e. In the data selection Sankey diagrams, the green labels and columns indicate values considered positive for the metric being measured. The blue labels and columns indicate values that are negative but are still valid and necessary to calculate an overall rate. Red labels and columns indicate values that cause those F2 counts to be excluded, and black labels and columns indicate neutral values that do not affect the metric calculation. If one or more values cause data to be excluded, the streams between columns representing those progeny are coloured red.



(a)

Figure S4.6: **Drive inheritance by drive-carrying parent phenotype.** (a) Drive allele inheritance rate separated by the drive-carrying parent's nuclease phenotype. "No" = no somatic phenotype, "Yes_KO" = a knock-out somatic phenotype, "Yes_Mos" = a mosaic somatic phenotype. Lines are drawn between paired crosses that are considered equivalent, apart from the respective X-axis factor. This generally represents the highest-quality comparative data available.

Formula	Random effects
~ 1 + Species + D1.Sex + D0.Nuclease.Sex + D0.NG_Unsorted + Nuclease.Promoter.5UTR	~ 1 | Publication/obsID

(a)

term	type	estimate	std.error	statistic	p.value
intercept	summary	-0.42	0.73	-0.58	0.56
SpeciesA. gambiae	summary	1.59	0.84	1.89	0.06
SpeciesA. stephensi	summary	1.80	1.01	1.78	0.07
SpeciesD. melanogaster	summary	1.68	0.82	2.05	0.04
D1.SexMale	summary	-0.48	0.18	-2.70	0.01
D0.Nuclease.SexMale	summary	0.28	0.23	1.26	0.21
D0.NG_UnsortedNuc;gRNA X -;-	summary	0.41	0.31	1.32	0.19
Nuclease.Promoter.5UTRbgcn	summary	0.94	0.85	1.11	0.27
Nuclease.Promoter.5UTREwald	summary	0.85	0.95	0.90	0.37
Nuclease.Promoter.5UTRexu	summary	-0.12	0.62	-0.20	0.85
Nuclease.Promoter.5UTRnanos	summary	1.15	0.81	1.43	0.15
Nuclease.Promoter.5UTRnup50	summary	0.89	0.68	1.30	0.19
Nuclease.Promoter.5UTRRcd1rp-aTub	summary	-0.22	1.67	-0.13	0.89
Nuclease.Promoter.5UTRsds3	summary	1.76	0.79	2.22	0.03
Nuclease.Promoter.5UTRshu	summary	1.63	0.83	1.96	0.05
Nuclease.Promoter.5UTRtrunk	summary	0.04	0.80	0.05	0.96
Nuclease.Promoter.5UTRubiq	summary	0.38	0.80	0.48	0.63
Nuclease.Promoter.5UTRvasa	summary	1.37	0.90	1.52	0.13
Nuclease.Promoter.5UTRvasa2	summary	1.55	1.17	1.32	0.19
Nuclease.Promoter.5UTRzpg	summary	1.20	0.81	1.49	0.14

(b)

I2_Total	I2_Publication	I2_Publication/obsID
99.69	20.34	79.35

(c)

Table S4.4: **Attributes of the fitted gene drive inheritance multivariate model.** (a) Model moderator specification and random effects. Analysis was performed using the rma.mv() function from the metafor package. A detailed explanation of individual factors/moderators can be found in supplemental section 4.7. (b) Model results on the log-odds scale. This component displays the coefficients and their corresponding standard errors, which provide information on the magnitude and direction of the effects of the model predictors on the outcome variable. Note that redundant moderators were removed from the model. (c) Model I^2 index[205]. Shown is the total amount of variance not attributable to sampling error and this variance subdivided over the model's random effects.

D1.Sex	response	std.error	df	null	statistic	p.value
Female	0.92	0.05	Inf	0.50	3.43	0.00
Male	0.88	0.08	Inf	0.50	2.68	0.01

(a)

term	contrast	null.value	odds.ratio	std.error	df	null	statistic	p.value
D1.Sex	Female / Male	0.00	1.61	0.29	Inf	1.00	2.70	0.01

(b)

Table S4.5: **Estimated marginal means for drive parent sex.** (a) Estimated marginal means. Values are backtransformed to proportions from log-odds. Infinite degrees of freedom correspond to the estimates being tested against the standard normal distribution. (b) Tukey's post hoc pairwise comparisons of the estimated marginal means.

D0.Nuclease.Sex	response	std.error	df	null	statistic	p.value
Female	0.89	0.07	Inf	0.50	2.88	0.00
Male	0.91	0.06	Inf	0.50	3.18	0.00

(a)

term	contrast	null.value	odds.ratio	std.error	df	null	statistic	p.value
D0.Nuclease.Sex	Female / Male	0.00	0.75	0.17	Inf	1.00	-1.26	0.21

(b)

Table S4.6: **Estimated marginal means for nuclease carrying grandparent's sex.** (a) Estimated marginal means. Values are backtransformed to proportions from log-odds. Infinite degrees of freedom correspond to the estimates being tested against the standard normal distribution. (b) Tukey's post hoc pairwise comparisons of the estimated marginal means.

D0.NG_Unsorted	response	std.error	df	null	statistic	p.value
Nuc;- X -;gRNA	0.84	0.09	Inf	0.50	2.44	0.01
Nuc;gRNA X -;-	0.89	0.07	Inf	0.50	3.04	0.00

(a)

term	contrast	null.value	odds.ratio	std.error	df	null	statistic	p.value
D0.NG_Unsorted	(Nuc;-X-;gRNA)/(Nuc;gRNAX-;-)	0.00	0.66	0.21	Inf	1.00	-1.32	0.19

(b)

Table S4.7: **Estimated marginal means for grandparent deposition form.** Nuc;-X -;gRNA indicates a cross in which a nuclease-carrying individual, male or female, was crossed to a gRNA-carrying individual. Nuc;gRNA X -;- indicates a cross where an individual, male or female, carrying both the nuclease and gRNA transgene was crossed to an individual carrying neither. (a) Estimated marginal means. Values are backtransformed to proportions from log-odds. Infinite degrees of freedom correspond to the estimates being tested against the standard normal distribution. (b) Tukey's post hoc pairwise comparisons of the estimated marginal means.

Formula	Random effects
~ 1	~ 1 | Publication/Cross.Name/obsID

(a)

term	type	estimate	std.error	statistic	p.value
overall	summary	-2.19	0.47	-4.64	0.00

(b)

pred	ci.lb	ci.ub	pi.lb	pi.ub
0.10	0.04	0.22	0.00	1.00

(c)

I2_Total	I2_Publication	I2_Publication/Cross.Name	I2_Publication/Cross.Name/obsID
99.36	3.19	18.12	78.05

(d)

Table S4.8: **Attributes of the intercept-only nuclease phenotype multivariate model.** (a) Model moderator specification and random effects. Analysis was performed using the rma.mv() function from the metafor package. (b) Model results on the log-odds scale. This component displays the coefficients and their corresponding standard errors, which provide information on the magnitude and direction of the effects of the model predictors on the outcome variable. (c) Results of the model back-transformed from log-odds to F2 somatic phenotype rate. The backtransformation provides a more interpretable representation of the model results. (d) Model I^2 index[205]. Shown is the total amount of variance not attributable to sampling error and this variance subdivided over the model's random effects.

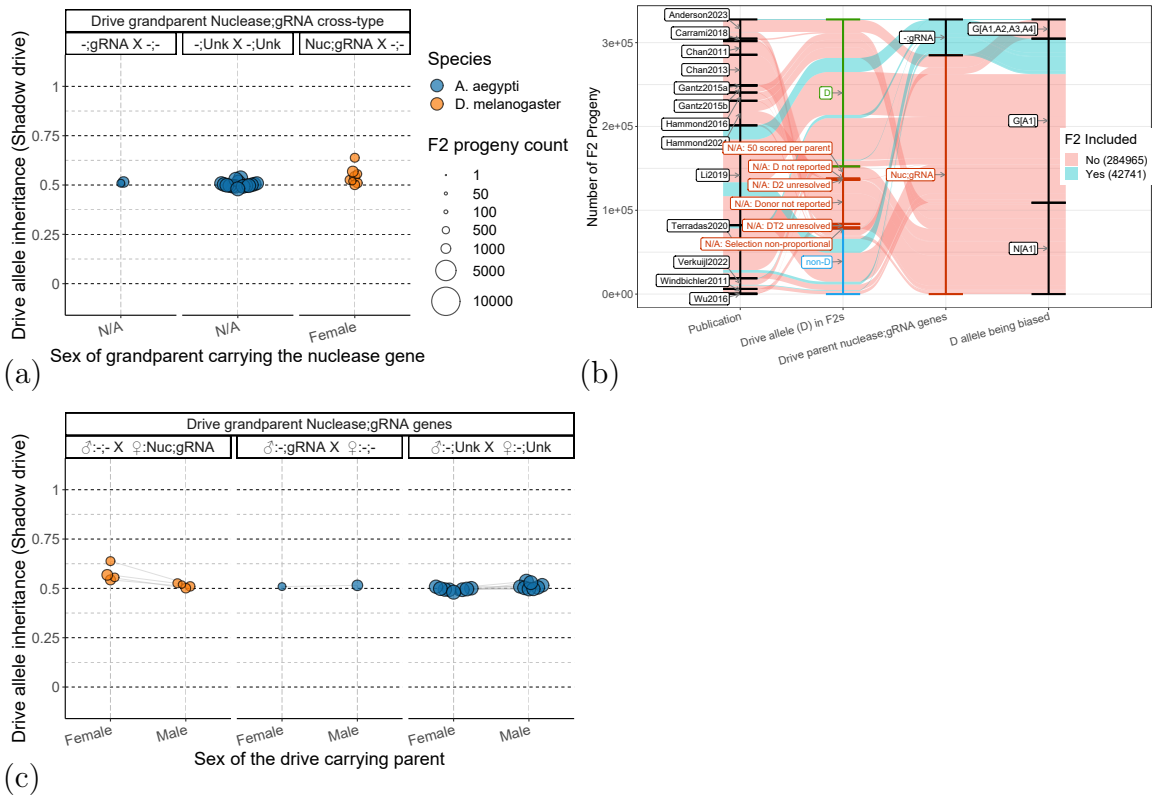


Figure S4.7: **Shadow drive has been reported to depend on the sex of the individuals receiving the deposition.** Shadow drive is the biasing of the gene drive allele in the absence of one or more expressed nuclease components (invariably the nuclease protein). (a) Shadow drive by the sex of the nuclease-carrying grandparent. In the current data set, we do not have data for both sexes. We have included crosses where no grandparent carried the nuclease gene and cases where the nuclease carrying grandparent was not specified (shadow drive not expected). (b) Sankey diagram indicating the subset of recorded F2 progeny for which shadow drive-based gene drive inheritance rates were calculated. In the data selection Sankey diagrams, the green labels and columns indicate values considered positive for the metric being measured. The blue labels and columns indicate values that are negative but are still valid and necessary to calculate an overall rate. Red labels and columns indicate values that cause those F2 counts to be excluded, and black labels and columns indicate neutral values that do not affect the metric calculation. If one or more values cause data to be excluded, the streams between columns representing those progeny are coloured red. (c) Shadow drive by the sex of the drive-carrying parent that received maternal nuclease deposition. In these crosses, the drive parent does not carry the nuclease transgene.

Formula	Random effects
~ 1 + F1.NG_Unsorted + F1.Possible.Germline.Mutation.Expected + F2.KO.Req + F2.NG + F1.Nuclease.Sex + Nuclease.Promoter.5UTR + F2.Sex	~ 1 | Publication/Cross.Name/obsID

(a)

term	type	estimate	std.error	statistic	p.value
intercept	summary	-8.36	4.45	-1.88	0.06
F1.NG_UnsortedNuc;gRNA X -;-	summary	1.43	3.28	0.43	0.66
F2.KO.Req2	summary	1.51	3.86	0.39	0.69
F2.NG-;gRNA	summary	3.45	3.78	0.91	0.36
F2.NGNuc;-	summary	-0.26	0.53	-0.50	0.62
F2.NGNuc;gRNA	summary	6.32	3.76	1.68	0.09
F2.NGUnk;-	summary	-0.48	3.25	-0.15	0.88
F2.NGUnk;gRNA	summary	6.11	4.96	1.23	0.22
F1.Possible.Germline.Mutation.ExpectedYes	summary	-0.03	0.91	-0.03	0.97
F1.Nuclease.SexMale	summary	-3.28	0.28	-11.70	0.00
Nuclease.Promoter.5UTRbgcn	summary	3.32	1.11	3.00	0.00
Nuclease.Promoter.5UTREwald	summary	-0.02	1.59	-0.01	0.99
Nuclease.Promoter.5UTRexu	summary	1.41	0.74	1.89	0.06
Nuclease.Promoter.5UTRnanos	summary	1.15	1.53	0.75	0.45
Nuclease.Promoter.5UTRnup50	summary	4.27	0.83	5.14	0.00
Nuclease.Promoter.5UTRsds3	summary	3.99	1.20	3.34	0.00
Nuclease.Promoter.5UTRshu	summary	3.72	1.42	2.61	0.01
Nuclease.Promoter.5UTRtrunk	summary	1.66	0.94	1.77	0.08
Nuclease.Promoter.5UTRubiq	summary	3.14	0.95	3.32	0.00
Nuclease.Promoter.5UTRvasa	summary	4.44	3.12	1.42	0.16
Nuclease.Promoter.5UTRzpg	summary	-0.31	1.51	-0.21	0.84
F2.SexMale	summary	-0.04	0.30	-0.14	0.89
F2.SexMale&Female	summary	0.57	3.26	0.18	0.86

(b)

I2_Total	I2_Publication	I2_Publication/Cross.Name	I2_Publication/Cross.Name/obsID
99.08	38.77	2.19	58.13

(c)

Table S4.9: **Attributes of the fitted somatic phenotype multivariate model.** (a) Model moderator specification and random effects. Analysis was performed using the rma.mv() function from the metafor package. Note that redundant predictors were dropped from the model. A detailed explanation of individual factors/moderators can be found in supplemental section 4.7. (b) Model results on the log-odds scale. This component displays the coefficients and their corresponding standard errors, which provide information on the magnitude and direction of the effects of the model predictors on the outcome variable. (c) Model I^2 index[205]. Shown is the total amount of variance not attributable to sampling error and this variance subdivided over the model's random effects.

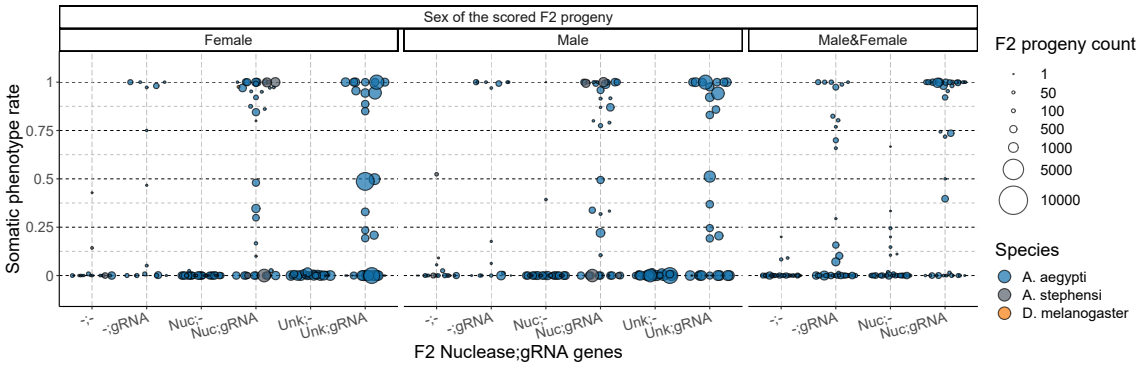
F2.Sex	response	std.error	df	null	statistic	p.value
Female	0.99	0.04	Inf	0.50	1.16	0.24
Male	0.99	0.04	Inf	0.50	1.14	0.25
Male&Female	0.99	0.03	Inf	0.50	1.03	0.30

(a)

term	contrast	null.value	odds.ratio	std.error	df	null	statistic	adj.p.value
F2.Sex	Female / Male	0.00	1.04	0.31	Inf	1.00	0.14	0.99
F2.Sex	Female / Male&Female	0.00	0.56	1.84	Inf	1.00	-0.18	0.98
F2.Sex	Male / Male&Female	0.00	0.54	1.77	Inf	1.00	-0.19	0.98

(b)

Table S4.10: **Estimated marginal means for F2 sex.** (a) Estimated marginal means. Values are backtransformed to proportions from log-odds. Infinite degrees of freedom correspond to the estimates being tested against the standard normal distribution. Male& Female are cases where the sex was not reported. (b) Tukey's post hoc pairwise comparisons of the estimated marginal means.



(a)

Figure S4.8: **Somatic phenotype rates increase dramatically when the F2 progeny are capable of active nuclease expression.** (a) Somatic phenotype rate compared between F2s that can and cannot express all components of the nuclease. The most common comparison is between split-drive crosses where gRNA-positive F2 progeny did or did not inherit the nuclease-expressing gene. In most cases, the gRNA disrupts the target gene, and as such, we cannot fairly compare the presence or absence of the gRNA in nuclease-positive progeny. Note that the F2 genotype depends on the F2's nuclease and gRNA genotype. As such, paired comparisons cannot be made for this graph because a difference in gRNA or nuclease status necessary means a difference in F2 genotype.

F1.Nuclease.Sex	response	std.error	df	null	statistic	p.value
Female	1.00	0.01	Inf	0.50	1.59	0.11
Male	0.96	0.17	Inf	0.50	0.77	0.44

(a)

term	contrast	null.value	odds.ratio	std.error	df	null	statistic	p.value
F1.Nuclease.Sex	Female / Male	0.00	26.48	7.41	Inf	1.00	11.70	0.00

(b)

Table S4.11: **Estimated marginal means for nuclease carrying parent sex.** (a) Estimated marginal means. Values are backtransformed to proportions from log-odds. Infinite degrees of freedom correspond to the estimates being tested against the standard normal distribution. (b) Tukey's post hoc pairwise comparisons of the estimated marginal means.

F1.NG_Unsorted	response	std.error	df	null	statistic	p.value
Nuc;- X -;gRNA	0.98	0.08	Inf	0.50	0.91	0.36
Nuc;gRNA X -;-	1.00	0.02	Inf	0.50	1.27	0.20

(a)

term	contrast	null.value	odds.ratio	std.error	df	null	statistic	p.value
F1.NG_Unsorted	(Nuc;-X-;gRNA)/(Nuc;gRNAX-;-)	0.00	0.24	0.79	Inf	1.00	-0.43	0.66

(b)

Table S4.12: **Estimated marginal means for parent deposition form.** (a) Estimated marginal means. Values are backtransformed to proportions from log-odds. Infinite degrees of freedom correspond to the estimates being tested against the standard normal distribution. Nuc;- X -;gRNA indicates a cross in which a nuclease-carrying individual, male or female, was crossed to a gRNA-carrying individual. Nuc;gRNA X -;- indicates a cross where an individual, male or female, carrying both the nuclease and gRNA transgene was crossed to an individual carrying neither. (b) Tukey's post hoc pairwise comparisons of the estimated marginal means.

4.7 Supplemental Methods

Database variables

Factors that must be identical for a cross to be considered paired are indicated with a † symbol (this applies to all metrics including drive inheritance). Additional pairing factors specific to the somatic phenotype metric are indicated with a ‡ symbol. Additional factors have been created by performing modifications, extrapolations, simplifications, and combinations of these factors.

Publication metadata

†Publication: A unique short identifier of the source publication. Data type: string (e.g., "Anderson2023").

Long.Name: First author and title of the article. Data type: string.

URL: A web address of the publication record. Data type: string.

DOI: A Digital Object Identifier (DOI) for the publication. Data type: string.

Cross: A unique identifier for each cross within a publication. Generally ordered by the appearance in the source publication. Data type: integer.

ID: A unique identifier for each cross in the whole database. It is composed of the "Cross" and "Publication" separated by an underscore. Data type: string (e.g., "01_Anderson2023").

Cross.Name: A descriptive name for the cross can be used to identify it from the other crosses in that publication. It generally mirrors the distinction or intended investigation in the source publication. Data type: string (e.g., "D1.F_vasa_TS1").

Row.Data.Type: Data type of the row. If set to "CrossData", this indicates that the row contains the values extracted from the publication. If set to "Justification", this indicates that the row consists of strings in each column that indicate the source/justification for the cross-data with the same ID. Data type: string (i.e., "CrossData", "Justification").

Source: The principle reference within the publication for the F2 (phenotype) data. Data type: string (e.g., "Fig1", "TableS1").

Notes: Miscellaneous information such as important assumptions relevant to the interpretation of this specific cross or publication. Data type: string.

†Species: The name of the organism being studied. Data type: string (e.g., "D. melanogaster").

Parent and grandparent factors

D1.Nuc.Pheno.Gene: The gene associated with the nuclease activity phenotype. Data type: string (e.g. "ebony", "white", "yellow", "N/A").

D1.Nuc.Pheno.Type: The description of the type of phenotype associated with the nuclease activity phenotype. Data type: string (e.g. "Somatic Phenotype", "Female to Pseudomale").

†D1.Nuc.Pheno: A phenotype associated with nuclease activity. Most often, a phenotype that is informative about the cutting of the target allele in an individual heterozygous for the drive allele and a target allele. Data type: string (i.e., "Yes", "Yes_KO", "Yes_Mos", "No", "N/A").

†**D1.Sex:** D1 is the F1 parent that is being evaluated (family tree shown in Fig 4.2e), and D1.Sex is the sex of that group/individual. In almost all cases, one sex is of interest for the gene drive process in the cross (e.g., drive-carrying males crossed to wild-type females). In most crosses, this is an individual or group of individuals that are double heterozygous for the gRNA and Cas9 drive element. For preparation / control crosses (e.g., gRNA individuals crossed to Cas9 carriers), the sex carrying the element designed to be biased is designated as D1. Data type: string (i.e., "Male", "Female"). Note, that this factor is removed for pairing of the somatic phenotype metric ‡.

D1.Allele.Biased: A functional description of the allele of interest that has the capacity to be biased (but not necessarily in this cross). See Fig S4.1 for an illustration of the genotype notation system that we use. This allele will always be at the first Loci. Data type: string (e.g., $N^{A}T$).

D1.Cutting.Expected: The expectation of whether a target allele in the D1 individual could be cut. This only takes into account the presence of Cas9 and at least one gRNA gene and the presence of a valid target. Sex-specific promoters are not taken into account. Data type: string ("Yes", "No: No Cas9").

D1.GESP: Notation of the possibility of a nuclease component being biased in the D1 individual. This is a proposed explanation for the grand parent enhanced somatic phenotype (GESP) effect described in Verkuijl et al.[1]. Data type: string (i.e., "Nuc;gRNA", "-;gRNA", "Nuc;-", "-;-").

‡**D1.GESP.Expected:** Combined factor: D1.GESP only in cases where cutting may be expected.

†**T1.Cutting.Expected:** Same as D1.Cutting.Expected but for the T1 parent.

T1.GESP: Same as D1.GESP but for the T1 parent.

‡**T1.GESP.Expected:** Combined factor: T1.GESP only in cases where cutting may be expected.

(‡ for F2) "X".Zygote.Genotype: "X" = "Male0.D1", "Female0.D1", "Male0.T1", "Female0.T1", "D1", "T1", "F2" (family tree shown in Fig 4.2e). A functional description of the genotype at each locus in the "X" individual(s) as zygotes. See Fig S4.1 for an illustration of the genotype notation system that we use. This contains information on the number of gRNAs, from which locus they are expressed, and which locus they target. In addition, it describes from what locus the nuclease is expressed and whether resistance alleles are present. Data type: string (e.g., N^{A}/T).

(‡ for F2) "X".NG: "X" = "Male0.D1", "Female0.D1", "Male0.T1", "Female0.T1", "D1", "T1", "F2" (family tree shown in Fig 4.2e). Notation of the presence of the nuclease gene and at least one gRNA capable of cutting the nuclease phenotype gene. Data type: string (i.e., "Nuc;gRNA", "-;gRNA", "Nuc;-", "-;-").

‡**F1.NG_Unsorted:** A combination of the F1 (D1 and T1) genotype descriptions. This description is "unsorted" in regards to sex. This means that if the sexes of the nuclease carrying F1s are swapped, the value is the same. Data type: string (i.e., "Nuc;gRNA X -;-", "Nuc;- X -;gRNA").

†**D0.NG_Unsorted:** A combination of the D0 genotype descriptions. This description is "unsorted" in regards to sex. This means that if the sexes of the

nuclease carrying D0 grandparents are swapped, the value is the same. Data type: string (i.e., "Nuc;gRNA X -;-", "Nuc;- X -;gRNA").

†D0.Nuclease.Sex: Derived from Male0.D1.NG and Female0.D1.NG. The sex of the nuclease-carrying grandparent. Data type: string (i.e., "Male", "Female", "Both", "None").

D1.T1.Cross.Scale: The number of D1 and T1 individuals crossed and in what way. Data type: string (e.g., "Pool: M:10 X F:51").

D1.Parent: An identifier for the different parents or pool within a single cross. Data type: string (e.g., "1", "2", "Pool").

F2 factors

†Stage.Screened: Life stage for which the data apply. Data type: string (e.g., "Larvae/pupae").

F2.Count: The number of F2 progeny that make up the specific grouping that this row represents. Use to calculate numerical outcomes. In some cases, rows with 0 observations are included to maintain a consistent data structure during data recording. Data type: integer.

‡F2.KO.Req: Number of allele required to be disrupted in this F2 genotype to generate a somatic phenotype. Data type: string (e.g., "0", "1", "1-2", "2").

(‡ for F2) F2.Sex: The sex of the F2 progeny. Data type: string (i.e., "Male", "Female", "Unk").

F2.Mol.Check: Notes if additional molecular assays were performed that helped narrow down the specific genotype. Data type: string (i.e., "Yes", "No").

F2.Seq.File: A URL link to a sequencing file associated with an additional assay of this group or individual F2. Data type: string.

‡D1.Possible.Germline.Mutation: Notation of the possibility of a resistance mutation having been inherited by this F2. Data type: string ("Yes", "No").

F2.D1.Allele.Inherited: The drive locus allele inherited by the F2 progeny from the D1 parent. This classification takes into account all the phenotypic and molecular data. See Fig S4.1 for an illustration of the genotype notation system that we use. Data type: string (e.g., "D", "1", "2", "T", "D; 2", "1; T").

‡T1.Possible.Germline.Mutation: Same as D1.Possible.Germline.Mutation but for the T1 parent.

F2.T1.Allele.Inherited: Same as F2.D1.Allele.Inherited but for the T1 parent.

F2.Nuc.Pheno: A phenotype associated with nuclease activity. Most often, a phenotype that is informative about the cutting of the target allele in an individual heterozygous for the drive allele and a target allele. Data type: string (i.e., "Yes", "Yes_KO", "Yes_Mos", "No", "N/A").

F2.Nuc.Pheno.Informative: A judgement of whether the presence or lack of a nuclease-associated phenotype is informative about the nuclease expression/deposition pattern. Note, to limit confusion, we exclude rare cases where the phenotype mediating gRNA is inherited from a different F1 parent than the drive gRNA (e.g., Wu2016[201]). Data type: string ("Yes", "No", "N/A").

F2.Nuclease: An indication of whether this group of F2 has inherited the nuclease protein gene. Data type: string (e.g.,

"Yes", "No").

F2.D1.D: An indication of whether the allele inherited from the D1 parent is the drive allele (the allele that is informative of inheritance bias). D describes all possible types of drive alleles (even those that do not carry nuclease components). Note that in some cases, the assay performed cannot differentiate a drive element from a non-functional resistance mutation. In those cases, we note "N/A" (even if in a subset of progeny the absence of a drive element can be determined) to avoid biasing the estimate of the drive inheritance rate. Data type: string (e.g., "non-D", "D", "N/A:D2-indistinguishable").

F2.D1.Chromosome: An indication of whether the allele inherited from the D1 parent is on the donor or recipient chromosome. This can be predicted by a marker (gene or sequence) closely linked to the drive element but not contained within the drive homology arms. Note that if no marker is present, we record "N/A" even if the lack of a drive element implies that the recipient chromosome was inherited. In these cases, the analysis cannot be performed because the drive alleles on the recipient chromosome cannot be differentiated from those on the donor chromosome. Data type: string (e.g., "Recipient", "Donor", "N/A").

F2.D1.D.Chromosome: A combination of F2.D1.D and F2.D1.Chromosome that is also determined when the data are recorded. This is used to calculate homing and background recombination rates. Data type: string (e.g., "non-D-Donor", "D-Donor", "non-D-Recipient", "D-Recipient").

F2.D1.D.or.2.Recipient: Metric included to accommodate the Chan et al.[68] and Chan et al.[78] data. Indicates drive or type-2 resistance alleles on the recipient chromosome. (e.g., "D-Recipient", "2-Recipient").

F2.D1.T: An indication of whether the drive locus allele is an intact target allele or any other allele. Note that in many cases, the assay performed cannot differentiate a target allele from a functional resistance mutation. In those cases, we note "N/A" to exclude them from biasing the estimate (even for cases where "non-T" could be correctly allocated). Data type: string (i.e., "non-T", "T", "N/A").

F2.D1.1andT: An indication of whether the drive locus allele is an intact target allele or a functional resistance allele. Data type: string (i.e., "non-1T", "1T", "N/A").

F2.D1.1or2: An indication of whether the drive locus allele is a functional or non-functional resistance allele. This can generally only be determined with molecular assays. Data type: string (i.e., "1", "2", "N/A").

F2.D.Allele.Type: A note on whether the drive element associated with the nuclease phenotype disrupts the function of the gene it associates with. "Trans" indicates the nuclease phenotype is generated by a gene distinct from the drive element. Data type: string (e.g. "KO", "Rescue", "Trans" "N/A").

†**F2.Nuc.Pheno.Locus:** The locus identifier associated with the nuclease activity phenotype. Data type: string (e.g. "A", "C").

†**F2.Nuc.Pheno.Gene:** The gene associated with the nuclease activity phenotype. Data type: string (e.g. "ebony", "white", "yellow").

†**F2.Nuc.Pheno.FitType:** The descrip-

tion of the type of phenotype associated with the nuclease activity phenotype. Somatic.Phenotype: white, yellow, ebony. Female.to.Pseudomale.Conversion: tra. Data type: string (e.g. "Somatic.Phenotype", "Female.to.Pseudomale.Conversion").

†**F2.Nuc.Pheno.Sufficiency:** A note on whether the nuclease activity phenotype gene is expected to be haplosufficent or haploinsufficient. Data type: string (e.g., "Haplosufficient", "Haploinsufficient").

Targeting genes

Loci.Count: The number of loci that are directly relevant to the interpretation of this cross. Data type: integer.

Loci: Identifier (e.g. gene name or chromosome position) of the loci. With multiple loci, they are separated by a ";" symbol (e.g. white; ebony). These loci are referred to by alphabetical designations (i.e. A, B, C, etc.). The locus of the allele that can be biased is always A. Data type: string.

†**Target.Gene:** A list of genes targeted by the gRNAs/DNA-binding protein involved in this cross. Data type: string (e.g., "A:white; B:yellow"). D.Target.Gene is a derived factor that only includes the drive allele locus.

†**Target.FitType:** A description of the expected consequence of loss of function of the target gene(s). Data type: string (e.g., "A:synthetic.neutral; C:putative.neutral"). D.Target.FitType is a derived factor that only includes the drive allele locus.

†**Target.Sufficiency:** A note on whether the target gene is expected to be haplosufficent or haploinsufficient. Data type: string (e.g., "A:Haplosufficient", "A:Haploinsufficient"). D.Target.Sufficiency is a derived factor that only includes the drive allele locus.

Target.Multiplex: The number of gRNAs/DNA-binding protein that target each locus in this cross. Data type: string (e.g., "A:1; B:2").

†**gRNA.Target.Sequence:** The DNA-binding protein or gRNA+PAM target sequence. The expected cut site is indicated with a "/" symbol. Multiple gRNAs are separated by the locus they target (e.g. "A1:...; B1:..."). Multiplexing of the same locus is differentiated by numbers (e.g., "A1:...; A2:...") Data type: string (e.g., "A1:GGCTACGCCGGCT ACAT/TGANGG; C1:GCCACAA TTGTCGATCG/TCANGG"). D.gRNA. Target.Sequence is a derived factor that only includes the drive allele locus.

†**gRNA.Target.Site:** A description of the gRNA target sequence within the target gene. Data type: string (e.g., "A1:CDS", "A1:Promoter"). D.gRNA.Target.Site is a derived factor that only includes the drive allele locus.

†**gRNA.Promoter:** A list of the gRNA promoters involved in this cross. The names and sequences of the promoters are matched to those described in Anderson et al. 2020[156], For sequence specific proteins the protein promoter is listed. Data type: string (e.g., "A1:DmU6-2(CR32867) ; C1:DmU6-3(CR31539)"). D.gRNA.Promoter is a derived factor that only includes the drive allele locus.

gRNA.Scaffold: A list of the gRNA scaffold/backbone sequences involved in this cross. This is set to "N/A" for cutting by sequence-specific proteins. Data type: string (e.g., "A1:WT-76nt").

Nuclease factors:

The nuclease and gRNA factors described here will not be strictly limited to the F1 parents. For example, they may describe nuclease expression in an F0 grandparent if shadow drive is being evaluated and the D1 parent only carries a gRNA. This implicit labelling method has so far not led to a conflict, but if it does, those crosses may have to be excluded. Alternatively, an additional layer of labelling can be introduced to differentiate distinct nuclease and targeting profiles within the same cross.

†**Nuclease.Insertion.Variant:** An identifier of equivalent nuclease expressing elements inserted at different loci. Note that we do not consider independently derived HDR events as insertion variants (e.g., 10.1 and 10.2 in Gantz 2015[53]). If a transgene is reused in a subsequent study, we use the same insertion identifier from the first publication. Data type: string (e.g., "C1_shu", "C2_shu", "N/A: No Cas9").

†**Nuclease.Promoter.5UTR:** The nuclease promoter and the 5′UTR involved in this cross. Data type: string (e.g., "vasa").

Nuclease.Promoter.5UTR.Size: The size in bp of the nuclease promoter and the 5′UTR involved in this cross. Data type: integer (e.g., 2000).

†**Nuclease.Gene:** The nuclease gene involved in this cross. Data type: string (e.g., "Cas9", "SceI").

Nuclease.CDS: A description of the coding sequence of the nuclease gene. Data type: string (e.g., "3xFLAG-NLS-Cas9-NLS-T2A-eGFP").

Nuclease.Codon.Optimisation: The codon optimisation used for the nuclease gene involved in this cross. Data type: string (e.g., "Mammalian", "Insect").

Nuclease.3UTR: The 3′UTR involved in this cross. Data type: string (e.g., "vasa").

Nuclease.3UTR.Size: The size in bp of the 3′UTR involved in this cross. Data type: integer (e.g., 500).

Sequences

Drive.Inserted.Link: A URL link to a sequence file of the drive element. Data type: string.

Drive.Plasmid.Link: A URL link to a sequence file of the drive element plasmid. Data type: string.

Target.Locus.Sequence.Link: A URL link to a sequence file of the target-site. Data type: string.

Nuclease.Inserted.Link: A URL link to a sequence file of the nuclease expressing element (this may be the same as Drive.Inserted.Link). Data type: string.

Nuclease.Inserted.Link: A URL link to a sequence file of the drive element plasmid (this may be the same as the Drive.Plasmid.Link). Data type: string.

Additional.Elements.Inserted: A URL link to the sequence file(s) of additional drive elements relevant to this cross. Data type: string.

Additional.Elements.Plasmid: A URL link to the sequence file(s) of additional drive element plasmids relevant to this cross. Data type: string.

General factors

D1.Allelic.or.HACK: Indicating whether the region that is intended to be homed from is in the same position on the homologous chromosome (allelic), or located elsewhere in the genome (HACK[194])

as in the case of Terradas et al. 2021[116]. Note that anything other than an allelic position is excluded from current analyses. Data type: string (e.g., "Allelic").

D1.Chromosome.Marker: Indication of the presence of a marker (gene or sequence) linked to the drive element that can be used to distinguish homing from other forms of inheritance bias. Data type: string (i.e., "M-Locus", "Cas9-Insertion").

†**Temperature:** A description of the temperature at which the D1 individuals were kept. Data type: string (e.g., "25C", "N/A"). This is included to differentiate a set of paired crosses in Chan et al.[78].

Experimental methods

It is common for a subset of progeny to be selected for molecular investigation, for example, all progeny of a particular parent. These data may reasonably be taken as representative of the cross and used to estimate a particular metric (e.g., type-1 to WT allele ratio). In another case, all progeny from a single parent may by chance present with a phenotype that resolves the zygotic genotypes. The latter case will be more likely with a low progeny count and outlier data where ambiguous genotypes are absent by chance. To avoid this, we need to specify when alleles have been resolved through a particular experiment as opposed to by chance. In addition, we found that we have to do this independently for different metrics. For example, in some studies[68;78;86] only data for the recipient chromosome are reported. This allows some metrics to be calculated (fraction of recipient chromosomes carrying the drive), but not others (ratio of donor to recipient chromosomes). Extrapolating multiple metrics from the small subset of factors did not allow us to capture this complexity and was found to lead to hard-to-anticipate errors in extrapolation. As such, the metrics listed in Fig 4.3d are recorded as a separate factor during data recording to ensure that the metrics are only calculated where appropriate. If a particular metric is not recorded, we have briefly justified why.

We have attempted to make the data set the closest possible reflection of the experiments performed. As such, we have avoided extrapolations or assumptions that may have been made in the original publication. The most common example of this is measuring the ratio of certain repair outcomes in a subset of progeny, and then extrapolating this to other progeny. We only report the ratio of these outcomes in the cross and progeny where they have been measured. Another example is the previously mentioned issue of omitting the number of F2 that inherited the drive-carrying donor chromosome as this (presumably) did not substantially diverge from the expected 50% Mendelian inheritance rates. Unfortunately, this limit on extrapolation often causes data to be discarded because common metrics (e.g., Drive inheritance rate) cannot be calculated without making a certain, on the face of it reasonable, extrapolation. Part of the goal of this project is to make the large body of gene drive literature more accessible and allow researchers to more easily identify useful prior results that then can be evaluated more comprehensively by reading the associated publication. As such, while these results will be omitted from the standard drive inheritance analyses we perform, we have included a range of metrics in the associate interactive tool,

including those that are a reflection of the limitations of experimental methods that nonetheless highlight the contribution of this work.

Another limitation of our system comes from the fact that we assign all metrics to individual progeny simultaneously. This comes into conflict when certain details are omitted when in-depth investigations such as sequencing are described. For example, in [108], a subset of the progeny of certain crosses was sequenced but their sex and Cas9 transgene status are not described (and in the context of that paper there is no reason why they should be). However, the sex and Cas9 status of the progeny are described when it comes to drive inheritance. Due to our database structure, we cannot accommodate those sequencing results without being forced to specify the sex of the progeny and the Cas9 status as that has been filled in from the drive inheritance data. Currently, this causes us to omit the sequencing results. A solution may be to allocate those data to specific progeny with a known Cas9 and sex but, for each metric, specify if certain factors should be treated as unknown when that metric is used.

Chapter 5

A CRISPR endonuclease gene drive reveals distinct mechanisms of inheritance bias

Sebald A. N. Verkuijl[1,2,*], Estela Gonzalez[2,3,*], Ming Li[4], Joshua X. D. Ang[2,3], Nikolay P. Kandul[4], Michelle A. E. Anderson[2,3], Omar S. Akbari[4], Michael B. Bonsall[1], Luke Alphey[2,3]

1: Mathematical Ecology Research Group, Department of Biology, University of Oxford, 11a Mansfield Road, Oxford, OX13SZ, U.K.
2: Arthropod Genetics, The Pirbright Institute, Ash Road, Pirbright GU24 0NF, U.K.
3: Current Address- The Department of Biology, University of York, York, U.K.
4: School of Biological Sciences, Department of Cell and Developmental Biology, University of California San Diego, La Jolla, California, United States, 92093
*: These authors contributed equally to this work

In the previous chapter, we performed several analyses using recombinatorially linked markers. In many cases, these markers are coincidental and not investigated in the original publication. In one of the publications for which we performed this analysis, we noted that the inheritance bias they reported from males did not coincide with the molecular hallmarks of homing. Instead, the inheritance bias occurred through a different mechanism. This group had already shared with us the transgenic lines used in that study. We set out to further investigate this unexpected mechanism with follow-up experiments.

5.1 Abstract

CRISPR/Cas gene drives can bias transgene inheritance through different mechanisms. Homing drives are designed to replace a wild-type allele with a copy of a drive element on the homologous chromosome. In *Aedes aegypti*, the sex-determining locus is closely linked to the *white* gene, which was previously used as a target for a homing drive element (w^{GDe}). Here, through a novel analysis using this linkage, we show that in males inheritance bias of w^{GDe} did not occur by homing, but rather through increased

propagation of the donor drive element. We test the same w^{GDe} drive element with transgenes expressing Cas9 with germline regulatory elements *sds3*, *bgcn*, and *nup50*. We only find inheritance bias through homing, even with the identical *nup50*-Cas9 transgene. We propose that DNA repair outcomes may be more context-dependent than anticipated and that other previously reported homing drives may, in fact, bias their inheritance through other mechanisms.

5.2 Introduction

Genetic modification of wild populations through gene drive may be a means of addressing some of the most pressing public health challenges in the world. Gene drive is the ability of a genetic element to bias its own inheritance, allowing it to spread a genetic change throughout a population[63]. There are many examples of natural gene drives that act through different inheritance biasing mechanisms[64]. Some types of gene drive function through the action of enzymes that create sequence-specific DNA breaks (DNA endonucleases), and various context-dependent cellular repair mechanisms exist to resolve DNA breaks[104]. Correspondingly, nuclease-based gene drives can function through different mechanisms including inheritance bias through a copying mechanism (homing drives) and drives that cause the loss of non-drive-bearing gametes or offspring (here referred to as meiotic drive).

Generally, in diploid organisms, each parent contributes one chromosome of each homologous pair and each allele has a 50% chance of being passed on to a given progeny (Mendelian inheritance). Synthetic homing and meiotic endonuclease gene drives both rely on selectively creating double-strand DNA breaks on the non-drive-bearing homolog. Through different mechanisms, this results in an inheritance bias of an allele or genomic region and, for meiotic drive, potentially the entire chromosome. Meiotic endonuclease drives lower the inheritance of the competing chromosome within a pair by damaging it, such that gametes carrying the non-drive chromosome are eliminated during gametogenesis or, in some cases, produce non-viable offspring. This includes the disruption of specific essential genes in toxin-antidote meiotic drives[61;162;208], or through more structural damage, such as chromosome 'shredder' meiotic drives[145;146]. Natural sex-linked meiotic drive systems have been reported in *Aedes* and *Culex* mosquitoes[209;210]. Synthetic shredder endonuclease meiotic drives have generally sought to exploit large-scale, potentially repeating sequence differences between chromosome pairs to increase the damage done to the chromosome that does not carry the drive[145;146].

For most reports of synthetic homing drives, the method of quantifying inheritance bias (phenotypic scoring of progeny carrying a marker gene in the drive allele) cannot differentiate between the underlying inheritance bias mechanism. However, a small subset of reports of homing drives have had marked chromosomes[50;75;91;93;95;103], especially pre-CRISPR[78;85–87], which may allow homing and meiotic inheritance bias to be differentiated. Through the use of a coincidental chromosomal marker, we observed evidence for meiotic drive in male *A. aegypti* with a homing CRISPR gene drive design

reported by Li et al.[95].

Li et al. tested the inheritance biasing ability of a set of homing split drive systems comprising a guide RNA (gRNA) expressing element inserted into the *white* gene (w^{GDe}) and one of five secondary site transgene insertions expressing Cas9 under the control of various promoters from genes expressed in the mosquito germline. The *white* gene is tightly linked to the sex-determining region of *A. aegypti* which allows the sex of the progeny to function as a chromosomal marker (donor/recipient) in the progeny of male drive carriers. While three of the Cas9 regulatory regions resulted in drive activity in females, only *nup50* expressing Cas9 resulted in a statistically significant increased inheritance of the drive from male drive parents[95]. We re-analysed the results of Li et al. for *nup50* males taking into account the sex linkage and found that the observed inheritance bias in males seemingly proceeded exclusively through meiotic drive.

We set out to test the hypothesis that the meiotic drive observed with the *nup50* expression pattern is a more general phenomenon and also occurs with other *A. aegypti* gene drives that show activity in males. We repeated the w^{GDe} and *nup50*-Cas9 crosses with lines provided by the original authors, and performed crosses with Cas9 expression under the control of putative transcription regulatory regions of two additional *A. aegypti* germline genes. The first, suppressor of defective silencing 3 (*sds3*) has been shown, by dsRNA-induced knockdown in *Anopheles gambiae*, to be necessary for normal development of the ovarian follicles and testes, without other obvious defects[211]. The second, benign gonial cell neoplasm protein (*bgcn*) is involved in the regulation and promotion of gametogenesis in both sexes[212] and has been described in the context of gene drive in *Drosophila melanogaster* with the I-SceI nuclease[78].

For each line that expresses Cas9, we report the degree of inheritance bias of the w^{GDe} element for both sexes and, in males, the mechanism of inheritance bias. For *sds3*, *bgcn*, and *nup50*-Cas9, we find an increase in recombination events indicative of homing. Furthermore, by scoring somatic eye phenotypes, we also find strong evidence of zygotic/somatic expression, maternal deposition and an effect of the Cas9 carrying grandparent's sex on w^{GDe} inheriting grand-offspring phenotypes.

5.3 Results

5.3.1 Inheritance of wGDe is biased by bgcn, sds3 and nup50-Cas9

To assess the degree and, in males, the mechanism of inheritance bias, we bred transgenic *A. aegypti* mosquitoes to create and analyse a split drive arrangement. In this split drive, the w^{GDe} allele expresses a gRNA targeting the wildtype *white* gene (w^{+}) at the site corresponding to where the drive element has been inserted and disrupts its protein coding sequence (Fig 5.1a and Fig S5.1). The *white* gene is

located on chromosome one, near the dominant acting male determining allele M such that males are M/m and females m/m.

To generate individuals in which drive can occur, the w^{GDe} element is combined with the other component of the split drive, a separate transgene that expresses Cas9 under the control of regulatory sequences from an endogenous germline-specific gene, either *nup50*, *bgcn*, or *sds3*. Individuals carrying a single copy of both the w^{GDe} and Cas9 transgenes (double heterozygotes) were generated in two ways: by crossing parental F_0 female w^{GDe} homozygotes to male Cas9 individuals (Fig 5.1b Top) or with the reciprocal cross (Fig 5.1b Bottom). The double heterozygous offspring (F_1) were in turn crossed to the Liverpool wild type strain, and their progeny (F_2) were collected and their fluorescence and phenotype scored (Tables S5.1-S5.7). For each condition, Fisher's Exact tests were performed comparing the w^{GDe} inheritance rates to those in the absence of any Cas9 element for male (52%, 620/1203) or female (51%, 308/605) parents (Table S5.8). All Cas9 expressing lines were able to bias the inheritance of the w^{GDe} element in at least one cross (Table S5.9 and Fig 5.1c).

For *sds3*, F_1 drive females with maternal Cas9 propagated the w^{GDe} element to 67% (118/176, p-value: 0.050^{*}) of their progeny (Table S5.1) and for *bgcn*, F_1 drive males with maternal Cas9 the propagation rate was 66% (257/389, p-value: 0.010^{**}) (Table S5.2). For *nup50* (Table S5.3), all four crosses had significantly increased inheritance rates, and to a similar degree as reported to the identical crosses in Li et al.[95]. The *nup50* double heterozygous males passed along the w^{GDe} element to 64% (1159/1819, p-value: 0.001^{***}) of their progeny with paternal Cas9 and to 63% (1852/2926, p-value: $<0.001^{***}$) of their progeny with maternal Cas9. For *nup50* drive females the propagation rate was 69% (952/1377, p-value: $<0.001^{***}$) for paternal Cas9 and 70% (1055/1501, p-value: $<0.001^{***}$) for maternal Cas9.

For *nup50*-Cas9, the progeny were collected individually from F_1 parents (Table S5.4-S5.7). There was considerable variation between the inheritance rate from different parents carrying the same drive (Fig 5.1c), a notable feature that has been reported in other articles on homing drives[48;92;95;103;114]. Due to this overdispersion, we cannot reliably determine if there is a statistical difference in the inheritance rate between the different Cas9 regulatory elements. However, because this overdispersion is expected only to occur if the drive is functional, our method for determining a difference from the control remains valid, albeit with a potentially inflated false negative rate.

5.3.2 Eye phenotype reveals the source of nuclease activity

All progeny were evaluated for eye pigment defects that may result from embryonic or later somatic biallelic disruption of the *white* gene by the w^{GDe} element and NHEJ mutations. Since the double heterozygote drive-carrying parents were crossed to wildtype individuals, each progeny inherited at least one dominant functional *white* allele from the non-drive parent, and, if the w^{GDe} element is not inherited, potentially an additional one from the drive parent. The biallelic loss of function of the *white* gene must therefore occur through deposition into, or somatic expression by, F_2 individuals.

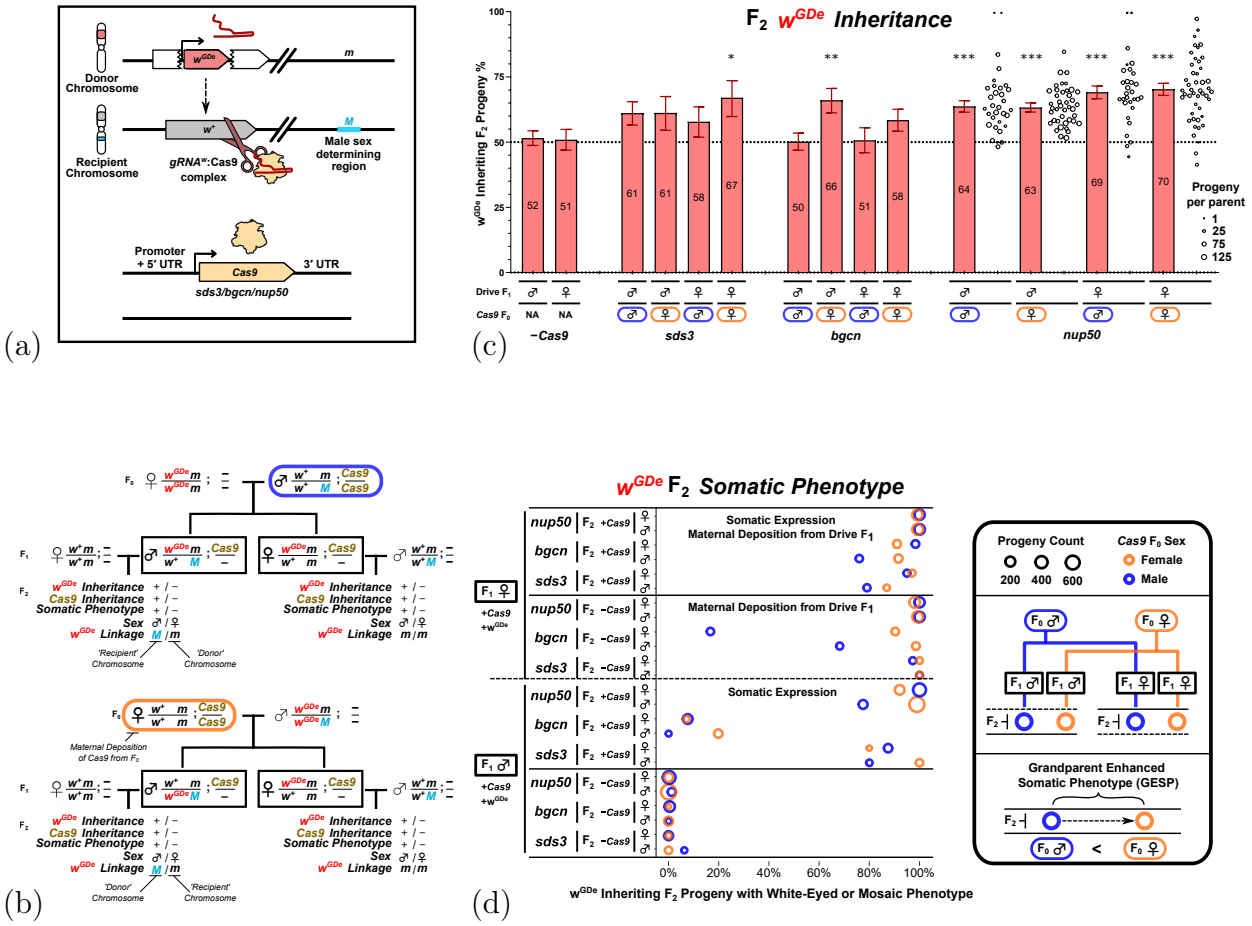


Figure 5.1: **Gene drive element (w^{GDe}) inheritance and somatic eye phenotype in the progeny of double heterozygote split drive carriers.** Illustration of gRNA:Cas9 split drive system. The gene drive element $^{\mathrm{GDe}}$ is inserted into, and disrupts, the *white* gene which is tightly linked to the sex-determining region (M or m). **b**. Breeding schemes for the four crosses per Cas9 expression variant. The solid boxes indicate the F_1 genotypes that may bias the inheritance of w^{GDe} in their germline. The upper family tree shows the double heterozygous F_1 with paternally contributed Cas9 and maternally contributed w^{GDe}, m-linked in both F_1 males and females. The bottom family tree shows the double heterozygous F_1 with paternally contributed w^{GDe}, M-linked in male and m-linked in female F_1s, and maternally contributed Cas9. **c** F_1 drive parent germline inheritance bias of w^{GDe} when combined with a *sds3*, *bgcn*, or *nup50*-Cas9 expressing element. The horizontal dotted line indicates the expected Mendelian 50% inheritance. For *nup50*, individual crosses were performed, and each circle represents the percentage of w^{GDe} positive progeny from an individual parent. Data are presented as mean values with the Wilson confidence intervals for the binomial proportion calculated for the pooled progeny count, which does not take into account the potential lack of independence due to parent-by-parent batch effects. Stars indicate the p-value thresholds from two-sided Fisher's exact tests to the matched drive sex F_1 −Cas9 condition. The p-values and number of progeny scored are presented in Table S5.9. Source data are provided as a Source Data file. **d** The percentage of w^{GDe} inheriting F_2 progeny, initially of the w^{+}/w^{GDe} genotype, that display a mosaic or total loss of eye pigment phenotype due to disruption of their w^{+} allele. The circle size indicates the number of progeny, and circle colour indicates if the Cas9 carrying F_0 grandparent was male (Blue) or female (Orange). Progeny from F_1 drive females are indicated with 'Maternal Deposition'. Progeny that inherited both a w^{GDe} allele and Cas9 element are indicated with 'Somatic Expression'. White phenotype rates for the F_2 progeny that did not inherit w^{GDe} are shown in Fig S5.3.

Consistent with this, the progeny of the −Cas9 control crosses did not present with a *white* phenotype (Table S5.8).

For male double heterozygote *sds3*-Cas9 crosses, of the F_2 progeny (♂ and ♀ pooled) that inherited both the w^{GDe} and the Cas9 element, 86% (111/129) presented with a mutant somatic phenotype if the Cas9 carrying F_0 was male, or 98% (61/62) if the Cas9 carrying F_0 grandparent was female (F_1:♂, +Cas9 in Fig 5.1d and Table S5.10). For *bgcn*-Cas9 this was 7% (14/196) or 17% (22/129), and for *nup50*-Cas9 this was 95% (586/615) or 98% (924/946). However, if only the w^{GDe} element was inherited, no cross had more than 1% of the pooled ♂ and ♀ F_2 progeny present with a somatic phenotype, presumably resulting from the lack of paternal Cas9 deposition through the sperm (F_1:♂, −Cas9 in Fig 5.1d and Table S5.10). For each cross, this was a significant difference (Table S5.10) indicating somatic expression, without substantial paternal deposition of Cas9/Cas9:gRNA^w. In contrast to the <1% rate observed in the progeny of F_1 drive males, the crosses with female double heterozygotes where only the w^{GDe} element was inherited, 40% (39/98) of progeny presented with a somatic phenotype if the Cas9 carrying F_0 was male, while 95% (124/131) if the Cas9 carrying F_0 grandparent was female. An astounding 99% (75/76) or 100% (61/61) of the *sds3* and 100% (462/462) or 99% (528/535) of the *nup50* progeny presented with somatic phenotypes (F_1:♀, −Cas9 in Fig 5.1d and Table S5.11). This indicates strong maternal deposition of Cas9/Cas9:gRNA^w. For each cross, this was a significant difference (Table S5.11). Maternal Cas9 deposition without substantial paternal deposition has been reported for many other drive systems[48–50;53;82;90;92;94;95;114;115;118;129;161].

5.3.3 Grandparent Enhanced Somatic Phenotype

Surprisingly, in the w^{GDe} inheriting progeny, we observed a trend where a higher fraction of the progeny exhibited a somatic phenotype when the Cas9-carrying grandparent was female as opposed to male (F_0:♂vs F_0:♀in Fig 5.1d). Contrasting each male F_0 Cas9 carrying grandparent cross with the equivalent cross with a female F_0 Cas9 (each row in Fig 5.1d) showed, for female F_0 Cas9, an average 5.2% (sd:14.4%) percentage point increase in white/mosaic eyed phenotype among $+w^{GDe}$ F_2 progeny. While maternal deposition from a Cas9 carrying grandparent may increase the number of w^{GDe} and NHEJ mutated alleles passed along by the F_1 parental generation to $-w^{GDe}$ progeny (S5.3), this should not, in contrast to what we observe (Fig 5.1d), influence the phenotype of the progeny that inherit the w^{GDe} element. If the w^{GDe} element is inherited there is no opportunity to inherit a germline NHEJ mutation that was created due to deposition from the grandparent into the parent. We created a generalised linear model that included Cas9 promoter, F_2 Cas9 status, F_2 sex, F_1 drive parent sex, and F_0 Cas9 carrying grandparent sex (Table S5.12). The sex of F_0 Cas9 carrying parent had a significant influence on the fraction of white/mosaic eyed $+w^{GDe}$ F_2 progeny. We termed this phenomenon Grandparent Enhanced Somatic Phenotype (GESP). All other factors were also significant, apart from the sex of the F_2 progeny.

5.3.4 Sex of the F_2 progeny reveals the mechanism of inheritance bias

In *A. aegypti*, the *white* gene is tightly linked to the sex-determining locus. This locus comprises two forms, a dominant male determining allele M and a corresponding m allele, such that males are M/m and females m/m. While the molecular basis of sex determination in this mosquito is not fully understood, M is associated with *Nix*, a gene shown to be involved in sex determination[213]. Analogous to an XY chromosome system, male offspring of an M/m male always carry the paternal M allele and female offspring the paternal m, with no such distinction between the two m alleles of the mother. For the male parent, if the initial linkage of w^{GDe} to m or M is known (determined by the sex of the w^{GDe} carrying grandparent), the sex of the progeny can be used as an indication of whether an observed inheritance bias is due to new recombination events (homing), or increased inheritance of the original drive-carrying chromosome (meiotic drive) (Fig 5.2a). To this end, we stratified the w^{GDe} inheritance by the sex of the F_2 progeny for each of the double heterozygous male parents (Fig. 5.2b).

The background recombination rate of w^{GDe} and sex in the absence of any Cas9 element was 1.08% (13/1203) (Table S5.8) and was compared by Fisher's Exact tests to the recombination rate from w^{GDe} Cas9 male double heterozygotes (Table S5.13). As reported above, only one cross each of the *sds3* and *bgcn* double heterozygotes showed a significant increase in overall w^{GDe} inheritance. However, quantifying conversion with marked chromosomes is much more sensitive than measuring overall w^{GDe} inheritance rate.

For the *sds3* double heterozygous males with paternal Cas9 contribution (and therefore in our crosses a m linked w^{GDe} element) 48% of progeny (216/450) inherited the recipient chromosome as determined by their sex (♂) and 9% of progeny (41/450) were w^{GDe} positive males. This allows us to estimate the fraction of recipient chromosomes that were converted by the combined effect of homing and background recombination: 41/216 = 19% (p-value: $<0.001^{***}$). The same was true for maternally contributed Cas9 where 9% of progeny were w^{GDe} females, indicating a homing rate of 19% (19/102 p-value: $<0.001^{***}$). For *bgcn* males with paternal or maternal Cas9, we found homing rates of 5% (24/464 p-value: $<0.001^{***}$) and 20% (32/160 p-value: $<0.001^{***}$), respectively. This large difference in the rate of homing between crosses with maternal vs. paternal F_0 Cas9 suggests that for *bgcn* maternally deposited Cas9 may contribute more to homing than autonomously expressed Cas9. Low expression with high maternal deposition by *bgcn*-Cas9 is also consistent with the observed phenotype rates of *white* (Fig 5.1d).

For male *nup50* double heterozygote males with paternal Cas9 contribution, 24% (210/869 p-value: $<0.001^{***}$) of the recipient chromosomes were converted by homing. For maternally contributed Cas9 this was 23% (315/1387 p-value: $<0.001^{***}$) of recipient chromosomes. We also performed this analysis on the *nup50* crosses reported in Li et al. (Table S5.14). Despite a significant increase in inheritance of the w^{GDe}

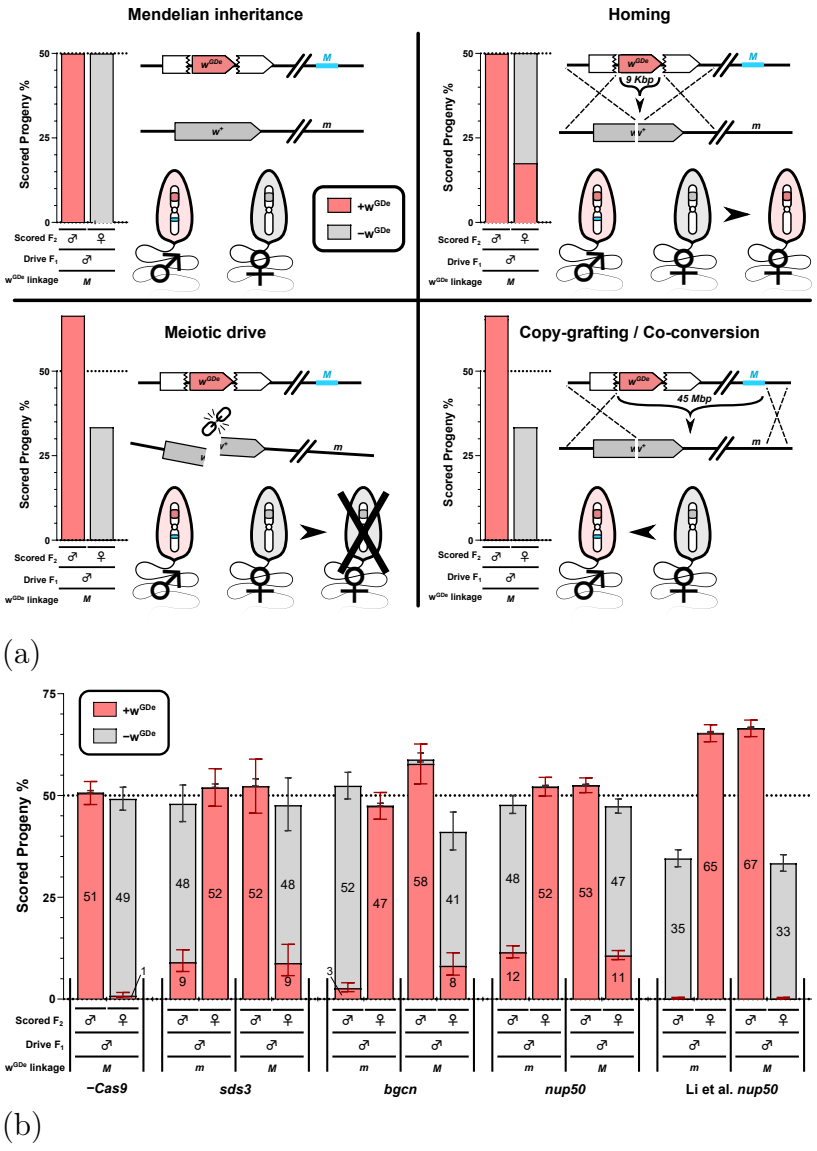


Figure 5.2: **Separating w^{GDe} inheritance by $\mathbf{F_2}$ sex allows different mechanisms of inheritance bias to be distinguished.** Illustration of how homing, meiotic drive and copy-grafting/co-conversion are expected to influence the observed sex-linkage of an M (shown in Blue) linked w^{GDe} element in the progeny of male drive double heterozygous parents. The expected sex-linkage is exactly opposite for an m linked w^{GDe} element. **b** Parental germline inheritance bias of w^{GDe} when combined with no Cas9, *nup50*, *bgcn*, or an *sds3*-Cas9 expressing element. We included the *nup50* results from Li et al. that use the identical w^{GDe} and *nup50*-Cas9 line. For each of the three Cas9 regulatory elements, the w^{GDe} inheritance from male double heterozygotes is reported in pairs of columns segregated by the sex of the F_2 progeny. In each case, the first pair of columns are the results for when w^{GDe} is m-linked (m), and the second pair are the results for when w^{GDe} is M-linked (M). Data are presented as mean values with the Wilson confidence intervals for the binomial proportion calculated for the pooled progeny count. The overlaid numbers are the percentage (cumulative within each column) of the indicated F_2 sex and w^{GDe} status among all progeny from that cross. The number of progeny scored are presented in Table S5.3. Source data are provided as a Source Data file.

element, there was no evidence of an increased recombination rate: 0% (3/690 p-value: 1.0^{ns}) for paternal Cas9 and 0% (3/688 p-value: 1.0^{ns}) for maternal Cas9 contribution. Instead, there was a significant bias in favour of the w^{GDe} linked sex corresponding to the donor chromosome. For paternally contributed Cas9, 65% (1306/1996 p-value: $<0.001^{***}$) of progeny were female, >99% of which were w^{GDe} positive. For maternally contributed Cas9 67% (1371/2059 p-value: $<0.001^{***}$) of progeny were male, >99% of which were w^{GDe} positive (Table S5.15). This sex bias should not occur through homing, instead, this is consistent with a meiotic drive mechanism where some of the non-w^{GDe} chromosomes are lost, or conversion of a very large region encompassing both w^{GDe} and the sex-determining region (Fig 5.2a). For the crosses performed for this study, including the *nup50* line, no significant difference in sex, and by extension recipient vs donor chromosome inheritance, was detected (Table S5.15). For *bgcn* with maternal F_0 Cas9, 59% of all F_2s inherited the donor chromosome (male), but this did not reach our significance threshold due to the relatively low number of progeny scored for this cross.

5.4 Discussion

In this study, we report the efficiency and mechanisms of three CRISPR-Cas9 nuclease gene drives targeting the *white* gene, expanding the set of tools to develop genetic control strategies for the public-health relevant *A. aegypti* mosquito. In our study, *sds3*, *bgcn*, and *nup50* expressed Cas9 each resulted in increased inheritance of the w^{GDe} drive element, with the primary mechanism being homing. Additionally, for each promoter, we find evidence of maternal deposition and somatic expression and, unexpectedly, an effect of the Cas9 carrying grandparent's sex on the grand-offspring phenotypes that we termed Grandparent Enhanced Somatic Phenotype (GESP). In line with Li et al., we find the *white* locus to be a good drive target, allowing for efficient transmission bias and convenient readout of an easily-scored visible recessive phenotype[95]. In addition, the insertion site allows for effective transgene expression from a sex-linked locus, which may be of particular use for future drives and other genetic control approaches. For the *bgcn* drive in males, the recipient chromosome conversion rate was much higher with maternally contributed Cas9 (19%) compared to paternally contributed Cas9 (5%). These results suggest that, in at least males, the *bgcn* drive may substantially function through maternally contributed Cas9. Homing through Cas9 deposition in the absence of expressed Cas9 ('shadow drive') has been reported for other drives[103;115;118], but to our knowledge, not as the primary means of inheritance bias for a drive. We find *nup50* and *sds3*-Cas9 capable of directing transmission bias in females and males, and we did not find that maternal deposition from the Cas9-carrying grandmother negatively influenced the homing rate observed in males. It is important to note that in our crosses, only Cas9 could be maternally deposited into the F_1 double heterozygotes, maternal deposition of Cas9 protein and the gRNA simultaneously may be much less conducive to shadow drive[1].

For all drives, the almost complete absence of any somatic phenotype in individuals that did not inherit the w^{GDe} element (Fig S5.3) could indicate that, while maternal

deposition of the Cas9 occurs, the gRNA^{w} or gRNA^{w}:Cas9 complex are either not deposited or are rapidly degraded. However, progeny that did not inherit the w^{GDe} element instead inherited the (initially) w^{+} allele from the double heterozygous parent. For *white* eye phenotypes to occur in these individuals, up to two functional w^{+} alleles may need to be disrupted by deposition instead of one; direct comparison of the rates of somatic mutation between offspring that do and do not inherit the $gRNA^{w}$ transgene are therefore potentially misleading. Furthermore, some non-w^{GDe} progeny may have inherited a *white* allele that contained a functional, but cut resistant, NHEJ mutation (type-1 resistant mutation) which would make biallelic disruption impossible.

For the $-w^{GDe}$ F_2 progeny, maternal deposition from the F_0 grandmother could increase their probability of inheriting a mutated w allele. As such, GESP does not apply and only refers to $+w^{GDe}$ F_2 progeny where the sex of the w^{GDe} or Cas9 carrying grandparent seemingly influences their propensity to present with a somatic phenotype. Although deposition from an F_0 grandparent may explain a change in the quantity of w^{GDe} alleles passed along by the F_1 drive parent, it does not appear to explain a change in the phenotype of those F_2 progeny that inherited a drive element. One possible explanation for GESP may be an increased maternal deposition rate of Cas9:gRNA complexes from increased gRNA expression in w^{GDe} homozygous germline cells compared to w^{GDe} heterozygous germline cells. Consistent with this, for *bgcn*-Cas9 the w^{GDe} homing rate was higher when the Cas9 carrying F_0 grandparent was female. A similar analysis of a single drive element (containing both Cas9 and a gRNA) found that maternal deposition rates were lower when drive conversion in the maternal germline was less[90]. However, in our split drive system, only the gRNA-expressing element is biased, the Cas9-expressing element remains heterozygous regardless if homing has occurred or not. It may be that different mechanisms, such as genomic imprinting or transgenerational persistence of deposited Cas9 mRNA/protein, contribute to GESP.

For *nup50* the overall inheritance biasing rate and somatic/embryonic drive activity closely match those reported by Li et al.[95] and underscore its potential utility for systems such as precision-guided SIT[133]. However, an important finding of our work is the propensity of this drive to function through two different mechanisms. The selective inheritance or elimination of a chromosome is generally achieved by creating multiple DNA breaks on the target chromosome[146;214–216] (e.g. X-shredder) or by disrupting an essential gene[61;162]. Meiotic drive through a single cut in a non-essential gene as found by Li et al. and reported here is noteworthy. An explanation could be the chromosomal location of the induced double-stranded break. A single cut has been shown to be sufficient for inheritance bias through the loss of a chromosome in yeast when it is targeted to a centromere, while nearby sites were not sufficient[217]. Chromosome loss has also been found to be a frequent outcome of allele-specific editing of a pericentromeric site in human embryos[176]. The *white* gene is located relatively near the centromere. However, a centromere effect does not explain the difference in results from this study and that of Li et al., which instead suggests subtle differences in the rearing conditions or background genetics of the mosquito strains may have a significant influence on the underlying mechanisms. Gene drive

assessment performed in *D. melanogaster* with different genetic backgrounds has revealed differences in drive activity but changes in the underlying mechanism were not investigated[90]. The *nup50*-Cas9 and w^{GDe} transgenic lines used in this study are derived from those described in Li et al., but the crosses to assess homing were made to Liverpool (LVP) strains maintained for a long period of time in different insectaries. Mosquito colonies maintained in laboratories can suffer from founder and drift effects, affecting their genetic background and reducing their heterozygosity[218]. Moreover, genetic variability in *A. aegypti* colonies of the same strain but reared in different laboratories has been documented[219]. There may also be methodological factors that could allow the same biological processes to manifest differently (e.g., different screening timings with genotype-dependent mortality rates).

A limitation of our study is that we cannot rule out that the sex bias we report for Li et al. *nup50*-Cas9 is due to copying of the estimated 45Mbp[220;221] region comprising both the w^{GDe} and the sex-determining region (Fig 5.2a). However, the large distance between the w^{GDe} drive and the M/m locus leads us to believe that this is unlikely, as co-conversion in similar contexts is generally reported to be on the scale of 100s of base pairs[4;85;103;106;222]. Furthermore, a substantial fraction of conversion tracts have been reported to be unidirectional in *A. aegypti*[4]. This suggests that even if large-scale co-conversion was favoured, some repair events should still have caused recombination between w^{GDe} and the sex-determining locus if co-conversion occurred primarily in the other direction relative to the sex-determination locus. Finally, several studies have reported partial homing events[47;48;81;82;90;99]. These partial homing events are seemingly due to sequences in the drive element (such as the gRNA gene) having undesired homology to the recipient chromosome (shown for w^{GDe} in Fig S5.1) and result in only part of the drive element being copied over. These reports of partial homing are inconsistent with a single DNA break-inducing large-scale homing beyond the (immediately) adjacent regions of homology.

There are additional phenomena that can lead to biased inheritance with a sex-linked transgene. In particular, alleles with sex-specific lethal effects may be clustered within the neighbourhood of the sex-determining region in *A. aegypti* and can become linked to a transgene[223]. However, the meiotic drive we report shows a reciprocal sex bias depending on the linkage of the w^{GDe} element with the M or m locus and the use of a split drive system demonstrates that the effects depend on Cas9 activity and are not simply due to the w^{GDe} insertion or a linked allele. A more comprehensive analysis of (even more distal) sequence differences between donor and recipient chromosomes after DNA repair may further inform the exact mechanism of inheritance bias. However interpretation of such data must be done with caution, donor chromosome sequences (including the drive element) may incorrectly appear homozygous when NHEJ mutations cause the binding sites of a PCR primer to be blocked on the recipient chromosome. This issue has been raised in several analyses[159;175;176;224] and highlights potential pitfalls for identifying homing events with these types of molecular assays. We highlight these cases specifically because we believe such genetic assays are worth perusing, but should be informed by this prior work to reduce the chance of misinterpretation.

To our knowledge, for gene drives designed to function through homing, recipient/donor chromosome markers have been used with non-CRISPR nucleases in *D. melanogaster*[78;86;87] and *A. gambiae*[85] and with CRISPR-Cas9 in *D. melanogaster*[91;103], *A. aegypti*[95] and *Mus musculus*[75;93]. There may be additional cases in which a split drive element can coincidentally act as a chromosome marker[50;116]. In *D. melanogaster*, some studies have noted a reduced inheritance of the recipient chromosome, however, these may be attributable to genotype specific fitness effects instead of DNA damage-induced loss of the recipient chromosome[50;103]. Xu et al. have performed the most extensive investigation of homing drives with marked chromosomes and found a mix of homing and bias through chromosome damage[91].

In light of our results, re-evaluation of the *A. gambiae* I-SceI gene drive reported by Windbichler et al. may suggest that a meiotic drive effect in homing drive designs is more widespread in mosquitoes[85]. Their drive-carrying line had a small marker (*Not*I restriction site) located approximately 0.7 kilobases from the I-SceI cut-site on the recipient chromosome, but not on the donor drive chromosome. They reported 86% inheritance of the drive element from heterozygote males. However, drive alleles that included the *Not*I site only accounted for around half the increased drive allele inheritance. The authors attributed this discrepancy to co-conversion, where homing of the drive element also replaced the nearby *Not*I marker. A combined meiotic drive and homing effect would seem to provide an alternative explanation. In the *M. musculus* drive reported by Grunwald et al. the recipient chromosome had a linked coat colour marker that allowed the homing events to be precisely tracked[93]. In females, *vasa*-Cre induced CAG-Cas9 expression resulted in homing rates of 42% (36/86) and 11% (5/47) depending on the Cas9 insertion site. In males, no homing was observed with any drive. However, for the *vasa* drives, males passed along the donor drive chromosome to 63% (45/71) and 54% (49/91) of their progeny, potentially indicating a meiotic drive mechanism in that sex. It should be noted that detecting meiotic drive using this method is less sensitive than detecting homing, and more progeny would need to be scored to have confidence in this trend. Together, these results suggest that a meiotic mechanism in drives intended to function through homing may be more common than currently realised. Distinguishing these mechanisms requires linked markers; for some organisms, this type of in-depth investigation may best be reserved for drives that after initial tests warrant further development.

Our work further expands the Cas9 expression patterns that have been tested in the context of mosquito gene drives. It is notable that the drives with a homing design reported in *Anopheles* mosquitoes *A. gambiae*[47;51;92;225] and *A. stephensi*[53;82;161] almost invariably have a dramatically higher conversion rate than those found in *A. aegypti*[95;96]. It is not clear what underlies this difference. However, the fact that the modest conversion rate for *nup50*-Cas9 males remains stable despite a change in the mechanism may limit possible explanations. This stability suggests that the factors that negatively affect the conversion rate in *A. aegypti* are not specific to either homing or meiotic drive. Moreover, it also indicates that the difference in conversion rate observed between mosquito species is probably not due to the species favouring one mechanism over the other. Yet, the difference in mechanism between homing and

meiotic drive through gamete destruction has important practical implications.

First, the loss of gametes through a meiotic-drive mechanism may negatively affect mating competitiveness by lowering the number of viable gametes, though in some cases gametes may be produced in sufficient excess for this not to be significant. The homing mechanism functions through conversion and should not affect gamete numbers. For the *nup50* meiotic drive reported by Li et al., male *nup50*-Cas9 fecundity was tested and found to not differ from wildtype[95].

Second, on a 'per cut' basis, meiotic drive is moderately less efficient than homing. When meiotic drive removes a non-drive gamete/embryo, it thereby benefits the remaining gametes/embryos. These may, in addition, to drive-carrying gametes, include other wildtype and cut-resistant alleles carried by gametes that were not destroyed. In contrast, homing converts a non-drive gamete to a drive gamete, which does not benefit any of the leftover non-drive gametes making homing more efficient.

Third, the linkage between different drive components may vary significantly depending on the mechanisms: for instance, if in a split drive system the Cas9 is located near the gRNA element homing would still only increase the number of gRNA alleles, but not the Cas9 alleles. However, meiotic drive would increase the inheritance of both the gRNA and Cas9 element. This could theoretically cause a split drive or daisy-chain drive[59] to spread more than anticipated. Locating each element on separate chromosomes would prevent this, and our data suggest that this may be a wise precaution to increase the predictability of their invasiveness. Although, if anticipated or identified in early-stage field trials, a meiotic drive-induced linkage between elements could also be leveraged, lowering the required release frequencies[169]. Nonetheless, in regards to risk-assessment of rare recombination events, the genomic distance at which two split drive elements become strongly linked is presumably still much more permissive for a meiotic drive mechanism as opposed to a homing mechanism.

Last, in the case of Li et al.'s *white* targeting *A. aegypti* drive, its linkage to the sex-determining locus caused an otherwise neutral replacement drive to act, in males, like a sex-biasing suppression drive. This might be desirable for some applications, but surely detrimental if the intended application were different. Most of these concerns apply even if the actual mechanism is co-conversion/copy-grafting of a large chromosome segment as opposed to meiotic drive.

The work presented in this chapter shows both the utility of the meta-analysis project and its limitations. It allowed us to identify an interesting part of gene drive biology for further investigation. However, we specifically fail to replicate that fundamental part of gene drive biology. None of the factors we currently record in the meta-analysis goes towards explaining this difference, and it suggests that we should be more humble in what conclusions we draw. There may be interesting and statistically significant relationships that are true for the data, but we need to question whether that remains true for the biology and beyond the simplified conditions of genetic crosses. There is likely a significant degree of noise that we will not be able to reduce by simply

recording even more design factors about the gene drives we are interested in studying. Expanding the space of known unknowns may be the most important contribution of the work performed in the previous chapter.

5.5 Methods

DNA constructs

The sequence and insertion site of the 3xP3-tdTomato carrying gRNA element (Fig S5.1) and *nup50* lines are described in Li et al.[95] and the *bgcn*-Cas9[172] and *sds3*-Cas9 constructs were produced by making several alterations to those original plasmids, provided by Omar Akbari[164]. These plasmids contain, within piggyBac terminal sequences, Cas9 expressed by *nup50* followed by a T2A self-cleaving peptide and EGFP and an OpIE2-DsRED cassette. To improve the visibility of the fluorescent marker, this was replaced with PUb-mCherry-SV40 for *bgcn* and *sds3*. To reproduce the germline-specific expression patterns predicted for these genes, the Cas9:EGFP coding sequence is preceded and followed by the non-coding sequences flanking the endogenous *bgcn* or *sds3* gene's open reading frame, followed by an additional P10 3^{prime}UTR. The *bgcn* and *sds3* constructs use a Cas9 that is insect codon optimised[172]. The *nup50* line makes use of a human codon optimised Cas9[164].

Mosquito lines

No ethical approval was required for working with invertebrate species. *A. aegypti* Liverpool strain (WT) was a gift from Jarek Krzywinski. The *nup50*-Cas9 and white gRNA expressing element w^{GDe} ($\mathrm{w}^{\mathrm{U6b\text{-}GDe}}$) lines were provided by Omar Akbari[95]. The *sds3*-Cas9 line was generated by standard embryo microinjection with a hyper-active *piggy*Bac transposase helper[172]. At Pirbright, the *nup50*-Cas9 line was maintained as a mix of homozygotes and heterozygotes with periodic selective elimination of wildtypes; the w^{GDe} element line was provided as homozygous and maintained in our facilities by screening for the white-eye phenotype (homozygous knockout of *white*) and the fluorescent marker. Cas9 expressing lines generated at the Pirbright facilities were maintained as heterozygotes, usually by crossing transgenic males to WT females and selecting for the fluorescent marker.

All mosquito lines were reared in an insectary facility under constant conditions of 28, 65-75% relative humidity and 12:12 light/dark cycle (1h dawn/1h dusk). Larvae were fed ground TetraMin flake fish food (TetraMin) while adults were provided with 10% sucrose solution *ad libitum*. Defibrinated horse blood (HB034, TCS Bioscience) was provided using a Hemotek membrane feeding system (6W1 system, Hemotek Ltd) covered with Parafilm (HS234526A, Bemis).

Crosses for homing assessment

Male and female adults, homozygous for w^{GDe} were crossed with mosquitoes of the Cas9 lines. Their progeny were screened as late larvae under fluorescence using a Leica MZ165FC microscope. The eye phenotype was also evaluated. Double heterozygous mosquitoes carrying both transgenes were then crossed to WT mosquitoes. Inheritance of the transgenes as well as eye phenotype, was again assessed under a fluorescence microscope. For *nup50*-Cas9, double heterozygotes were individually crossed. For *bgcn*-Cas9 and *sds3*-Cas9 multiple double heterozygotes were crossed simultaneously to WT of the opposite sex. The exact number and phenotype of the progeny of each cross are shown in Tables S5.2-S5.3. The individual cross-data for *nup50*-Cas9 are shown in Table S5.4-S5.7. In some cases, F_1 double heterozygotes produced from the same cross presented with a different fluorescent marker or eye pigment phenotypes. In each case, these were noted in the cross tables, and examples of the phenotypes are shown in Fig S5.2.

Statistical methods

Statistical analysis of w^{GDe} inheritance bias

For each F_1 sex, the w^{GDe} inheritance rate in the absence of a Cas9 expressing element (Table S5.8) was used as the baseline inheritance. This was 52% (620/1203) for males and 51% (308/605) for females. These rates were used as the expected outcome in a two-sided Fisher's exact test with the w^{GDe} inheritance from F_1 parents that carried the w^{GDe} and one of the Cas9 expressing elements. A significant difference in w^{GDe} inheritance is taken as evidence for drive activity. See Table S5.9.

Statistical analysis of somatic expression and parental deposition

For each Cas9 line, the fraction of mosaic eyed (ME) or white-eyed (WE) progeny among the F_2 offspring inheriting w^{GDe} but not the Cas9 ($+w^{\mathrm{GDe}}$;$-$Cas9) from F_1 drive males served as a control for the frequency of such phenotypes in the absence of somatic expression or maternal deposition. For somatic expression, the ME/WE fraction of the F_2 progeny harbouring both the Cas9 and w^{GDe} elements from F_1 drive males was compared to the control cross using a two-sided Fisher's exact test (Table S5.10). For maternal deposition, the F_2 progeny harbouring only the w^{GDe} element from F_1 drive females was compared to the control (Table S5.11).

Statistical analysis of the influence of factors on the fraction of mosaic and white-eyed progeny

A generalised linear model with binomial errors was created that included Cas9 promoter (*sds3*, *bgcn*, *nup50*), F_2 Cas9 status ($+/-$), F_2 sex (♂/♀), F_1 drive parent sex (♂/♀), and F_0 Cas9 carrying grandparent sex (♂/♀). The response variable was the proportion of ME and WE progeny among the all the F_2 progeny from that cross

and F_2 sex (48 conditions). The analysis was performed in R version 4.0.2 using the glm function. See Table S5.12.

Statistical analysis of homing and meiotic drive

For homing, the background recombination rate (calculated from the F_1 $+w^{GDe}$;−Cas9 male cross Table S5.8) is used as the expected outcome in a two-sided Fisher's exact test. For the control cross (in the absence of possible Cas9 mediated inheritance bias) the w^{GDe} allele was provided by the male F_0 grandparent and therefore M-linked in the F_1 males. In the absence of recombination, all F_2 males should be w^{GDe} positive, and all F_2 females should be w^{GDe} negative. Out of the 1203 progeny scored, we saw 13 (1.08%) recombination events. 2 out of 609 F_2 males were w^{GDe} negative, and 11 out of 581 F_2 females were w^{GDe} positive. For the crosses including a Cas9 element, a statistically significant increase in recombination rate between the recipient/donor chromosome marker and the drive element was taken as evidence of homing (Table S5.13). For meiotic drive, a statistically significant difference in the inheritance of either the recipient or donor chromosome (i.e., F_2 sex) is taken as evidence for meiotic drive (Table S5.15). The progeny sex ratio is compared to the sex-ratio in the absence of a Cas9 expressing element (Table S5.8).

Data availability statement

All the datasets generated during the current study are included in the supplementary information/source data file. The Li et al. data used in this study are available in the supplemental files of the original article https://doi.org/10.7554/eLife.51701.

Code availability statement

R version 4.0.2 was used for data analysis. The script for analysis and additional files are available at: `https://osf.io/q4cj5/`

Acknowledgements

Biotechnology and Biological Sciences Research Council (BBSRC) supported S.A.N.V. (BB/M011224/1), L.A. (BBS/E/I/00007033, BBS/E/I/00007038, and BBS/E/I/00007039) and M.B.B. (BB/H01814X/1, BB/L00948X/1, BB/V008110/1). M.A.E.A., E.G., J.A. and L.A. were supported through an award from Defense Advanced Research Project Agency (DARPA) Safe Genes program (to MIT N66001-17-2-4054), as was M.B.B (HR00111-16-2-0005). O.S.A, N.P.K., and M.L were supported in part by funding from the DARPA Safe Genes Program Grant (HR0011-17-2-0047) and an NIH awards (R01AI151004 and RO1AI148300). R01AI151004) awarded to O.S.A.. The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the U.S. Government.

Author contributions statement

M.A.E.A. and L.A. conceived of the experiments, E.G. conducted the crosses and data collection, S.A.N.V. and M.B.B analysed the results and S.A.N.V. wrote the manuscript. S.A.N.V. performed the initial analysis that noted the unexpected inheritance mechanism, and J.A contributed. O.S.A, M.L, and N.P.K. provided transgenic lines. All authors reviewed and provided comments on the manuscript.

Competing interests

O.S.A. is a founder of Agragene, Inc., and has an equity interest. The terms of this arrangement have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. All other authors declare no competing interests.

5.6 Supplemental data

sds3 Cross	F_1s crossed (♂x♀)	Drive F_1 phenotype	$+w^{GDe}$;+Cas9				$+w^{GDe}$;−				−/+Cas9				−/−			
			♀		♂		♀		♂		♀		♂		♀		♂	
			ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT
F_1: ♂w^{GDe};*sds3*-Cas9 x ♀LVP F_0: ♂Cas9 x ♀w^{GDe}	10 x 50	Mosaic eyes	91	13	20	5	0	130	1	15	0	0	0	76	0	0	0	99
F_1: ♂LVP x ♀w^{GDe};*sds3*-Cas9 F_0: ♂Cas9 x ♀w^{GDe}	10 x 25	Mosaic eyes	38	2	34	9	36	1	39	0	0	22	0	23	0	33	0	38
F_1: ♂w^{GDe};*sds3*-Cas9 x ♀LVP F_0: ♂w^{GDe} x ♀Cas9	10 x 50	Mosaic eyes	4	1	57	0	0	14	0	55	1	40	0	0	0	42	0	0
F_1: ♂LVP x ♀w^{GDe};*sds3*-Cas9 F_0: ♂w^{GDe} x ♀Cas9	10 x 21	Mosaic eyes	33	1	20	3	28	0	33	0	0	12	0	16	0	12	1	17

Table S5.1: F_2 progeny of *sds3*-Cas9 and w^{GDe} double heterozygotes crossed to LVP wildtype. Mosaic eyes (ME), white eyes (WE). Drive F_1 phenotypes are shown in Fig S5.2.

bgcn Cross	F_1s crossed (♂x♀)	Drive F_1 phenotype	$+w^{GDe}$;+Cas9				$+w^{GDe}$;−				−/+Cas9				−/−			
			♀		♂		♀		♂		♀		♂		♀		♂	
			ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT
F_1: ♂w^{GDe};*bgcn*-Cas9 x ♀LVP F_0: ♂Cas9 x ♀w^{GDe}	10 x 50	Dark eyes	11	55	0	1	1	114	0	2	0	0	0	58	0	1	0	101
	10 x 51	Mosaic eyes	3	115	0	11	0	121	0	10	0	0	0	128	0	0	0	153
		Total	14	170	0	12	1	235	0	12	0	0	0	186	0	1	0	254
F_1: ♂LVP x ♀w^{GDe};*bgcn*-Cas9 F_0: ♂Cas9 x ♀w^{GDe}	10 x 25	Dark eyes	36	1	32	0	2	37	27	0	0	28	0	41	0	32	0	42
	3 x 7	Mosaic eyes	25	0	6	12	7	8	3	14	0	14	0	12	0	20	0	15
		Total	61	1	38	12	9	45	30	14	0	42	0	53	0	52	0	57
F_1: ♂w^{GDe};*bgcn*-Cas9 x ♀LVP F_0: ♂w^{GDe} x ♀Cas9	5 x 25	Dark eyes	0	8	6	51	0	3	0	60	0	35	0	2	0	40	0	1
	10 x 50	Mosaic eyes	2	18	14	30	0	1	0	64	0	18	1	0	0	35	0	0
		Total	2	26	20	81	0	4	0	124	0	53	1	2	0	75	0	1
F_1: ♂LVP x ♀w^{GDe};*bgcn*-Cas9 F_0: ♂w^{GDe} x ♀Cas9	3 x 7	Dark eyes	31	5	43	4	18	6	32	1	0	24	11	17	0	21	0	35
	7 x 20	Mosaic eyes	40	2	44	4	38	0	36	0	0	19	0	34	0	20	0	35
		Total	71	7	87	8	56	6	68	1	0	43	11	51	0	41	0	70

Table S5.2: F_2 progeny of *bgcn*-Cas9 and w^{GDe} double heterozygotes crossed to LVP wild type. Mosaic eyes (ME), white eyes (WE). The ratio between the amount of dark and mosaic-eyed drive F_1s crossed is roughly proportional to the ratio of those phenotypes generated from the F_0 crosses.

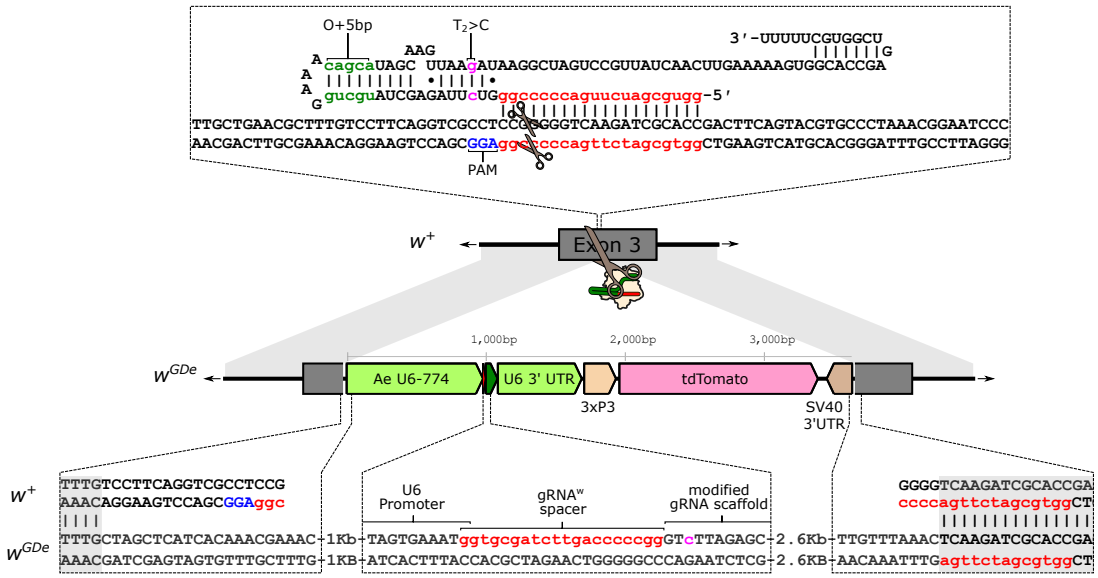


Figure S5.1: Detailed illustration of the w^{GDe} element and regions of homology with the w^{+} locus after a Cas9:gRNA mediated double-strand DNA break. The gRNA spacer sequence is printed in lowercase red letters as RNA and DNA. The three-base pair protospacer adjacent motif (PAM) is printed in Blue letters. Non-standard modifications to the gRNA scaffold sequence are printed in lowercase green (O+5bp) or magenta (T_2>C) letters and are based on Dang et al.[226] and similar modifications have been described in Feng et al.[227]. Regions of homology within the gRNA scaffold, between the gRNA and w^{+}, between the cut w^{+} locus and between the w^{GDe} locus are indicated with vertical lines (|). 22bp of the *white* exon 3 sequence is absent from the sequences flanking the w^{GDe} element. Once w^{+} exon 3 is cut, 18nt and 4nt need to be resected from the free DNA ends for them to become fully homologous to the sequences flanking the w^{GDe} element. The DNA sequence of the gRNA spacer is in the opposite orientation (reverse complement) relative to the same sequences in the flanking homology sequences. This may make it less likely that the gRNA spacer is involved in a partial homing event. The U6 promoter and 3′UTR sequence contain sequences found repeated elsewhere in the genome, in addition to their homology to the endogenous U6b (AAEL017774) gene. Truncations of Pol III promoters suggest that the U6 promoter and 3'UTR sequence used in the w^{GDe} element are likely substantially larger than the minimum sequence required to maintain efficient RNA expression[156].

nup50 Cross	F_1s crossed (♂x♀)	Drive F_1 phenotype	$+w^{GDe}$;+Cas9				$+w^{GDe}$;−				−/+Cas9				−/−			
			♀		♂		♀		♂		♀		♂		♀		♂	
			ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT
F_1: ♂w^{GDe};*nup50*-Cas9 x ♀LVP F_0: ♂Cas9 x ♀w^{GDe}	See Table S5.4	Mosaic eyes	486	0	100	29	1	462	1	80	0	1	0	306	0	0	0	353
F_1: ♂LVP x ♀w^{GDe};*nup50*-Cas9 F_0: ♂Cas9 x ♀w^{GDe}	See Table S5.5	Mosaic eyes, Brighter	22	0	23	0	15	0	22	0	0	11	0	6	0	10	0	11
		Mosaic eyes, Dimmer	209	0	236	0	207	0	218	0	0	101	0	94	0	91	0	101
		Total	231	0	259	0	222	0	240	0	0	112	0	100	0	101	0	112
F_1: ♂w^{GDe};*nup50*-Cas9 x ♀LVP F_0: ♂w^{GDe} x ♀Cas9	See Table S5.6	Mosaic eyes, Brighter	106	3	395	8	0	80	0	387	0	243	0	0	0	266	0	0
		Mosaic eyes, Dimmer	57	11	366	0	0	58	0	381	0	281	0	1	0	282	0	1
		Total	163	14	761	8	0	138	0	768	0	524	0	1	0	548	0	1
F_1: ♂LVP x ♀w^{GDe};*nup50*-Cas9 F_0: ♂w^{GDe} x ♀Cas9	See Table S5.7	Mosaic eyes, Brighter	115	2	119	2	109	3	115	1	1	40	0	42	0	34	1	38
		Mosaic eyes, Dimmer	122	0	160	0	144	2	160	1	0	65	0	65	1	78	2	79
		Total	237	2	279	2	253	5	275	2	1	105	0	107	1	112	3	117

Table S5.3: F_2 progeny of *nup50*-Cas9 and w^{GDe} double heterozygotes crossed to LVP wild type. Mosaic eyes (ME), white eyes (WE). Drive F_1 phenotypes are shown in Fig S5.2.

Cross	Drive F_1 phenotype	$+w^{GDe}$;+Cas9				$+w^{GDe}$;−				−/+Cas9				−/−			
		♀		♂		♀		♂		♀		♂		♀		♂	
		ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT
F_1: ♂w^{GDe};*nup50*-Cas9 x ♀LVP F_0: ♂Cas9 x ♀w^{GDe}	Mosaic	15		5			15		1				8				6
	Mosaic	12		7			11		7				17				19
	Mosaic	29		6	1		23		5				25				15
	Mosaic	34		4			34		4				13				21
	Mosaic	10		1	2		11		1				3				13
	Mosaic	21		8			24		2				14				21
	Mosaic	16		4			8						7				4
	Mosaic	15			1		9		4				12				19
	Mosaic	15		4			19		5				10				14
	Mosaic	1															
	Mosaic	5			2		8		2				7				3
	Mosaic						1										
	Mosaic	18					15		1				10				17
	Mosaic	3					6						3				4
	Mosaic	5		2	2		6		2				4				2
	Mosaic	18		7	2		24		6				11				13
	Mosaic	1											1				0
	Mosaic	19		4	2		18		8				6				8
	Mosaic	12					8						6				6
	Mosaic	19		4			15		9				2				7
	Mosaic	22		1			33		2				8				20
	Mosaic	23		5			11		5				9				9
	Mosaic	7		1			3						5				3
	Mosaic	10		1	2	1	15	1	3				8				9
	Mosaic	32		9			21		1				9				20
	Mosaic	16		2	3		23		3				15				12
	Mosaic	21		5	2		17						25				13
	Mosaic	21		4			11		3		1		7				8
	Mosaic	5		6			14						12				13
	Mosaic	15		1	3		9						11				8
	Mosaic	24		5	3		21		2				13				23
	Mosaic	22		4	4		29		4				25				23
	Total	486	0	100	29	1	462	1	80	0	1	0	306	0	0	0	353

Table S5.4: F_2 progeny of individual male *nup50*-Cas9 and w^{GDe} double heterozygotes crossed to LVP wild type. *nup50*-Cas9 from paternal F_0. Mosaic eyes (ME), white eyes (WE). Drive F_1 phenotypes are shown in Fig S5.2.

Cross	Drive F_1 phenotype	$+w^{GDe}$;+Cas9				$+w^{GDe}$;−				−/+Cas9				−/−			
		♀		♂		♀		♂		♀		♂		♀		♂	
		ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT
F_1: ♂LVP x ♀w^{GDe};*nup50*-Cas9 F_0: ♂Cas9 x ♀w^{GDe}	Mosaic, Dimmer	6		5		8		2			3		4		2		2
	Mosaic, Dimmer	10		9		10		7			1		3		3		4
	Mosaic, Dimmer	7		6		7		6			4				4		5
	Mosaic, Dimmer	20		12		0		11					5		3		4
	Mosaic, Dimmer	7		12		8		15			5		4		5		5
	Mosaic, Dimmer	5		15		9		20			5				9		4
	Mosaic, Dimmer	12		14		11		10			8		5		4		8
	Mosaic, Dimmer	3				1									1		
	Mosaic, Dimmer	3		4		4		6			8		3		4		3
	Mosaic, Dimmer	7		6		2		6			4		4		6		5
	Mosaic, Dimmer	9		6		13		7			2		3		4		8
	Mosaic, Dimmer					2											
	Mosaic, Dimmer	5		9		17		13			10		5		3		4
	Mosaic, Dimmer	17		16		9		14			5		4		3		4
	Mosaic, Dimmer	7		17		5		12			3		5		5		5
	Mosaic, Dimmer	4		8		13		2			2		7		4		7
	Mosaic, Dimmer	3		11		3		2			6		4		1		2
	Mosaic, Dimmer	13		5		10		5			5		2		6		3
	Mosaic, Dimmer	19		16		7		8			6		6		3		4
	Mosaic, Dimmer	4		4		8		6			6		6		5		5
	Mosaic, Dimmer	6		5		3		5			2		1		6		2
	Mosaic, Dimmer	12		15		18		14			12		7		3		2
	Mosaic, Dimmer	9		15		16		18			2		7		2		3
	Mosaic, Dimmer	9		10		8		11					2				4
	Mosaic, Dimmer	12		16		15		18			2		7		5		8
	Mosaic, Brighter	5		7		4		6			1		3		3		4
	Mosaic, Brighter	6		5		4		1			2				3		3
	Mosaic, Brighter	8		11		5		14			6		2		3		3
	Mosaic, Brighter	2															
	Mosaic, Brighter	1				2		1			2		1		1		1
	Sum	231	0	259	0	222	0	240	0	0	112	0	100	0	101	0	112

Table S5.5: F_2 progeny of individual female *nup50*-Cas9 and w^{GDe} double heterozygotes crossed to LVP wild type. *nup50*-Cas9 from paternal F_0. Mosaic eyes (ME), white eyes (WE). Drive F_1 phenotypes are shown in Fig S5.2.

Cross	Drive F_1 phenotype	$+w^{GDe}$;+Cas9				$+w^{GDe}$;−				−/+Cas9				−/−			
		♀		♂		♀		♂		♀		♂		♀		♂	
		ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT
F_1: ♂ w^{GDe};*nup50*-Cas9 x ♀LVP F_0: ♂ w^{GDe} x ♀Cas9	Mosaic, Brighter	7		26			6		21		13				15		
	Mosaic, Brighter	4		17			1		15		9				14		
	Mosaic, Brighter	3		14			5		12		2				4		
	Mosaic, Brighter	7		35			10		25		11				12		
	Mosaic, Brighter	6		21			4		26		11				14		
	Mosaic, Brighter	4		26			3		22		24				17		
	Mosaic, Brighter	9		13			3		27		6				16		
	Mosaic, Brighter	4		11			6		6		4				12		
	Mosaic, Brighter	2		10	1		3		10		7				14		
	Mosaic, Brighter		3	17			1		14		11				15		
	Mosaic, Brighter	7		21			5		17		6				9		
	Mosaic, Brighter	11		12			4		14		9				7		
	Mosaic, Brighter	9		25			5		26		14				14		
	Mosaic, Brighter	6		33			7		26		32				25		
	Mosaic, Brighter	4		17			3		10		11				7		
	Mosaic, Brighter	6		12			2		15		11				10		
	Mosaic, Brighter	1		17			3		22		7				12		
	Mosaic, Brighter	2		12	7		4		17		15				7		
	Mosaic, Brighter	5		18					19		19				9		
	Mosaic, Brighter	1		17			2		16		6				15		
	Mosaic, Brighter	8		21			3		27		15				18		
	Mosaic, Dimmer			25			1		28		23				15		
	Mosaic, Dimmer	3		26					31		30				26		
	Mosaic, Dimmer	3		27			3		20		15				23		
	Mosaic, Dimmer	2		17			3		24		24				18		
	Mosaic, Dimmer	4		23			2		22		13				13		
	Mosaic, Dimmer	2		20			3		15		10		1		11		
	Mosaic, Dimmer	3		16			1		25		18				14		
	Mosaic, Dimmer	16		29			13		32		21				17		
	Mosaic, Dimmer		2	7					7		2				4		
	Mosaic, Dimmer			22			4		13		14				19		
	Mosaic, Dimmer	3		14			3		16		16				14		
	Mosaic, Dimmer	1		11					12		4				5		
	Mosaic, Dimmer	1		12			1		20		9				7		
	Mosaic, Dimmer		4	12					8		3				10		
	Mosaic, Dimmer	3		30			5		28		17				22		
	Mosaic, Dimmer	6		21			8		31		27				22		
	Mosaic, Dimmer	1	2	13			4		12		9				14		
	Mosaic, Dimmer	5	1	9			4		9		8				8		
	Mosaic, Dimmer	3		18			2		18		14				12		1
	Mosaic, Dimmer	1	2	14			1		10		4				8		
	Sum	163	14	761	8	0	138	0	768	0	524	0	1	0	548	0	1

Table S5.6: F_2 progeny of individual male *nup50*-Cas9 and w^{GDe} double heterozygotes crossed to LVP wild type. *nup50*-Cas9 from maternal F_0. Mosaic eyes (ME), white eyes (WE). Drive F_1 phenotypes are shown in Fig S5.2.

Cross	Drive F_1 phenotype	$+w^{GDe}$;+Cas9				$+w^{GDe}$;−				−/+Cas9				−/−			
		♀		♂		♀		♂		♀		♂		♀		♂	
		ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT
F_1: ♂LVP x ♀w^{GDe};*nup50*-Cas9 F_0: ♂w^{GDe} x ♀Cas9	Mosaic, Brighter	0				1							1				
	Mosaic, Brighter	2		3		9		7			3		3		2		1
	Mosaic, Brighter	3		4		5		2			0						1
	Mosaic, Brighter	10	1	13		6	1	11	1	1	3		1				1
	Mosaic, Brighter	12		9		6		8			2		3		5		5
	Mosaic, Brighter	1		6		6		6			2		2		4		1
	Mosaic, Brighter	16	1	13		6	2	4			1						
	Mosaic, Brighter	2		2		3		3			0				1		
	Mosaic, Brighter	4		6		5		5			9		2		4		3
	Mosaic, Brighter	4		8	2	6		4			2		3				
	Mosaic, Brighter	5		5		3		7					3		3		3
	Mosaic, Brighter	3		9		5		3			3		1		2		1
	Mosaic, Brighter	18		16		10		16			5		3		6		8
	Mosaic, Brighter	4		2		3		2			1		1				1
	Mosaic, Brighter	4		5		10		5					3			1	2
	Mosaic, Brighter	6		3		5		3					3		1		3
	Mosaic, Brighter	2				1		3					1				3
	Mosaic, Brighter	7		4		5		10			2		1		2		1
	Mosaic, Brighter	3		3		4		5			3		6		1		2
	Mosaic, Brighter	5		1		5		3					2				1
	Mosaic, Brighter	1		3		2		7					2		1		1
	Mosaic, Brighter	3		4		3		1			4		1		2		
	Mosaic, Dimmer	9		8		9		8			2		5		6		3
	Mosaic, Dimmer	2		4		7		1			3		4		4		
	Mosaic, Dimmer	7		4		5		7			4		3	1	3		5
	Mosaic, Dimmer	6		10		7		3			2		3		2		4
	Mosaic, Dimmer	10		10		3		11			3		2		3		8
	Mosaic, Dimmer	4		6		6		12					2		1		1
	Mosaic, Dimmer	6		6		1		8			5		4		2		4
	Mosaic, Dimmer	6		5		6		9			5		1		2		2
	Mosaic, Dimmer	8		7		12		3			4		3		2	1	1
	Mosaic, Dimmer	2		2		5		7			1		4		3		
	Mosaic, Dimmer	13		17		14		9			4		2		3		10
	Mosaic, Dimmer	5		8		9		10			4		2		4		4
	Mosaic, Dimmer	6		10			2	3	1							1	3
	Mosaic, Dimmer	2		5				6									2
	Mosaic, Dimmer	3		5		4					1		1		2		3
	Mosaic, Dimmer	8		8		10		8			5		3		7		9
	Mosaic, Dimmer	5		5		2		12			3		2		3		3
	Mosaic, Dimmer	3		3		4		3			2		2		3		3
	Mosaic, Dimmer			2		3		6			2		1		1		2
	Mosaic, Dimmer	4		6		9		12			1		3		5		5
	Mosaic, Dimmer	1		6		8		1			5		5		9		
	Mosaic, Dimmer	2		2		2		6			3		7		3		4
	Mosaic, Dimmer	5		8		9		9			2		4		6		3
	Mosaic, Dimmer	5		13		9		6			4		2		4		
	Sum	237	2	279	2	253	5	275	2	1	105	0	107	1	112	3	117

Table S5.7: F_2 progeny of individual female *nup50*-Cas9 and w^{GDe} double heterozygotes crossed to LVP wildtype. *nup50*-Cas9 from maternal F_0. Mosaic eyes (ME), white eyes (WE). Drive F_1 phenotypes are shown in Fig S5.2.

Cas9	Drive F_1 phenotype	$+w^{GDe}$;+Cas9				$+w^{GDe}$;−				−/+Cas9				−/−			
Cross		♀		♂		♀		♂		♀		♂		♀		♂	
		ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT
F_1: ♂w^{GDe} x ♀LVP F_0: ♂w^{GDe} x ♀ LVP	Dark eyes	0	0	0	0	0	11	0	609	0	0	0	0	0	581	0	2

F_1: ♂LVP x ♀w^{GDe} F_0: ♂w^{GDe} x ♀ LVP	Dark eyes	0	0	0	0	0	145	0	163	0	0	0	0	0	147	0	150

Table S5.8: F_2 progeny of w^{GDe} heterozygotes (no Cas9) crossed to LVP wild type. Mosaic eyes (ME), white eyes (WE)

Cas9	Cas9 F_0	Drive F_1	$+w^{GDe}$ F_2	Total F_2	%	p-value	Threshold	Odds ratio
-Cas9	N/A	♂	620	1203	51.5%	1.00E+00	ns	1
-Cas9	N/A	♀	308	605	50.9%	1.00E+00	ns	1
sds3	♂	♂	275	450	61.1%	6.58E-02	ns	1.2
sds3	♀	♂	131	214	61.2%	1.57E-01	ns	1.2
sds3	♂	♀	159	275	57.8%	2.98E-01	ns	1.1
sds3	♀	♀	118	176	67.0%	4.95E-02	*	1.3
bgcn	♂	♂	444	885	50.2%	7.32E-01	ns	1.0
bgcn	♀	♂	257	389	66.1%	9.75E-03	**	1.3
bgcn	♂	♀	210	414	50.7%	1.00E+00	ns	1.0
bgcn	♀	♀	304	520	58.5%	1.75E-01	ns	1.1
nup50	♂	♂	1159	1819	63.7%	6.32E-04	***	1.2
nup50	♀	♂	1852	2926	63.3%	3.76E-04	***	1.2
nup50	♂	♀	952	1377	69.1%	1.67E-04	***	1.4
nup50	♀	♀	1055	1501	70.3%	6.61E-05	***	1.4

Table S5.9: Overall w^{GDe} inheritance bias. in each case, a two-sided Fisher's exact test is performed using the −Cas9 condition with the matching drive F_1 sex. Significance thresholds: * for ≤0.05, ** for ≤0.01, *** for ≤0.001.

Cas9	Cas9 F_0	Drive F_1	F_2 Cas9	F_2 w^{GDe}	ME/WE	Total	%	Control Difference	Ctrl ME/WE	Ctrl Total	Ctrl %	p-value	Threshold	Odds ratio
sds3	♂	♂	Yes	Yes	111	129	86.0%	Cas9 Positive Vs Negative	1	146	0.7%	1.58E-27	***	124.6
sds3	♀	♂	Yes	Yes	61	62	98.4%	Cas9 Positive Vs Negative	0	69	0.0%	1.65E-15	***	Inf
bgcn	♂	♂	Yes	Yes	14	196	7.1%	Cas9 Positive Vs Negative	1	248	0.4%	2.07E-04	***	17.6
bgcn	♀	♂	Yes	Yes	22	129	17.1%	Cas9 Positive Vs Negative	0	128	0.0%	1.04E-06	***	Inf
nup50	♂	♂	Yes	Yes	586	615	95.3%	Cas9 Positive Vs Negative	2	544	0.4%	1.45E-118	***	258.7
nup50	♀	♂	Yes	Yes	924	946	97.7%	Cas9 Positive Vs Negative	0	906	0.0%	9.62E-205	***	Inf

Table S5.10: Two-sided Fisher's exact test for somatic expression. Mosaic eyes (ME), white eyes (WE). Significance thresholds: * for ≤0.05, ** for ≤0.01, *** for ≤0.001.

Cas9	Cas9 F_0	Drive F_1	F_2 Cas9	F_2 w^{GDe}	ME/WE	Total	%	Control Difference	Ctrl ME/WE	Ctrl Total	Ctrl %	p-value	Threshold	Odds ratio
sds3	♂	♀	No	Yes	75	76	98.7%	Drive F_1 ♀ vs ♂	1	146	0.7%	1.80E-26	***	141.9
sds3	♀	♀	No	Yes	61	61	100.0%	Drive F_1 ♀ vs ♂	0	69	0.0%	7.36E-16	***	Inf
bgcn	♂	♀	No	Yes	39	98	39.8%	Drive F_1 ♀ vs ♂	1	248	0.4%	1.49E-18	***	97.9
bgcn	♀	♀	No	Yes	124	131	94.7%	Drive F_1 ♀ vs ♂	0	128	0.0%	1.61E-28	***	Inf
nup50	♂	♀	No	Yes	462	462	100.0%	Drive F_1 ♀ vs ♂	2	544	0.4%	2.29E-115	***	271.3
nup50	♀	♀	No	Yes	528	535	98.7%	Drive F_1 ♀ vs ♂	0	906	0.0%	8.43E-178	***	Inf

Table S5.11: Two-sided Fisher's exact test for maternal deposition. Mosaic eyes (ME), white eyes (WE). Significance thresholds: * for ≤0.05, ** for ≤0.01, *** for ≤0.001.

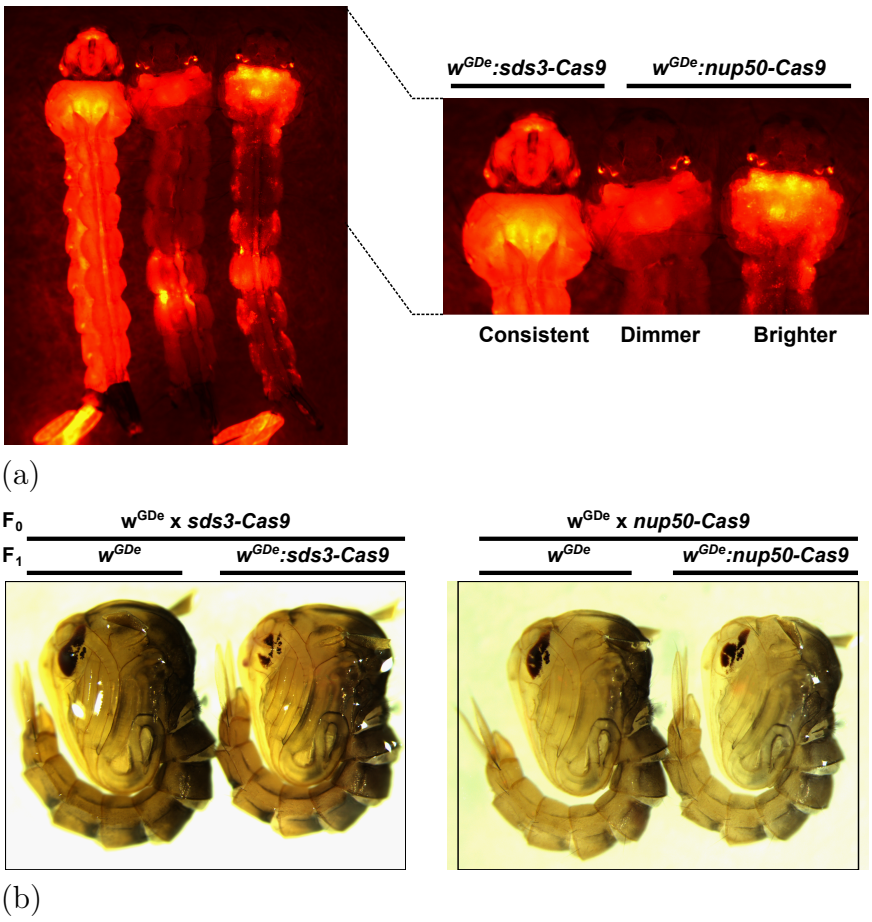
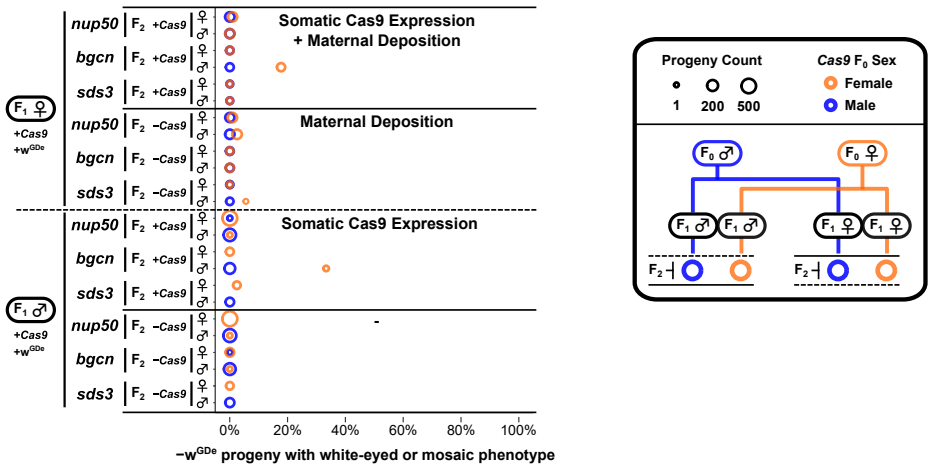


Figure S5.2: Fluorescent marker phenotype and *white* gene eye phenotype noted among the F_1s. a 'Dimmer' and 'Brighter' OpIE2-DsRED phenotype noted in the *nup50*-Cas9 line. The *sds3* and *bgcn*-Cas9 used PUb-mCherry-SV40 and had a consistent phenotype. b Examples of the F_1 dark-eyed (WT) and mosaic eyed phenotype.

Predictors	Estimate	S.E.	z-value	p-value	Threshold
bgcn-Cas9	-3.5012	0.2201	-15.909	<2e-16	***
nup50-Cas9	0.9341	0.1995	4.683	2.83E-06	***
+Cas9	6.2214	0.1759	35.365	<2e-16	***
Cas9 F_0 female	0.8699	0.1442	6.033	1.61E-09	***
gRNA:Cas9 F_1 female	8.0606	0.2280	35.351	<2e-16	***
Scored F_2 female	0.2328	0.1397	1.667	0.0955	ns
Constant	-5.0034	0.2673	-18.721	<2e-16	***

Table S5.12: The results of the binomial generalised linear model applied to mosaic eyed (ME) or white-eyed (WE) phenotype fraction among all $+w^{GDe}$ F_2 progeny in Table S5.1-S5.3. No interaction terms were specified. The *sds3*-Cas9, -Cas9, Cas9 F_0 male, gRNA:Cas9 F_1 male, F_2 male cross serves as the reference result. Significance thresholds: * for $\leq$0.05, ** for $\leq$0.01, *** for $\leq$0.001.



(a)

Figure S5.3: Somatic eye phenotype in the $-w^{\mathrm{GDe}}$ progeny of double heterozygote split drive carriers. The percentage of $-w^{\mathrm{GDe}}$ progeny that display a mosaic or total loss of eye pigment phenotype. F_2 progeny are segregated by the drive-carrying F_1's sex (♂, ♀), the F_2's Cas9 transgene inheritance (−Cas9, +Cas9), the Cas9 regulatory sequences (*sds3*, *bgcn*, *nup50*), and the F_2's sex (♂, ♀). The circle size indicates the number of progeny that make up that group, and circle colour indicates if the Cas9 carrying F_0 grandparent was male (Blue) or female (Orange). The set of progeny that came from drive F_1 carrying females are indicated with 'Maternal Deposition'. The set of progeny that inherited a Cas9 element are indicated with 'Somatic Cas9 Expression'. Due to the sex-linkage of the w^{GDe} element from F_1 drive males the number of F_2 progeny are unequally distributed among the groups for each sex. This sex-bias is more pronounced for the $-w^{\mathrm{GDe}}$ progeny than for the $+w^{\mathrm{GDe}}$ progeny due to the the effects of homing. Circles are absent if no progeny matched that grouping. Within matched crosses (each row), differences in the white phenotype rate corresponding to the Cas9 carrying F_0's sex cannot be solely attributed to a grandparent enhanced somatic phenotype for $-w^{\mathrm{GDe}}$ progeny. White phenotype rates for $+w^{\mathrm{GDe}}$ progeny are shown in Fig 5.1d. Note that inheritance of the w^{GDe} element also prevents the potential inheritance from the drive parent of an undamaged wild-type *white* gene or a *white* gene mutation that retains its function (r1) in pigment production. As such, any difference in white phenotype rates between $+w^{\mathrm{GDe}}$ and $-w^{\mathrm{GDe}}$ progeny cannot be solely attributed to gRNA expression from w^{GDe} in the F_2 progeny.

Cas9	Drive F_1	Cas9 F_0	Positives	Total Events	%	p-value	Threshold	Odds ratio
-Cas9	♂	N/A	11	592	1.9%	1.95E-01	ns	1.7
sds3	♂	♂	41	216	19.0%	6.79E-22	***	17.5
sds3	♂	♀	19	102	18.6%	4.93E-13	***	17.2
bgcn	♂	♂	24	464	5.2%	3.99E-06	***	4.8
bgcn	♂	♀	32	160	20.0%	2.89E-19	***	18.4
nup50	♂	♂	210	869	24.2%	3.50E-57	***	22.3
nup50	♂	♀	315	1387	22.7%	1.02E-61	***	21.0
Li et al. *nup50*	♂	♂	3	690	0.4%	1.93E-01	ns	0.4
Li et al. *nup50*	♂	♀	3	688	0.4%	1.94E-01	ns	0.4

Table S5.13: Homing significance test based on the recombination of the $+w^{\mathrm{GDe}}$ allele and sex-determining region. The total events are all female F_2s for the ♀ Cas9 F_0 (and therefore ♂ w^{GDe} F_0) crosses. Positives are $+w^{\mathrm{GDe}}$ female F_2s which should only occur through recombination. For ♂ Cas9 F_0 crosses male F_2s are considered. In each case, a two-sided Fisher's exact test is performed with 13/1203 as the expected outcome (this includes recombination of the *white* allele in the -Cas9 cross). Significance thresholds: * for $\leq$0.05, ** for $\leq$0.01, *** for $\leq$0.001.

Li et al. *nup50*	$+w^{GDe}$				$-w^{GDe}$			
Cross	♀		♂		♀		♂	
	ME/WE	WT	ME/WE	WT	ME/WE	WT	ME/WE	WT
F_1: ♂w^{GDe};*nup50*-Cas9 x ♀LVP F_0: ♂Cas9 x ♀w^{GDe}	650	654	0	3	0	2	0	687
F_1: ♂LVP x ♀w^{GDe};*nup50*-Cas9 F_0: ♂Cas9 x ♀w^{GDe}	757	0	790	0	0	162	0	175
F_1: ♂w^{GDe};*nup50*-Cas9 x ♀LVP F_0: ♂w^{GDe} x ♀Cas9	0	3	701	669	0	685	0	1
F_1: ♂LVP x ♀w^{GDe};*nup50*-Cas9 F_0: ♂w^{GDe} x ♀Cas9	743	0	763	0	3	175	1	185

Table S5.14: F_2 progeny of *nup50*-Cas9 and w^{GDe} double heterozygotes crossed to wild type taken from Li et al. Mosaic eyes (ME), white eyes (WE). Data are, in row order, from Li et al. supplemental files 4f, 4e, 4d, and 4c.

Cas9	drive F_1	Cas9 F_0	♂ F_2s	♂+♀ F_2s	%	p-value	Threshold	Odds ratio
-Cas9	♂	N/A	611	1203	50.8%	1.00E+00	ns	1.00
sds3	♂	♂	216	450	48.0%	5.65E-01	ns	0.95
sds3	♂	♀	112	214	52.3%	8.49E-01	ns	1.03
bgcn	♂	♂	464	885	52.4%	6.77E-01	ns	1.03
bgcn	♂	♀	229	389	58.9%	1.29E-01	ns	1.16
nup50	♂	♂	869	1819	47.8%	3.48E-01	ns	0.94
nup50	♂	♀	1539	2926	52.6%	5.57E-01	ns	1.04
Li et al. *nup50*	♂	♂	690	1996	34.6%	7.91E-09	***	0.68
Li et al. *nup50*	♂	♀	1371	2059	66.6%	8.10E-06	***	1.31

Table S5.15: Meiotic drive significance test based on sex bias in F_2 progeny. In each case, a two-sided Fisher's exact test is performed with 611/1203 as the expected outcome. Significance thresholds: * for $\leq$0.05, ** for $\leq$0.01, *** for $\leq$0.001.

Chapter 6

Conclusions

The central theme of this thesis has been understanding the interaction of transgene design with the nuclease process that underlies gene drive performance. We have shown that daisy-chain gene drives can be powerful tools for limiting genetic modification to target populations. However, their effectiveness is highly dependent on the efficiency and fidelity of the inheritance biasing process. We next performed a systematic meta-analysis of experimental homing gene drive data. This allowed us to investigate on a large scale the effect of sex and deposition on many different gene drive designs. Based on the analysis performed for our meta-analysis, we discovered that homing gene drives may not always function through homing. We analyse new experimental results that indicate that DNA repair outcomes underlying gene drive inheritance can vary between experimental conditions.

Potentially, the most important question that our work raises is about the predictability of fundamental features of gene drives, such as their inheritance biasing mechanism. In chapter 5, we attempt to replicate previous work with individuals derived from the same transgenic lines. We find that inheritance bias occurred through separate mechanisms in our replication attempt. This difference in the mechanism depends on factors that we do not currently understand. We would probably have ascribed this difference to any of a range of design features if we had not performed an exact replication cross. I think this type of misattribution frequently occurs in the literature.

We have developed a framework for systematically analysing different design factors. We found that anything beyond the most basic gene drive design elements cannot be interrogated with the current approach to gene drive testing. Many experimental, biological, and transgene design factors are simply not independently varied. As such, ascribing an improvement in performance to a different nuclease promoter or target gene is often not justified given the many (under-emphasised) other differences between experiments. For example, we and others have discussed the observation that homing gene drives seem to function disproportionately well in *Anopheles* mosquitoes. Based on the work presented in our meta-analysis, I am less confident that this is actually a reflection of species biology and not due to a technical design bottleneck.

Most of the papers I identified as relevant in our literature search came out during the period of my PhD. The body of literature is becoming more expansive than any student can usefully absorb. I think it is vital that we develop systems to more effectively make use of this body of work to identify the gaps in our understanding and perform deliberate replications informed by previous work. The systematic and detailed evaluation of published data has led to the discovery of multiple surprising/anomalous observations in the published literature. The work in chapter 5 led to new insights in collaboration with the original authors. However, in a different case, it sadly led to the retraction of a publication due to oversights in the experimental methods. This outcome is highly damaging, especially to the junior scientist involved. During data extraction, I observed less severe discrepancies in other published works. This is an inevitable part of the human endeavour of science, but current publication systems are not well suited to correct these kinds of errors. The correction process is hurt by using the same processes to signal misconduct, increasing the stakes and stigma for correcting genuine errors[228]. A less all-or-nothing approach to scientific corrections and evolving documents would aid in keeping the scientific record clean. Replications can also be applied to computational modelling, as I did in chapter 3. I think it would be a valuable contribution to the field to do this more broadly both as a validation of the models and as a way to make explicit different assumptions and approaches.

Our findings of unexpected repair outcomes after CRISPR editing fit with the work studying the use of nuclease for gene therapy. There is increasing evidence that Cas9 can cause large deletions and genomic rearrangements that can often go undetected by common PCR-based molecular assays[67;159;229]. Furthermore, gene editing can select for cells with deregulated responses to DNA damage (p53 is often implicated)[230;231]. Nuclease gene drives are intended to generate DNA damage over many generations on a population scale. It may be worth investigating whether nuclease-based gene drives can similarly apply selective pressures[105] against DNA damage checkpoints of the individuals that carry them. Selection for individuals with a dysregulated DNA response may in turn, lead to higher intrinsic mutation rates. Normally, this would be detrimental to the target population, as the germline mutation rate will have been fine-tuned by evolution[232]. However, it may be adaptive in a scenario in which humans are suddenly (on an evolutionary timescale) actively trying to eradicate them through various means. The actual selective processes are more complex than that presented above, and it may instead be that a homing drive selects for higher-fidelity repair leading to a lower mutation rate. For a suppression drive, it also requires the gene drive to be successful enough to spread and select for DNA damage insensitivity but not be so successful as actually to achieve its goal. Follow-up computational modelling can be useful to explore if the above mechanisms are a plausible route towards unintended outcomes.

In our modelling of daisy-chain gene drives, we focused on the spreading of genetic trait through a target population (population replacement). While we performed a set of simulations of different suppression strategies, these mostly highlighted how much more challenging suppression will be with a self-limiting approach. Pursuing a replacement strategy introduces significant additional complexity to any genetic

biocontrol intervention. Regulators and sponsors need extensive evidence on safety and efficacy to consider the approval of any transgenic biocontrol intervention. Most genetic biocontrol field trials measure entomological outcomes in the form of a reduction in the pest population[30;38–40;233–235]. Measuring actual endpoints, such as disease burden or crop yield, is much more expensive and complex[236]. However, the link between insect population suppression and epidemiological endpoints often already has a strong evidence base from the use of pesticides[237]. There are no clear analogous interventions on which population replacement strategies can rely on to justify measuring only entomological outcomes (i.e., transgene introgression). Therefore, it may be significantly harder to prove with a field trial that population replacement is effective compared to population suppression.

A range of gene drive effector modifications are currently being developed[238;239], but it is untested how transgenic replacement approaches will be received by regulators and governments. In the United States, suppression and replacement strategies are regulated by separate agencies[240]. The persistence of the modification in the population may lead regulators to require more extensive safety studies. However, while replacement may have some drawbacks compared to suppression, our modelling clearly shows that it is much more practical to achieve[2]. Furthermore, the introduction of *wolbachia* into wild *Aedes* populations by the World Mosquito Programme is a replacement strategy and one of, if not the, largest genetic biocontrol projects currently in progress[241]. This clearly shows that replacement strategies can be effective and approved.

Daisy-chain gene drives and other self-limiting drives are extremely promising. However, their prominence in the public consciousness carries some risks. There are serious questions about whether we can technically make hyperefficient self-limiting systems and whether they will remain self-limiting once released. As such, society and regulators may currently be considering[242] the use of an idealised technology with fewer drawbacks that researchers fail to deliver. This may come at the cost of having a public discourse about whether those serious public health challenges merit the use of self-perpetuating gene drives or self-limiting systems that are not as efficient. I hope our work goes towards informing this discussion by critically evaluating the technical feasibility of certain designs. At the very least, our work highlights areas of technical development that are needed to make daisy-chain gene drives viable.

Gene drives based on programmable nucleases have the promise of functioning as a platform technology that can be implemented in many different species of pests. However, the efficiency of current homing gene drives in the academic literature is often hit-and-miss. Similar gene drive designs have worked well in one species and poorly in another. Recent work I contributed to during this PhD shows that after many attempts[3;95;96;172] high inheritance bias efficiencies can be achieved in *A. aegypti*[5]. While this is an important advance, many challenges remain, and if development proceeds similarly in other species, that will severely restrict what species this technology may practically be applied in. Moreover, it is not self-evident that shifting the general cost/benefit of genetic biocontrol through technological progress

will be sufficient.

Throughout the world, there are many irradiation-based genetic-biocontrol facilities operating for tens of species[243]. Each species has unique cost/benefit calculations given their rearing requirements, time to maturity, diet, sex-sorting requirements, irradiation sensitivity, dispersal range, promiscuity, density dependence, etc. The foundation of the field of synthetic gene drives is that genetic biocontrol will massively benefit from becoming more efficient through technological advancements. Successful SIT programmes have been implemented for *Cochliomyia hominivorax* (new world screwworm), *Pectinophora gossypiella* (pink bollworm), *Cydia pomonella* (codling moth), *Cactoblastis cactorum* (cactus moth), *Teia anartoides* (painted apple moth), and *Thaumatotibia leucotreta* (false codling moth)[18]. However, invariably these insects remain pests elsewhere in the world. Clearly, a variety of factors, such as geography, infrastructure, funding, and governance, contribute to this. However, it is likely difficult to identify cases where a more efficient technology is necessary and sufficient to overcome this.

While scaling is notoriously difficult for any biotechnology[244], the principle of genetic biocontrol has many favourable attributes in this regard. The standard approach to mass rearing already relies on non-specialised manual labour and largely standard equipment[245]. On the other hand, several scaling challenges can be a significant bottleneck, regardless of the underlying efficiency of the technology. The logistics of performing releases may be costly and not fully balanced by the suppression efficiency of the individual insect (e.g., releasing one insect every 100m will have limited savings compared to one hundred insects per 100m).

Instead of efficiency, demand (customer acquisition cost in commercial settings) may be the most substantial bottleneck for genetic biocontrol. There are a number of factors that can make the application model of genetic biocontrol complex. Delayed outcomes, regulation, costly to prove efficacy (field trials), often a mismatch between funders and beneficiaries, holdouts, freeloaders, and concerns over acceptance by end users in the case of crop pests (customers of the customers). Possibly most importantly, insect releases are generally not compatible with existing pest control (i.e., chemical pesticides) workflows, markets, and metrics (although they can be complementary in their mode of action). This likely creates enormous inertia and a high burden on genetic biocontrol programmes to educate stakeholders on the underlying principles and convince them that it is worth pursuing.

The increased adoption of genetic biocontrol may require substantial innovations in its business model, product, and outreach instead of (only) underlying technical innovation. This may include pest control as a service, supporting alternative participation structures (e.g., collectives instead of individual farmers), engaging with influencers (e.g., grower organisations), rebranding (green technology), and addressing perceived reputational risks by users. However, increasing the efficiency of the technology by orders of magnitude through the use of self-limiting gene drives may well give the headroom to overcome these non-technical challenges and costs. In addition, sentiments around genetic engineering and gene drive may shift in the future and may

even become a "selling" point as high-profile projects succeed.

In conclusion, the field of genetic biocontrol has already demonstrated its usefulness in combating a diverse range of insect pests through irradiation-based techniques. The future of genetic biocontrol hinges not only on technological advancements such as self-limiting gene drives but also on improving the way we communicate their potential and the systems we develop for informed decision-making.

Bibliography

1. Sebald A. N. Verkuijl, Joshua X. D. Ang, Luke Alphey, Michael B. Bonsall, and Michelle A. E. Anderson. The Challenges in Developing Efficient and Robust Synthetic Homing Endonuclease Gene Drives. *Frontiers in Bioengineering and Biotechnology*, 10, 3 2022. ISSN 2296-4185. doi: 10.3389/fbioe.2022.856981. URL https://www.frontiersin.org/article/10.3389/fbioe.2022.856981.

2. Sebald A N Verkuijl, Michelle A E Anderson, Luke Alphey, and Michael B Bonsall. Daisy-chain gene drives: The role of low cut-rate, resistance mutations, and maternal deposition. *PLOS Genetics*, 18(9):e1010370, 9 2022. doi: 10.1371/journal.pgen.1010370. URL https://doi.org/10.1371/journal.pgen.1010370.

3. Sebald A N Verkuijl, Estela Gonzalez, Ming Li, Joshua X. D. Ang, Nikolay P Kandul, Michelle A E Anderson, Omar S Akbari, Michael B Bonsall, and Luke Alphey. A CRISPR endonuclease gene drive reveals distinct mechanisms of inheritance bias. *Nature Communications*, 13(1):7145, 11 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-34739-y. URL https://www.nature.com/articles/s41467-022-34739-y.

4. Joshua Xin De Ang, Katherine Nevard, Rebekah Ireland, Deepak-Kumar Purusothaman, Sebald A N Verkuijl, Lewis Shackleford, Estela Gonzalez, Michelle A E Anderson, and Luke Alphey. Considerations for homology-based DNA repair in mosquitoes: Impact of sequence heterology and donor template source. *PLOS Genetics*, 18(2):e1010060, 2 2022. ISSN 1553-7404. doi: 10.1371/journal.pgen.1010060. URL https://doi.org/10.1371/journal.pgen.1010060.

5. Michelle A E Anderson, Estela Gonzalez, Joshua X D Ang, Lewis Shackleford, Katherine Nevard, Sebald A N Verkuijl, Matthew P Edgington, Tim Harvey-Samuel, and Luke Alphey. Closing the gap to effective gene drive in Aedes aegypti by exploiting germline regulatory elements. *Nature Communications*, 14(1):338, 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-36029-7. URL https://doi.org/10.1038/s41467-023-36029-7.

6. Zhi-Qiang Zhang. Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness (Addenda 2013). *Zootaxa*, 3703:1–82, 2013. ISSN 1175-5326 (Print). doi: 10.11646/zootaxa.3703.1.1.

7. E.-C. OERKE. Crop losses to pests. *The Journal of Agricultural Science*, 144(1):31–43, 2 2006. ISSN 0021-8596. doi: 10.1017/S0021859605005708.

8. Thomas W Culliney. Crop Losses to Arthropods. 2014.

9. Serge Savary, Laetitia Willocquet, Sarah Jane Pethybridge, Paul Esker, Neil McRoberts, and Andy Nelson. The global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution*, 3(3):430–439, 2019. ISSN 2397-334X. doi: 10.1038/s41559-018-0793-y. URL https://doi.org/10.1038/s41559-018-0793-y.

10. Christophe Diagne, Boris Leroy, Anne-Charlotte Vaissière, Rodolphe E Gozlan, David Roiz, Ivan Jarić, Jean-Michel Salles, Corey J A Bradshaw, and Franck Courchamp. High and rising economic costs of biological invasions worldwide. *Nature*, 592(7855): 571–576, 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03405-6. URL https://doi.org/10.1038/s41586-021-03405-6.

11. World Health Organization and UNICEF. Global vector control response 2017-2030. 2017. ISSN 9241512970.

12. World Health Organization and others. *World malaria report 2022*. World Health Organization, 2022.

13. Eben H Paxton, Richard J Camp, P Marcos Gorresen, Lisa H Crampton, David L Leonard, and Eric A VanderWerf. Collapsing avian community on a Hawaiian island. *Science Advances*, 2(9):e1600029, 2016. doi: 10.1126/sciadv.1600029. URL https://www.science.org/doi/abs/10.1126/sciadv.1600029.

14. Nathan Donley. The USA lags behind other agricultural nations in banning harmful pesticides. *Environmental Health*, 18(1): 44, 2019. ISSN 1476-069X. doi: 10.1186/s12940-019-0488-0. URL https://doi.org/10.1186/s12940-019-0488-0.

15. P Indira Devi, M Manjula, and R V Bhavani. Agrochemicals, Environment, and Human Health. *Annual Review of Environment and Resources*, 47(1):399–421, 10 2022. ISSN 1543-5938. doi: 10.1146/annurev-environ-120920-111015. URL https://doi.org/10.1146/annurev-environ-120920-111015.

16. Luke Alphey. Genetic Control of Mosquitoes. *Annual Review of Entomology*, 59(1):205–224, 1 2014. ISSN 0066-4170. doi: 10.1146/annurev-ento-011613-162002. URL https://doi.org/10.1146/annurev-ento-011613-162002.

17. E. F. Knipling. Possibilities of Insect Control or Eradication Through the Use of Sexually Sterile Males. *Journal of Economic Entomology*, 48(4):459–462, 8 1955. ISSN 1938-291X. doi: 10.1093/jee/48.4.459. URL http://academic.oup.com/jee/article/48/4/459/2205947/Possibilities-of-Insect-Control-or-Eradication.

18. Waldemar. Klassen and Christopher F Curtis. History of the Sterile Insect Technique. In V A Dyck, J Hendrichs, and A S Robinson, editors, *Sterile Insect Technique*, pages 3–36. Springer-Verlag, Berlin/Heidelberg, 2005. ISBN 978-1-4020-4051-1. doi: 10.1007/1-4020-4051-2{_}1. URL http://link.springer.com/10.1007/1-4020-4051-2_1.

19. A Bakri, K Mehta, and D R Lance. Sterilizing Insects with Ionizing Radiation. In V A Dyck, J Hendrichs, and A S Robinson, editors, *Sterile Insect Technique: Principles and Practice in Area-Wide Integrated Pest Management*, pages 233–268. Springer Netherlands, Dordrecht, 2005. ISBN 978-1-4020-4051-1. doi: 10.1007/1-4020-4051-2{_}9. URL https://doi.org/10.1007/1-4020-4051-2_9.

20. Victor A Dyck, Jorge Hendrichs, and Alan S Robinson. *Sterile insect technique: principles and practice in area-wide integrated pest management*. Taylor & Francis, 2021.

21. P Rendón, Donald McInnis, D Lance, and J Stewart. Medfly (Diptera: Tephritidae) genetic sexing: large-scale field comparison of males-only and bisexual sterile fly releases in Guatemala. *Journal of economic entomology*, 97(5):1547–1553, 2004. ISSN 1938-291X.

22. J D Mumford. Application of Benefit/Cost Analysis to Insect Pest Control Using the Sterile Insect Technique. In V A Dyck, J Hendrichs, and A S Robinson, editors, *Sterile Insect Technique*, pages 481–498. Springer-Verlag, Berlin/Heidelberg, 2005. ISBN 978-1-4020-4051-1. doi: 10.1007/1-4020-4051-2{_}18. URL https://doi.org/10.1007/1-4020-4051-2_18http://link.springer.com/10.1007/1-4020-4051-2_18.

23. Jacob E Crawford, David W Clarke, Victor Criswell, Mark Desnoyer, Devon Cornel, Brittany Deegan, Kyle Gong, Kaycie C Hopkins, Paul Howell, Justin S Hyde, Josh Livni, Charlie Behling, Renzo Benza, Willa Chen, Karen L Dobson, Craig Eldershaw, Daniel Greeley, Yi Han, Bridgette Hughes, Evdoxia Kakani, Joe Karbowski, Angus Kitchell, Erika Lee, Teresa Lin, Jianyi Liu, Martin Lozano, Warren MacDonald, James W Mains, Matty Metlitz, Sara N Mitchell, David Moore, Johanna R Ohm, Kathleen Parkes, Alexandra Porshnikoff, Chris Robuck, Martin Sheridan, Robert Sobecki, Peter Smith, Jessica Stevenson, Jordan Sullivan, Brian Wasson, Allison M Weakley, Mark Wilhelm, Joshua Won, Ari Yasunaga, William C Chan, Jodi Holeman, Nigel Snoad, Linus Upson, Tiantian Zha, Stephen L Dobson, F Steven Mulligan, Peter Massaro, and Bradley J White. Efficient production of male Wolbachia-infected Aedes aegypti mosquitoes enables large-scale suppression of wild populations. *Nature Biotechnology*, 38(4):482–492, 2020. ISSN 1546-1696. doi: 10.1038/s41587-020-0471-x. URL https://doi.org/10.1038/s41587-020-0471-x.

24. Célia Lutrat, David Giesbrecht, Eric Marois, Steve Whyard, Thierry Baldet, and Jérémy Bouyer. Sex Sorting for Pest Control: It's Raining Men! *Trends in Parasitology*, 35(8):649–662, 2019. ISSN 1471-4922. doi: https://doi.org/10.1016/j.pt.2019.06.001. URL https://www.sciencedirect.com/science/article/pii/S1471492219301321.

25. Carlos F Marina, Pablo Liedo, J Guillermo Bond, Adriana R Osorio, Javier Valle, Roberto Angulo-Kladt, Yeudiel Gómez-Simuta, Ildefonso Fernández-Salas, Ariane Dor, and Trevor Williams. Comparison of Ground Release and Drone-Mediated Aerial Release of Aedes aegypti Sterile Males in Southern Mexico: Efficacy and Challenges. *Insects*, 13(4), 3 2022. ISSN 2075-4450 (Print). doi: 10.3390/insects13040347.

26. Gabriel Carrasco-Escobar, Marta Moreno, Kimberly Fornace, Manuela Herrera-Varela, Edgar Manrique, and Jan E Conn. The use of drones for mosquito surveillance and control. *Parasites & Vectors*, 15(1):473, 2022. ISSN 1756-3305. doi: 10.1186/s13071-022-05580-5. URL https://doi.org/10.1186/s13071-022-05580-5.

27. R C Smith, M F Walter, R H Hice, D A O'Brochta, and P W Atkinson. Testis-specific expression of the beta2 tubulin promoter of Aedes aegypti and its application as a genetic sex-separation marker. *Insect molecular biology*, 16(1):61–71, 2 2007. ISSN 0962-1075 (Print). doi: 10.1111/j.1365-2583.2006.00701.x.

28. Xiaoying Zheng, Dongjing Zhang, Yongjun Li, Cui Yang, Yu Wu, Xiao Liang, Yongkang Liang, Xiaoling Pan, Linchao Hu, Qiang Sun, Xiaohua Wang, Yingyang Wei, Jian Zhu, Wei Qian, Ziqiang Yan, Andrew G. Parker, Jeremie R. L. Gilles, Kostas Bourtzis, Jérémy Bouyer, Moxun Tang, Bo Zheng, Jianshe Yu, Julian Liu, Jiajia Zhuang, Zhigang Hu, Meichun Zhang, Jun-Tao Gong, Xiao-Yue Hong, Zhoubing Zhang, Lifeng Lin, Qiyong Liu, Zhiyong Hu, Zhongdao Wu, Luke Anthony Baton, Ary A. Hoffmann, and Zhiyong Xi. Incompatible and sterile insect techniques combined eliminate mosquitoes. *Nature*, 572(7767):56–61, 8 2019. ISSN 0028-0836. doi: 10.1038/s41586-019-1407-9. URL http://www.nature.com/articles/s41586-019-1407-9.

29. Dean D. Thomas. Insect Population Control Using a Dominant, Repressible, Lethal Genetic System. *Science*, 287(5462):2474–2476, 3 2000. ISSN 00368075. doi: 10.1126/science.287.5462.2474. URL http://science.sciencemag.org/content/287/5462/2474.abstracthttps://www.sciencemag.org/lookup/doi/10.1126/science.287.5462.2474.

30. Siân A M Spinner, Zoe H Barnes, Alin Mirel Puinean, Pam Gray, Tarig Dafa'alla, Caroline E Phillips, Camila Nascimento de Souza, Tamires Fonseca Frazon, Kyla Ercit, Amandine Collado, Neil Naish, Edward Sulston, Gwilym C Ll. Phillips, Kelleigh K Greene, Mattia Poletto, Benjamin D Sperry, Simon A Warner, Nathan R Rose, Grey K Frandsen, Natalia C Verza, Kevin J Gorman, and Kelly J Matzen. New self-sexing Aedes aegypti strain eliminates barriers to scalable and sustainable vector control for governments and communities in dengue-prone environments. *Frontiers in Bioengineering and Biotechnology*, 10, 2022. ISSN 2296-4185. URL https://www.frontiersin.org/articles/10.3389/fbioe.2022.975786.

31. Blandine Massonnet-Bruneel, Nicole Corre-Catelin, Renaud Lacroix, Rosemary S Lees, Kim Phuc Hoang, Derric Nimmo, Luke Alphey, and Paul Reiter. Fitness of Transgenic Mosquito Aedes aegypti Males Carrying a Dominant Lethal Genetic System. *PLOS ONE*, 8(5):e62711, 5 2013. URL https://doi.org/10.1371/journal.pone.0062711.

32. M H Andreasen and C F Curtis. Optimal life stage for radiation sterilization of Anopheles males and their fitness for release. *Medical and veterinary entomology*, 19(3):238–244, 2005. ISSN 0269-283X.

33. Austin Burt and Anne Deredec. Self-limiting population genetic control with sex-linked genome editors. *Proceedings of the Royal Society B: Biological Sciences*, 285(1883):20180776, 2018. doi: 10.1098/rspb.2018.0776. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2018.0776.

34. Paola Pollegioni, Ace R North, Tania Persampieri, Alessandro Bucci, Roxana L Minuz, David Alexander Groneberg, Tony Nolan, Philippos-Aris Papathanos, Andrea Crisanti, and Ruth Müller. Detecting the population dynamics of an autosomal sex ratio distorter transgene in malaria vector mosquitoes. *Journal of Applied Ecology*, 57(10):2086–2096, 2020. doi: https://doi.org/10.1111/1365-2664.13702. URL https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/1365-2664.13702.

35. Tim Harvey-Samuel, Thomas Ant, and Luke Alphey. Towards the genetic control of invasive species. *Biological Invasions*, 19(6):1683–1703, 2017. ISSN 1573-1464. doi: 10.1007/s10530-017-1384-6. URL https://doi.org/10.1007/s10530-017-1384-6.

36. Robyn R Raban, John M Marshall, and Omar S Akbari. Progress towards engineering gene drives for population control. *The Journal of experimental biology*, 223(Pt Suppl 1), 2 2020. ISSN 1477-9145 (Electronic). doi: 10.1242/jeb.208181.

37. Michael Francisco. Armyworm meets Friendly moth. *Nature Biotechnology*, 39(5):532, 2021. ISSN 1546-1696. doi: 10.1038/s41587-021-00932-5. URL https://doi.org/10.1038/s41587-021-00932-5.

38. Anthony M. Shelton, Stefan J. Long, Adam S. Walker, Michael Bolton, Hilda L. Collins, Loïc Revuelta, Lynn M. Johnson, and Neil I. Morrison. First Field Release of a Genetically Engineered, Self-Limiting Agricultural Pest Insect: Evaluating Its Potential for Future Crop Protection. *Frontiers in Bioengineering and Biotechnology*, 7(January):1–15, 1 2020. ISSN 2296-4185. doi: 10.3389/fbioe.2019.00482. URL https://www.frontiersin.org/article/10.3389/fbioe.2019.00482/full.

39. Gregory S Simmons, Andrew R McKemey, Neil I Morrison, Sinead O'Connell, Bruce E Tabashnik, John Claus, Guoliang Fu, Guolei Tang, Mickey Sledge, Adam S Walker, Caroline E Phillips, Ernie D Miller, Robert I Rose, Robert T Staten, Christl A Donnelly, and Luke Alphey. Field Performance of a Genetically Engineered Strain of Pink Bollworm. *PLOS ONE*, 6(9):1–11, 2011. doi: 10.1371/journal.pone.0024110. URL https://doi.org/10.1371/journal.pone.0024110.

40. Angela F Harris, Derric Nimmo, Andrew R McKemey, Nick Kelly, Sarah Scaife, Christl A Donnelly, Camilla Beech, William D Petrie, and Luke Alphey. Field performance of engineered male mosquitoes. *Nature Biotechnology*, 29(11):1034–1037, 2011. ISSN 1546-1696. doi: 10.1038/nbt.2019. URL https://doi.org/10.1038/nbt.2019.

41. Z. S. Brown, M. S. Jones, and J. Mumford. Economic principles and concepts in area-wide genetic pest management. In *The economics of integrated pest management of insects*, pages 96–121. CABI, UK, 1 2019. doi: 10.1079/9781786393678.0096. URL http://www.cabidigitallibrary.org/doi/10.1079/9781786393678.0096.

42. World Health Organization. *Guidance framework for testing genetically modified mosquitoes*. World Health Organization, Geneva, 2nd ed edition, 2021. ISBN 9789240025233 (electronic version). URL https://apps.who.int/iris/handle/10665/341370.

43. Kenneth A Oye, Kevin Esvelt, Evan Appleton, Flaminia Catteruccia, George Church, Todd Kuiken, Shlomiya Bar-Yam Lightfoot, Julie McNamara, Andrea Smidler, and James P Collins. Regulating gene drives. *Science*, 345(6197):626–628, 2014. doi: 10.1126/science.1254287. URL https://www.science.org/doi/abs/10.1126/science.1254287.

44. Stephanie L James, Brinda Dass, and Hector Quemada. Regulatory and policy considerations for the implementation of gene drive-modified mosquitoes to prevent malaria transmission. *Transgenic Research*, 2023. ISSN 1573-9368. doi: 10.1007/s11248-023-00335-z. URL https://doi.org/10.1007/s11248-023-00335-z.

45. N/A. Containment Practices for Arthropods Modified with Engineered Transgenes Capable of Gene Drive Addendum 1 to the Arthropod Containment Guidelines, Version 3.2. *Vector-Borne and Zoonotic Diseases*, 22(1):3–17, 10 2021. ISSN 1530-3667. doi: 10.1089/vbz.2021.0035. URL https://www.liebertpub.com/doi/abs/10.1089/vbz.2021.0035.

46. Srinivasan Chandrasegaran and Dana Carroll. Origins of Programmable Nucleases for Genome Engineering. *Journal of Molecular Biology*, 428(5, Part B):963–989, 2016. ISSN 0022-2836. doi: https://doi.org/10.1016/j.jmb.2015.10.014. URL https://www.sciencedirect.com/science/article/pii/S0022283615006063.

47. Andrew Hammond, Roberto Galizi, Kyros Kyrou, Alekos Simoni, Carla Siniscalchi, Dimitris Katsanos, Matthew Gribble, Dean Baker, Eric Marois, Steven Russell, Austin Burt, Nikolai Windbichler, Andrea Crisanti, and Tony Nolan. A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nature Biotechnology*, 34(1):78–83, 1 2016. ISSN 1087-0156. doi: 10.1038/nbt.3439. URL http://www.nature.com/articles/nbt.3439https://www.nature.com/articles/nbt.3439.pdf.

48. Georg Oberhofer, Tobin Ivy, and Bruce A. Hay. Behavior of homing endonuclease gene drives targeting genes required for viability or female fertility with multiplexed guide RNAs. *Proceedings of the National Academy of Sciences*, 115(40):E9343–E9352, 10 2018. ISSN 0027-8424. doi: 10.1073/pnas.1805278115. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1805278115.

49. Nikolay P Kandul, Junru Liu, Jared B Bennett, John M Marshall, and Omar S Akbari. A confinable home-and-rescue gene drive for population modification. *eLife*, 10(39):24377–24383, 3 2021. doi: 10.7554/eLife.65939. URL https://elifesciences.org/articles/6593910.7554/eLife.65939.

50. Gerard Terradas, Anna B Buchman, Jared B Bennett, Isaiah Shriner, John M Marshall, Omar S Akbari, and Ethan Bier. Inherently confinable split-drive systems in Drosophila. *Nature Communications*, 12(1):1480, 12 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-21771-7. URL http://www.nature.com/articles/s41467-021-21771-7.

51. Kyros Kyrou, Andrew M. Hammond, Roberto Galizi, Nace Kranjc, Austin Burt, Andrea K. Beaghton, Tony Nolan, and Andrea Crisanti. A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. *Nature Biotechnology*, 36(11):1062–1066, 11 2018. ISSN 1087-0156. doi: 10.1038/nbt.4245. URL http://www.nature.com/articles/nbt.4245.

52. Astrid Hoermann, Tibebu Habtewold, Prashanth Selvaraj, Giuseppe Del Corsano, Paolo Capriotti, Maria Grazia Inghilterra, Temesgen M Kebede, George K Christophides, and Nikolai Windbichler. Gene drive mosquitoes can aid malaria elimination by retarding Plasmodium sporogonic development. *Science Advances*, 8(38):2022.02.15.480588, 9 2022. ISSN 2375-2548. doi: 10.1126/sciadv.abo1733. URL http://biorxiv.org/content/early/2022/02/17/2022.02.15.480588.abstracthttps://www.science.org/doi/10.1126/sciadv.abo1733.

53. Valentino M. Gantz, Nijole Jasinskiene, Olga Tatarenkova, Aniko Fazekas, Vanessa M. Macias, Ethan Bier, and Anthony A. James. Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proceedings of the National Academy of Sciences*, 112(49):E6736–E6743, 12 2015. ISSN 0027-8424. doi: 10.1073/pnas.1521077112. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1521077112.

54. Delphine Thizy, Lea Pare Toe, Charles Mbogo, Damaris Matoke-Muhia, Vincent Pius Alibu, S Kathleen Barnhill-Dilling, Tracey Chantler, Gershom Chongwe, Jason Delborne, Lydia Kapiriri, Esther Nassonko Kavuma, Sethlomo Koloi-Keaikitse, Ana Kormos, Katherine Littler, Dickson Lwetoijera, Roberta Vargas de Moraes, Noni Mumba, Lilian Mutengu, Sylvia Mwichuli, Silvia Elizabeth Nabukenya, Janet Nakigudde, Paul Ndebele, Carolyne Ngara, Eric Ochomo, Simon Odiwuor Ondiek, Stephany Rivera, Aaron J Roberts, Benjamin Robinson, Rodrick Sambakunsi, Abha Saxena, Naima Sykes, Brian B Tarimo, Nicki Tiffin, and Karen H Tountas. Proceedings of an expert workshop on community agreement for gene drive research in Africa - Co-organised by KEMRI, PAMCA and Target Malaria., 2021. ISSN 2572-4754 (Electronic).

55. John Min, Andrea L Smidler, Devora Najjar, and Kevin M Esvelt. Harnessing gene drive. *Journal of Responsible Innovation*, 5(sup1):S40–S65, 1 2018. ISSN 2329-9460. doi: 10.1080/23299460.2017.1415586. URL https://doi.org/10.1080/23299460.2017.1415586.

56. Josef Zapletal, Neda Najmitabrizi, Madhav Erraguntla, Mark A. Lawley, Kevin M. Myles, and Zach N. Adelman. Making gene drive biodegradable. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1818):20190804, 2 2021. ISSN 0962-8436. doi: 10.1098/rstb.2019.0804. URL https://royalsocietypublishing.org/doi/10.1098/rstb.2019.0804.

57. Sumit Dhole, Michael R Vella, Alun L Lloyd, and Fred Gould. Invasion and migration of spatially self-limiting gene drives: A comparative analysis. *Evolutionary Applications*, 11(5):794–808, 2018. ISSN 1752-4571. doi: https://doi.org/10.1111/eva.12583. URL https://doi.org/10.1111/eva.12583.

58. Jaye Sudweeks, Brandon Hollingsworth, Dimitri V Blondel, Karl J Campbell, Sumit Dhole, John D Eisemann, Owain Edwards, John Godwin, Gregg R Howald, Kevin P Oh, Antoinette J Piaggio, Thomas A A Prowse, Joshua V Ross, J Royden Saah, Aaron B Shiels, Paul Q Thomas, David W Threadgill, Michael R Vella, Fred Gould, and Alun L Lloyd. Locally Fixed Alleles: A method to localize gene drive to island populations. *Scientific Reports*, 9(1):15821, 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-51994-0. URL https://doi.org/10.1038/s41598-019-51994-0.

59. Charleston Noble, John Min, Jason Olejarz, Joanna Buchthal, Alejandro Chavez, Andrea L. Smidler, Erika A. DeBenedictis, George M. Church, Martin A. Nowak, and Kevin M. Esvelt. Daisy-chain gene drives for the alteration of local populations. *Proceedings of the National Academy of Sciences*, 116(17):8275–8282, 4 2019. ISSN 0027-8424. doi: 10.1073/pnas.1716358116. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1716358116.

60. James E. DiCarlo, Alejandro Chavez, Sven L. Dietz, Kevin M. Esvelt, and George M. Church. Safeguarding CRISPR-Cas9 gene drives in yeast. *Nature Biotechnology*, 33(12):1250–1255, 12 2015. ISSN 1087-0156. doi: 10.1038/nbt.3412. URL http://www.nature.com/articles/nbt.3412.

61. Jackson Champer, Esther Lee, Emily Yang, Chen Liu, Andrew G Clark, and Philipp W Messer. A toxin-antidote CRISPR gene drive system for regional population modification. *Nature Communications*, 11(1):1082, 12 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-14960-3. URL https://doi.org/10.1038/s41467-020-14960-3.

62. Keun Chae, Chanell Dawson, Collin Valentin, Bryan Contreras, Josef Zapletal, Kevin M Myles, and Zach N Adelman. Engineering a self-eliminating transgene in the yellow fever mosquito, Aedes aegypti. *PNAS Nexus*, page pgac037, 3 2022. ISSN 2752-6542. doi: 10.1093/pnasnexus/pgac037. URL https://doi.org/10.1093/pnasnexus/pgac037.

63. Luke S Alphey, Andrea Crisanti, Filippo (Fil) Randazzo, and Omar S Akbari. Opinion: Standardizing the definition of gene drive. *Proceedings of the National Academy of Sciences*, 117(49):30864–30867, 12 2020. ISSN 0027-8424. doi: 10.1073/pnas.2020417117. URL http://www.pnas.org/content/early/2020/11/10/2020417117.abstracthttp://www.pnas.org/lookup/doi/10.1073/pnas.2020417117.

64. Austin Burt and Robert Trivers. *Genes in Conflict: The Biology of Selfish Genetic Elements*, volume 1. Harvard University Press, 2006. ISBN 9780674017139.

65. Martin Jinek, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, 337(6096):816–821, 8 2012. ISSN 0036-8075. doi: 10.1126/science.1225829. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1225829https://www.sciencemag.org/lookup/doi/10.1126/science.1225829.

66. Andrew J Aguirre, Robin M Meyers, Barbara A Weir, Francisca Vazquez, Cheng-Zhong Zhang, Uri Ben-David, April Cook, Gavin Ha, William F Harrington, Mihir B Doshi, Maria Kost-Alimova, Stanley Gill, Han Xu, Levi D Ali, Guozhi Jiang, Sasha Pantel, Yenarae Lee, Amy Goodale, Andrew D Cherniack, Coyin Oh, Gregory Kryukov, Glenn S Cowley, Levi A Garraway, Kimberly Stegmaier, Charles W Roberts, Todd R Golub, Matthew Meyerson, David E Root, Aviad Tsherniak, and William C Hahn. Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer discovery*, 6(8): 914–929, 8 2016. ISSN 2159-8290 (Electronic). doi: 10.1158/2159-8290.CD-16-0154.

67. Michael Kosicki, Kärt Tomberg, and Allan Bradley. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology*, 36(8):765–771, 9 2018. ISSN 1087-0156. doi: 10.1038/nbt.4192. URL https://doi.org/10.1038/nbt.4192http://www.nature.com/articles/nbt.4192.

68. Yuk-Sang Chan, Daniel A. Naujoks, David S. Huen, and Steven Russell. Insect Population Control by Homing Endonuclease-Based Gene Drive: An Evaluation in Drosophila melanogaster. *Genetics*, 188(1):33–44, 5 2011. ISSN 0016-6731. doi: 10.1534/genetics.111.127506. URL http://www.genetics.org/lookup/doi/10.1534/genetics.111.127506.

69. Chun-Chieh Lin and Christopher J. Potter. Editing Transgenic DNA Components by Inducible Gene Replacement in Drosophila melanogaster. *Genetics*, 203(4):1613–1628, 8 2016. ISSN 0016-6731. doi: 10.1534/genetics.116.191783. URL http://www.genetics.org/lookup/doi/10.1534/genetics.116.191783.

70. Austin Burt. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1518):921–928, 5 2003. doi: 10.1098/rspb.2002.2319. URL https://doi.org/10.1098/rspb.2002.2319.

71. Sofie A O'Brien, Kyoungho Lee, Hsu-Yuan Fu, Zion Lee, Tung S Le, Christopher S Stach, Meghan G McCann, Alicia Q Zhang, Michael J Smanski, Nikunj V Somia, and Wei-Shou Hu. Single Copy Transgene Integration in a Transcriptionally Active Site for Recombinant Protein Synthesis. *Biotechnology Journal*, 13(10):1800226, 10 2018. ISSN 1860-6768. doi: https://doi.org/10.1002/biot.201800226. URL https://doi.org/10.1002/biot.201800226.

72. Joseph R Brady, Melody C Tan, Charles A Whittaker, Noelle A Colant, Neil C Dalvie, Kerry Routenberg Love, and J Christopher Love. Identifying Improved Sites for Heterologous Gene Integration Using ATAC-seq. *ACS Synthetic Biology*, 9(9):2515–2524, 9 2020. doi: 10.1021/acssynbio.0c00299. URL https://doi.org/10.1021/acssynbio.0c00299.

73. Heena Dhiman, Marguerite Campbell, Michael Melcher, Kevin D Smith, and Nicole Borth. Predicting favorable landing pads for targeted integrations in Chinese hamster ovary cell lines by learning stability characteristics from random transgene integrations. *Computational and Structural Biotechnology Journal*, 18:3632–3648, 2020. ISSN 2001-0370. doi: https://doi.org/10.1016/j.csbj.2020.11.008. URL https://www.sciencedirect.com/science/article/pii/S200103702030475X.

74. Alexander Nash, Giulia Mignini Urdaneta, Andrea K Beaghton, Astrid Hoermann, Philippos Aris Papathanos, George K Christophides, and Nikolai Windbichler. Integral gene drives for population replacement. *Biology Open*, 8(1):bio037762, 1 2019. ISSN 2046-6390. doi: 10.1242/bio.037762. URL https://doi.org/10.1242/bio.037762.

75. Alexander J Weitzel, Hannah A Grunwald, Ceri Weber, Rimma Levina, Valentino M Gantz, Stephen M Hedrick, Ethan Bier, and Kimberly L Cooper. Meiotic Cas9 expression mediates gene conversion in the male and female mouse germline. *PLOS Biology*, 19(12):e3001478, 12 2021. ISSN 1545-7885. doi: 10.1371/journal.pbio.3001478. URL https://doi.org/10.1371/journal.pbio.3001478https://dx.plos.org/10.1371/journal.pbio.3001478.

76. Sebald AN Verkuijl and Marianne G. Rots. The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies. *Current Opinion in Biotechnology*, 55:68–73, 2019. ISSN 18790429. doi: 10.1016/j.copbio.2018.07.005.

77. C Monteilhet, A Perrin, A Thierry, L Colleaux, and B Dujon. Purification and characterization of the in vitro activity of I-Sce I, a novel and highly specific endonuclease encoded by a group I intron. *Nucleic acids research*, 18(6):1407–1413, 3 1990. ISSN 0305-1048. doi: 10.1093/nar/18.6.1407. URL https://pubmed.ncbi.nlm.nih.gov/2183191https://www.ncbi.nlm.nih.gov/pmc/articles/PMC330504/.

78. Yuk-Sang Chan, David S. Huen, Ruth Glauert, Eleanor Whiteway, and Steven Russell. Optimising Homing Endonuclease Gene Drive Performance in a Semi-Refractory Species: The Drosophila melanogaster Experience. *PLoS ONE*, 8(1):e54130, 1 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0054130. URL https://dx.plos.org/10.1371/journal.pone.0054130.

79. John M. Marshall, Anna Buchman, Héctor M. Sánchez C., and Omar S. Akbari. Overcoming evolved resistance to population-suppressing homing-based gene drives. *Scientific Reports*, 7(1):3776, 12 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-02744-7. URL http://www.nature.com/articles/s41598-017-02744-7.

80. Charleston Noble, Jason Olejarz, Kevin M. Esvelt, George M. Church, and Martin A. Nowak. Evolutionary dynamics of CRISPR gene drives. *Science Advances*, 3(4):e1601964, 4 2017. ISSN 2375-2548. doi: 10.1126/sciadv.1601964. URL https://advances.sciencemag.org/lookup/doi/10.1126/sciadv.1601964.

81. Jackson Champer, Jingxian Liu, Suh Yeon Oh, Riona Reeves, Anisha Luthra, Nathan Oakes, Andrew G. Clark, and Philipp W. Messer. Reducing resistance allele formation in CRISPR gene drive. *Proceedings of the National Academy of Sciences*, 115(21):5522–5527, 5 2018. ISSN 0027-8424. doi: 10.1073/pnas.1720354115. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1720354115.

82. Thai Binh Pham, Celine Hien Phong, Jared B. Bennett, Kristy Hwang, Nijole Jasinskiene, Kiona Parker, Drusilla Stillinger, John M. Marshall, Rebeca Carballar-Lejarazú, and Anthony A. James. Experimental population modification of the malaria vector mosquito, Anopheles stephensi. *PLOS Genetics*, 15(12):e1008440, 12 2019. ISSN 1553-7404. doi: 10.1371/journal.pgen.1008440. URL https://dx.plos.org/10.1371/journal.pgen.1008440.

83. Silke Fuchs, William T Garrood, Anna Beber, Andrew Hammond, Roberto Galizi, Matthew Gribble, Giulia Morselli, Tin-Yu J Hui, Katie Willis, Nace Kranjc, Austin Burt, Andrea Crisanti, and Tony Nolan. Resistance to a CRISPR-based gene drive at an evolutionarily conserved site is revealed by mimicking genotype fixation. *PLOS Genetics*, 17(10):e1009740, 10 2021. ISSN 1553-7404. doi: 10.1371/journal.pgen.1009740. URL http://biorxiv.org/content/early/2021/07/26/2021.07.26.453797.abstracthttps://dx.plos.org/10.1371/journal.pgen.1009740.

84. Samuel E. Champer, Nathan Oakes, Ronin Sharma, Pablo García-Díaz, Jackson Champer, and Philipp W Messer. Modeling CRISPR gene drives for suppression of invasive rodents using a supervised machine learning framework. *PLOS Computational Biology*, 17(12):e1009660, 12 2021. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1009660. URL https://doi.org/10.1371/journal.pcbi.1009660https://dx.plos.org/10.1371/journal.pcbi.1009660.

85. Nikolai Windbichler, Miriam Menichelli, Philippos Aris Papathanos, Summer B. Thyme, Hui Li, Umut Y. Ulge, Blake T. Hovde, David Baker, Raymond J. Monnat, Austin Burt, and Andrea Crisanti. A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature*, 473(7346):212–215, 5 2011. ISSN 0028-0836. doi: 10.1038/nature09937. URL http://www.nature.com/articles/nature09937.

86. Yuk-Sang Chan, Ryo Takeuchi, Jordan Jarjour, David S. Huen, Barry L. Stoddard, and Steven Russell. The Design and In Vivo Evaluation of Engineered I-OnuI-Based Enzymes for HEG Gene Drive. *PLoS ONE*, 8(9):e74254, 9 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0074254. URL https://dx.plos.org/10.1371/journal.pone.0074254.

87. Alekos Simoni, Carla Siniscalchi, Yuk-Sang Chan, David S. Huen, Steven Russell, Nikolai Windbichler, and Andrea Crisanti. Development of synthetic selfish elements based on modular nucleases in Drosophila melanogaster. *Nucleic Acids Research*, 42(11):7461–7472, 6 2014. ISSN 0305-1048. doi: 10.1093/nar/gku387. URL https://academic.oup.com/nar/article/42/11/7461/1452254.

88. Valentino M. Gantz and Ethan Bier. The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations. *Science*, 348(6233):442–444, 4 2015. ISSN 0036-8075. doi: 10.1126/science.aaa5945. URL http://www.sciencemag.org/cgi/doi/10.1126/science.aaa5945https://www.sciencemag.org/lookup/doi/10.1126/science.aaa5945.

89. Kevin M. Esvelt, Andrea L. Smidler, Flaminia Catteruccia, and George M. Church. Concerning RNA-guided gene drives for the alteration of wild populations. *eLife*, 3(July2014):1–21, 7 2014. ISSN 2050-084X. doi: 10.7554/eLife.03401. URL https://elifesciences.org/articles/03401.

90. Jackson Champer, Riona Reeves, Suh Yeon Oh, Chen Liu, Jingxian Liu, Andrew G. Clark, and Philipp W. Messer. 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, 7 2017. ISSN 1553-7404. doi: 10.1371/journal.pgen.1006796. URL https://doi.org/10.1371/journal.pgen.1006796https://dx.plos.org/10.1371/journal.pgen.1006796.

91. Xiang-ru Shannon Xu, Emily A Bulger, Valentino M Gantz, Carissa Klanseck, Stephanie R. Heimler, Ankush Auradkar, Jared B. Bennett, Lauren Ashley Miller, Sarah Leahy, Sara Sanz Juste, Anna Buchman, Omar S Akbari, John M Marshall, and Ethan Bier. Active Genetic Neutralizing Elements for Halting or Deleting Gene Drives. *Molecular Cell*, 80(2):246–262, 10 2020. ISSN 10972765. doi: 10.1016/j.molcel.2020.09.003. URL https://doi.org/10.1016/j.molcel.2020.09.003https://linkinghub.elsevier.com/retrieve/pii/S1097276520306110.

92. Andrew M. Hammond, Kyros Kyrou, Marco Bruttini, Ace North, Roberto Galizi, Xenia Karlsson, Nace Kranjc, Francesco M. Carpi, Romina D'Aurizio, Andrea Crisanti, and Tony Nolan. The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLOS Genetics*, 13(10):e1007039, 10 2017. ISSN 1553-7404. doi: 10.1371/journal.pgen.1007039. URL http://dx.plos.org/10.1371/journal.pgen.1007039https://dx.plos.org/10.1371/journal.pgen.1007039.

93. Hannah A. Grunwald, Valentino M. Gantz, Gunnar Poplawski, Xiang-Ru S. Xu, Ethan Bier, and Kimberly L. Cooper. Super-Mendelian inheritance mediated by CRISPR–Cas9 in the female mouse germline. *Nature*, 566(7742):105–109, 2 2019. ISSN 0028-0836. doi: 10.1038/s41586-019-0875-2. URL http://dx.doi.org/10.1038/s41586-019-0875-2http://www.nature.com/articles/s41586-019-0875-2.

94. Chandran Pfitzner, Melissa A. White, Sandra G. Piltz, Michaela Scherer, Fatwa Adikusuma, James N. Hughes, and Paul Q. Thomas. Progress Toward Zygotic and Germline Gene Drives in Mice. *The CRISPR Journal*, 3(5):388–397, 10 2020. ISSN 2573-1599. doi: 10.1089/crispr.2020.0050. URL https://www.liebertpub.com/doi/10.1089/crispr.2020.0050.

95. Ming Li, Ting Yang, Nikolay P Kandul, Michelle Bui, Stephanie Gamez, Robyn Raban, Jared Bennett, Héctor M Sánchez C, Gregory C Lanzaro, Hanno Schmidt, Yoosook Lee, John M Marshall, and Omar S Akbari. Development of a confinable gene drive system in the human disease vector Aedes aegypti. *eLife*, 9(May):1–40, 1 2020. ISSN 2050-084X. doi: 10.7554/eLife.51701. URL https://elifesciences.org/articles/51701.

96. William Reid, Adeline E Williams, Irma Sanchez-Vargas, Jingyi Lin, Rucsanda Juncu, Ken E Olson, and Alexander W E Franz. Assessing single-locus CRISPR/Cas9-based gene drive variants in the mosquito Aedes aegypti via single-generation crosses and modeling. *G3 Genes|Genomes|Genetics*, 12(12):2021.12.08.471839, 12 2022. ISSN 2160-1836. doi: 10.1093/g3journal/jkac280. URL https://academic.oup.com/g3journal/article/doi/10.1093/g3journal/jkac280/6762083.

97. Xuejiao Xu, Tim Harvey-Samuel, Hamid Anees Siddiqui, Joshua Xin De Ang, Michelle Ellis Anderson, Christine M. Reitmayer, Erica Lovett, Philip T Leftwich, Minsheng You, and Luke Alphey. Toward a CRISPR-Cas9-based Gene Drive in the Diamondback Moth Plutella xylostella. *The CRISPR Journal*, page 2021.10.05.462963, 3 2022. ISSN 2573-1599. doi: 10.1089/crispr.2021.0129. URL http://biorxiv.org/content/early/2021/10/05/2021.10.05.462963.abstracthttps://www.liebertpub.com/doi/10.1089/crispr.2021.0129.

98. Tao Zhang, Michael Mudgett, Ratnala Rambabu, Bradley Abramson, Xinhua Dai, Todd P Michael, and Yunde Zhao. Selective inheritance of target genes from only one parent of sexually reproduced F1 progeny in Arabidopsis. *Nature Communications*, 12(1):3854, 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-24195-5. URL https://doi.org/10.1038/s41467-021-24195-5.

99. Eli M. Carrami, Kolja N. Eckermann, Hassan M. M. Ahmed, Héctor M. Sánchez C., Stefan Dippel, John M. Marshall, and Ernst A. Wimmer. Consequences of resistance evolution in a Cas9-based sex conversion-suppression gene drive for insect pest management. *Proceedings of the National Academy of Sciences*, 115(24):6189–6194, 6 2018. ISSN 0027-8424. doi: 10.1073/pnas.1713825115. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1713825115https://pnas.org/doi/full/10.1073/pnas.1713825115.

100. Sarah E Allen, Gabriel T Koreman, Ankita Sarkar, Bei Wang, Mariana F Wolfner, and Chun Han. Versatile CRISPR/Cas9-mediated mosaic analysis by gRNA-induced crossing-over for unmodified genomes. *PLOS Biology*, 19(1):e3001061, 1 2021. URL https://doi.org/10.1371/journal.pbio.3001061.

101. Erich Brunner, Ryohei Yagi, Marc Debrunner, Dezirae Beck-Schneider, Alexa Burger, Eliane Escher, Christian Mosimann, George Hausmann, and Konrad Basler. CRISPR-induced double-strand breaks trigger recombination between homologous chromosome arms. *Life Science Alliance*, 2(3):e201800267, 6 2019. doi: 10.26508/lsa.201800267. URL http://www.life-science-alliance.org/content/2/3/e201800267.abstract.

102. Meru J Sadhu, Joshua S Bloom, Laura Day, and Leonid Kruglyak. CRISPR-directed mitotic recombination enables genetic mapping without crosses. *Science (New York, N.Y.)*, 352(6289):1113–1116, 5 2016. ISSN 1095-9203. doi: 10.1126/science.aaf5124. URL https://pubmed.ncbi.nlm.nih.gov/27230379https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4933295/.

103. Annabel Guichard, Tisha Haque, Marketta Bobik, Xiang-Ru S. Xu, Carissa Klanseck, Raja Babu Singh Kushwah, Mateus Berni, Bhagyashree Kaduskar, Valentino M. Gantz, and Ethan Bier. Efficient allelic-drive in Drosophila. *Nature Communications*, 10(1):1640, 12 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-09694-w. URL http://www.nature.com/articles/s41467-019-09694-w.

104. Tarun S Nambiar, Lou Baudrier, Pierre Billon, and Alberto Ciccia. CRISPR-based genome editing through the lens of DNA repair. *Molecular Cell*, 82(2):348–388, 2022. ISSN 1097-2765. doi: https://doi.org/10.1016/j.molcel.2021.12.026. URL https://www.sciencedirect.com/science/article/pii/S109727652101131Xhttps://doi.org/10.1016/j.molcel.2021.12.026.

105. Anna M Langmuller, Jackson Champer, Sandra Lapinska, Lin Xie, Matthew Metzloff, Samuel E Champer, Jingxian Liu, Yineng Xu, Jie Du, Andrew G Clark, and Philipp W Messer. Fitness effects of CRISPR endonucleases in Drosophila melanogaster populations. *eLife*, 11:2021.05.13.444039, 9 2022. ISSN 2050-084X. doi: 10.7554/eLife.71809. URL http://biorxiv.org/content/early/2021/05/13/2021.05.13.444039.abstracthttps://elifesciences.org/articles/71809.

106. Zhiqian Li, Nimi Marcel, Sushil Devkota, Ankush Auradkar, Stephen M Hedrick, Valentino M Gantz, and Ethan Bier. CopyCatchers are versatile active genetic elements that detect and quantify inter-homolog somatic gene conversion. *Nature Communications*, 12(1):2625, 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-22927-1. URL https://doi.org/10.1038/s41467-021-22927-1.

107. Xiang-Ru Shannon Xu, Valentino Matteo Gantz, Natalia Siomava, and Ethan Bier. CRISPR/Cas9 and active genetics-based trans-species replacement of the endogenous Drosophila kni-L2 CRM reveals unexpected complexity. *eLife*, 6:1–23, 12 2017. ISSN 2050-084X. doi: 10.7554/eLife.30281. URL https://elifesciences.org/articles/30281.

108. Gerard Terradas, Anita Hermann, Anthony A James, William McGinnis, and Ethan Bier. High-resolution in situ analysis of Cas9 germline transcript distributions in gene-drive Anopheles mosquitoes. *G3 Genes|Genomes|Genetics*, page jkab369, 11 2021. ISSN 2160-1836. doi: 10.1093/g3journal/jkab369. URL https://doi.org/10.1093/g3journal/jkab369.

109. Chrysanthi Taxiarchi, Andrea Beaghton, Nayomi Illansinhage Don, Kyros Kyrou, Matthew Gribble, Dammy Shittu, Scott P. Collins, Chase L. Beisel, Roberto Galizi, and Andrea Crisanti. A genetically encoded anti-CRISPR protein constrains gene drive spread and prevents population suppression. *Nature Communications*, 12(1):3977, 12 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-24214-5. URL http://dx.doi.org/10.1038/s41467-021-24214-5http://www.nature.com/articles/s41467-021-24214-5https://doi.org/10.1038/s41467-021-24214-5.

110. Samuel E. Champer, Suh Yeon Oh, Chen Liu, Zhaoxin Wen, Andrew G. Clark, Philipp W. Messer, and Jackson Champer. Computational and experimental performance of CRISPR homing gene drive strategies with multiplexed gRNAs. *Science Advances*, 6(10):eaaz0525, 3 2020. ISSN 2375-2548. doi: 10.1126/sciadv.aaz0525. URL https://advances.sciencemag.org/lookup/doi/10.1126/sciadv.aaz0525.

111. L C Kadyk and L H Hartwell. Sister chromatids are preferred over homologs as substrates for recombinational repair in Saccharomyces cerevisiae. *Genetics*, 132(2):387–402, 10 1992. ISSN 0016-6731 (Print). doi: 10.1093/genetics/132.2.387.

112. James E. Haber. TOPping Off Meiosis. *Molecular Cell*, 57(4):577–581, 2015. ISSN 1097-2765. doi: https://doi.org/10.1016/j.molcel.2015.02.004. URL https://www.sciencedirect.com/science/article/pii/S1097276515000969.

113. Andrea Enguita-Marruedo, Marta Martín-Ruiz, Eva García, Ana Gil-Fernández, María Teresa Parra, Alberto Viera, Julio S Rufas, and Jesús Page. Transition from a meiotic to a somatic-like DNA damage response during the pachytene stage in mouse meiosis. *PLoS genetics*, 15(1):e1007439, 1 2019. ISSN 1553-7404 (Electronic). doi: 10.1371/journal.pgen.1007439.

114. Víctor López Del Amo, Alena L. Bishop, Héctor M. Sánchez C., Jared B. Bennett, Xuechun Feng, John M. Marshall, Ethan Bier, and Valentino M. Gantz. A transcomplementing gene drive provides a flexible platform for laboratory investigation and potential field deployment. *Nature Communications*, 11(1):352, 12 2020. ISSN 2041-1723. doi: 10.1038/s41467-019-13977-7. URL http://www.nature.com/articles/s41467-019-13977-7.

115. Jackson Champer, Joan Chung, Yoo Lim Lee, Chen Liu, Emily Yang, Zhaoxin Wen, Andrew G. Clark, and Philipp W. Messer. Molecular safeguarding of CRISPR gene drive experiments. *eLife*, 8:1–10, 1 2019. ISSN 2050-084X. doi: 10.7554/eLife.41439. URL https://elifesciences.org/articles/41439.

116. Gerard Terradas, Jared B Bennett, Zhiqian Li, John M Marshall, and Ethan Bier. Genetic conversion of a split-drive into a full-drive element. *Nature Communications*, 14(1):191, 1 2023. ISSN 2041-1723. doi: 10.1038/s41467-022-35044-4. URL https://doi.org/10.1038/s41467-022-35044-4.

117. Ramon Y. Birnbaum, E. Josephine Clowney, Orly Agamy, Mee J. Kim, Jingjing Zhao, Takayuki Yamanaka, Zachary Pappalardo, Shoa L. Clarke, Aaron M. Wenger, Loan Nguyen, Fiorella Gurrieri, David B. Everman, Charles E. Schwartz, Ohad S. Birk, Gill Bejerano, Stavros Lomvardas, and Nadav Ahituv. Coding exons function as tissue-specific enhancers of nearby genes. *Genome Research*, 22(6):1059–1068, 6 2012. ISSN 1088-9051. doi: 10.1101/gr.133546.111. URL http://genome.cshlp.org/lookup/doi/10.1101/gr.133546.111.

118. Nikolay P Kandul, Junru Liu, Anna Buchman, Valentino M Gantz, Ethan Bier, and Omar S. Akbari. Assessment of a Split Homing Based Gene Drive for Efficient Knockout of Multiple Genes. *G3; Genes|Genomes|Genetics*, 10(2):827–837, 2 2020. ISSN 2160-1836. doi: 10.1534/g3.119.400985. URL http://g3journal.org/lookup/doi/10.1534/g3.119.400985.

119. Attilio Pane, Marco Salvemini, Pasquale Delli Bovi, Catello Polito, and Giuseppe Saccone. The transformer gene in Ceratitis capitata provides a genetic basis for selecting and remembering the sexual fate. *Development (Cambridge, England)*, 129(15):3715–3725, 8 2002. ISSN 0950-1991 (Print).

120. Sojung Kim, Daesik Kim, Seung Woo Cho, Jungeun Kim, and Jin-Soo Kim. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Research*, 24(6):1012–1019, 6 2014. ISSN 1088-9051. doi: 10.1101/gr.171322.113. URL http://genome.cshlp.org/lookup/doi/10.1101/gr.171322.113.

121. Alexa Burger, Helen Lindsay, Anastasia Felker, Christopher Hess, Carolin Anders, Elena Chiavacci, Jonas Zaugg, Lukas M Weber, Raul Catena, Martin Jinek, Mark D Robinson, and Christian Mosimann. Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. *Development (Cambridge, England)*, 143(11):2025–2037, 6 2016. ISSN 1477-9129 (Electronic). doi: 10.1242/dev.134809.

122. Volker Hartenstein. *Atlas of Drosophila Development*. Cold Spring Harbor Laboratory Press, 1993. ISBN 978-0879694722. URL http://www.sdbonline.org/sites/fly/atlas/1819.htm.

123. Ayal Hendel, Rasmus O Bak, Joseph T Clark, Andrew B Kennedy, Daniel E Ryan, Subhadeep Roy, Israel Steinfeld, Benjamin D Lunstad, Robert J Kaiser, Alec B Wilkens, Rosa Bacchetta, Anya Tsalenko, Douglas Dellinger, Laurakay Bruhn, and Matthew H Porteus. Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nature Biotechnology*, 33(9):985–989, 2015. ISSN 1546-1696. doi: 10.1038/nbt.3290. URL https://doi.org/10.1038/nbt.3290.

124. Hanhui Ma, Li-Chun Tu, Ardalan Naseri, Maximiliaan Huisman, Shaojie Zhang, David Grunwald, and Thoru Pederson. CRISPR-Cas9 nuclear dynamics and target recognition in living cells. *Journal of Cell Biology*, 214(5):529–537, 8 2016. ISSN 0021-9525. doi: 10.1083/jcb.201604115. URL https://rupress.org/jcb/article/214/5/529/38728/CRISPRCas9-nuclear-dynamics-and-target-recognition.

125. Haifeng Wang, Muneaki Nakamura, Timothy R Abbott, Dehua Zhao, Kaiwen Luo, Cordelia Yu, Cindy M Nguyen, Albert Lo, Timothy P Daley, Marie La Russa, Yanxia Liu, and Lei S Qi. CRISPR-mediated live imaging of genome editing and transcription. *Science*, 365(6459):1301–1305, 9 2019. ISSN 0036-8075. doi: 10.1126/science.aax7852. URL http://www.ncbi.nlm.nih.gov/pubmed/31488703https://www.science.org/doi/10.1126/science.aax7852.

126. Andrea K. Beaghton, Andrew Hammond, Tony Nolan, Andrea Crisanti, and Austin Burt. Gene drive for population genetic control: non-functional resistance and parental effects. *Proceedings of the Royal Society B: Biological Sciences*, 286(1914):20191586, 11 2019. ISSN 0962-8452. doi: 10.1098/rspb.2019.1586. URL https://royalsocietypublishing.org/doi/10.1098/rspb.2019.1586.

127. Felicity Allen, Luca Crepaldi, Clara Alsinet, Alexander J. Strong, Vitalii Kleshchevnikov, Pietro De Angeli, Petra Páleníková, Anton Khodak, Vladimir Kiselev, Michael Kosicki, Andrew R. Bassett, Heather Harding, Yaron Galanty, Francisco Muñoz-Martínez, Emmanouil Metzakopian, Stephen P. Jackson, and Leopold Parts. Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nature Biotechnology*, 37(1):64–72, 1 2019. ISSN 1087-0156. doi: 10.1038/nbt.4317. URL https://www.nature.com/articles/nbt.4317.

128. Anob M. Chakrabarti, Tristan Henser-Brownhill, Josep Monserrat, Anna R. Poetsch, Nicholas M. Luscombe, and Paola Scaffidi. Target-Specific Precision of CRISPR-Mediated Genome Editing. *Molecular Cell*, 73(4):699–713, 2 2019. ISSN 10972765. doi: 10.1016/j.molcel.2018.11.031. URL https://doi.org/10.1016/j.molcel.2018.11.031https://linkinghub.elsevier.com/retrieve/pii/S1097276518310013.

129. Andrew Hammond, Xenia Karlsson, Ioanna Morianou, Kyros Kyrou, Andrea Beaghton, Matthew Gribble, Nace Kranjc, Roberto Galizi, Austin Burt, Andrea Crisanti, and Tony Nolan. Regulating the expression of gene drives is key to increasing their invasive potential and the mitigation of resistance. *PLOS Genetics*, 17(1):e1009321, 1 2021. ISSN 1553-7404. doi: 10.1371/journal.pgen.1009321. URL https://dx.plos.org/10.1371/journal.pgen.1009321https://doi.org/10.1371/journal.pgen.1009321.

130. Jackson Champer, Zhaoxin Wen, Anisha Luthra, Riona Reeves, Joan Chung, Chen Liu, Yoo Lim Lee, Jingxian Liu, Emily Yang, Philipp W Messer, and Andrew G Clark. CRISPR Gene Drive Efficiency and Resistance Rate Is Highly Heritable with No Common Genetic Loci of Large Effect. *Genetics*, 212(1):333–341, 5 2019. ISSN 1943-2631. doi: 10.1534/genetics.119.302037. URL https://doi.org/10.1534/genetics.119.302037.

131. Ana L Ramírez. Aedes aegypti (blood-fed), 2 2019. URL https://figshare.com/articles/figure/Aedes_aegypti_blood-fed_/7755161https://ndownloader.figshare.com/files/14434553.

132. Ana L Ramírez. (male) Aedes aegypti, 2 2019. URL https://figshare.com/articles/figure/_male_Aedes_aegypti/7699778https://ndownloader.figshare.com/files/14329748.

133. Nikolay P. Kandul, Junru Liu, Hector M. Sanchez C., Sean L. Wu, John M. Marshall, and Omar S. Akbari. Transforming insect population control with precision guided sterile males with demonstration in flies. *Nature Communications*, 10 (1):84, 12 2019. ISSN 2041-1723. doi: 10.1038/s41467-018-07964-7. URL http://dx.doi.org/10.1038/s41467-018-07964-7http://www.nature.com/articles/s41467-018-07964-7.

134. Philip A Eckhoff, Edward A Wenger, H Charles J Godfray, and Austin Burt. 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*, 114(2):E255 LP – E264, 1 2017. doi: 10.1073/pnas.1611064114. URL http://www.pnas.org/content/114/2/E255.abstract.

135. Ace R North, Austin Burt, and H Charles J Godfray. Modelling the suppression of a malaria vector using a CRISPR-Cas9 gene drive to reduce female fertility. *BMC Biology*, 18(1):98, 2020. ISSN 1741-7007. doi: 10.1186/s12915-020-00834-z. URL https://doi.org/10.1186/s12915-020-00834-z.

136. Jackson Champer, Isabel K Kim, Samuel E Champer, Andrew G Clark, and Philipp W Messer. Suppression gene drive in continuous space can result in unstable persistence of both drive and wild-type alleles. *Molecular ecology*, 30(4):1086–1101, 2 2021. ISSN 1365-294X (Electronic). doi: 10.1111/mec.15788.

137. Cathy Salles, Maryvonne Mével-Ninio, Alain Vincent, and François Payre. A Germline-Specific Splicing Generates an Extended Ovo Protein Isoform Required for Drosophila Oogenesis. *Developmental Biology*, 246(2):366–376, 6 2002. ISSN 00121606. doi: 10.1006/dbio.2002.0659. URL https://linkinghub.elsevier.com/retrieve/pii/S0012160602906597.

138. Hitoshi Tsujimoto, Kun Liu, Paul J. Linser, Peter Agre, and Jason L. Rasgon. Organ-Specific Splice Variants of Aquaporin Water Channel AgAQP1 in the Malaria Vector Anopheles gambiae. *PLoS ONE*, 8(9):e75888, 9 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0075888. URL https://dx.plos.org/10.1371/journal.pone.0075888.

139. Elizabeth R. Sutton, Yachuan Yu, Sebastian M. Shimeld, Helen White-Cooper, and Luke Alphey. Identification of genes for engineering the male germline of Aedes aegypti and Ceratitis capitata. *BMC Genomics*, 17(1):948, 12 2016. ISSN 1471-2164. doi: 10.1186/s12864-016-3280-3. URL http://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-3280-3.

140. Hélène Chassin, Marius Müller, Marcel Tigges, Leo Scheller, Moritz Lang, and Martin Fussenegger. A modular degron library for synthetic circuits in mammalian cells. *Nature Communications*, 10(1):2013, 12 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-09974-5. URL http://www.nature.com/articles/s41467-019-09974-5.

141. Megan E. Goeckel, Erianna M. Basgall, Isabel C. Lewis, Samantha C. Goetting, Yao Yan, Megan Halloran, and Gregory C. Finnigan. Modulating CRISPR gene drive activity through nucleocytoplasmic localization of Cas9 in S. cerevisiae. *Fungal Biology and Biotechnology*, 6(1):2, 12 2019. ISSN 2054-3085. doi: 10.1186/s40694-019-0065-x. URL https://doi.org/10.1186/s40694-019-0065-xhttps://fungalbiolbiotech.biomedcentral.com/articles/10.1186/s40694-019-0065-x.

142. Carlos M Loya, Cecilia S Lu, David Van Vactor, and Tudor A Fulga. Transgenic microRNA inhibition with spatiotemporal specificity in intact organisms. *Nature Methods*, 6(12):897–903, 2009. ISSN 1548-7105. doi: 10.1038/nmeth.1402. URL https://doi.org/10.1038/nmeth.1402.

143. Janice A Fischer, Edward Giniger, Tom Maniatis, and Mark Ptashne. GAL4 activates transcription in Drosophila. *Nature*, 332 (6167):853–856, 1988. ISSN 1476-4687. doi: 10.1038/332853a0. URL https://doi.org/10.1038/332853a0.

144. Astrid Hoermann, Sofia Tapanelli, Paolo Capriotti, Giuseppe Del Corsano, Ellen K G Masters, Tibebu Habtewold, George K Christophides, and Nikolai Windbichler. Converting endogenous genes of the malaria mosquito into simple non-autonomous gene drives for population replacement. *eLife*, 10:e58791, 2021. ISSN 2050-084X. doi: 10.7554/eLife.58791. URL https://doi.org/10.7554/eLife.58791.

145. Nikolai Windbichler, Philippos Aris Papathanos, and Andrea Crisanti. 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, 12 2008. ISSN 1553-7404. doi: 10.1371/journal.pgen.1000291. URL https://doi.org/10.1371/journal.pgen.1000291https://dx.plos.org/10.1371/journal.pgen.1000291.

146. Roberto Galizi, Lindsey A. Doyle, Miriam Menichelli, Federica Bernardini, Anne Deredec, Austin Burt, Barry L. Stoddard, Nikolai Windbichler, and Andrea Crisanti. A synthetic sex ratio distortion system for the control of the human malaria mosquito. *Nature Communications*, 5(1):3977, 9 2014. ISSN 2041-1723. doi: 10.1038/ncomms4977. URL http://www.nature.com/articles/ncomms4977.

147. Alekos Simoni, Andrew M Hammond, Andrea K Beaghton, Roberto Galizi, Chrysanthi Taxiarchi, Kyros Kyrou, Dario Meacci, Matthew Gribble, Giulia Morselli, Austin Burt, Tony Nolan, and Andrea Crisanti. A male-biased sex-distorter gene drive for the human malaria vector Anopheles gambiae. *Nature Biotechnology*, 38(9):1054–1060, 2020. ISSN 1546-1696. doi: 10.1038/s41587-020-0508-1. URL https://doi.org/10.1038/s41587-020-0508-1.

148. Thomas A A Prowse, Phillip Cassey, Joshua V Ross, Chandran Pfitzner, Talia A Wittmann, and Paul Thomas. Dodging silver bullets: good CRISPR gene-drive design is critical for eradicating exotic vertebrates. *Proceedings of the Royal Society B: Biological Sciences*, 284(1860):20170799, 8 2017. doi: 10.1098/rspb.2017.0799. URL https://doi.org/10.1098/rspb.2017.0799.

149. Matthew P Edgington, Tim Harvey-Samuel, and Luke Alphey. Population-level multiplexing: A promising strategy to manage the evolution of resistance against gene drives targeting a neutral locus. *Evolutionary Applications*, 13(8):1939–1948, 9 2020. ISSN 1752-4571. doi: https://doi.org/10.1111/eva.12945. URL https://doi.org/10.1111/eva.12945.

150. Sijie Liu and Daochun Kong. End resection: a key step in homologous recombination and DNA double-strand break repair. *Genome Instability & Disease*, 2(1):39–50, 2 2021. ISSN 2524-7662. doi: 10.1007/s42764-020-00028-5. URL https://doi.org/10.1007/s42764-020-00028-5http://link.springer.com/10.1007/s42764-020-00028-5.

151. Emily Yang, Matthew Metzloff, Anna M Langmüller, Xuejiao Xu, Andrew G Clark, Philipp W Messer, and Jackson Champer. A homing suppression gene drive with multiplexed gRNAs maintains high drive conversion efficiency and avoids functional resistance alleles. *G3 Genes|Genomes|Genetics*, 12(6):2021.05.27.446071, 5 2022. ISSN 2160-1836. doi: 10.1093/g3journal/jkac081. URL https://academic.oup.com/g3journal/article/doi/10.1093/g3journal/jkac081/6565321.

152. Víctor López Del Amo, Sara Sanz Juste, and Valentino M Gantz. A nickase Cas9 gene-drive system promotes super-Mendelian inheritance in Drosophila. *Cell Reports*, 39(8):110843, 5 2022. ISSN 22111247. doi: 10.1016/j.celrep.2022.110843. URL http://biorxiv.org/content/early/2021/12/02/2021.12.01.470847.abstracthttps://linkinghub.elsevier.com/retrieve/pii/S2211124722006167.

153. Víctor López Del Amo, Brittany S. Leger, Kurt J. Cox, Shubhroz Gill, Alena L. Bishop, Garrett D. Scanlon, James A. Walker, Valentino M. Gantz, and Amit Choudhary. Small-Molecule Control of Super-Mendelian Inheritance in Gene Drives. *Cell Reports*, 31(13):107841, 6 2020. ISSN 22111247. doi: 10.1016/j.celrep.2020.107841. URL https://www.biorxiv.org/content/10.1101/665620v1.fullhttps://linkinghub.elsevier.com/retrieve/pii/S2211124720308226.

154. Fillip Port and Simon L Bullock. Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs. *Nature Methods*, 13(10):852–854, 2016. ISSN 1548-7105. doi: 10.1038/nmeth.3972. URL https://doi.org/10.1038/nmeth.3972.

155. David J H F Knapp, Yale S Michaels, Max Jamilly, Quentin R V Ferry, Hector Barbosa, Thomas A Milne, and Tudor A Fulga. Decoupling tRNA promoter and processing activities enables specific Pol-II Cas9 guide RNA expression. *Nature Communications*, 10(1):1490, 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-09148-3. URL https://doi.org/10.1038/s41467-019-09148-3.

156. Michelle A. E. Anderson, Jessica Purcell, Sebald A. N. Verkuijl, Victoria C. Norman, Philip T. Leftwich, Tim Harvey-Samuel, and Luke S. Alphey. Expanding the CRISPR Toolbox in Culicine Mosquitoes: In Vitro Validation of Pol III Promoters. *ACS Synthetic Biology*, 9(3):678–681, 3 2020. ISSN 2161-5063. doi: 10.1021/acssynbio.9b00436. URL https://pubs.acs.org/doi/10.1021/acssynbio.9b00436.

157. Samuel H Sternberg, Sy Redding, Martin Jinek, Eric C Greene, and Jennifer A Doudna. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 507(7490):62–67, 2014. ISSN 1476-4687. doi: 10.1038/nature13011. URL https://doi.org/10.1038/nature13011.

158. Eva K. Brinkman, Tao Chen, Marcel de Haas, Hanna A. Holland, Waseem Akhtar, and Bas van Steensel. Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks. *Molecular Cell*, 70(5):801–813, 2018. ISSN 10974164. doi: 10.1016/j.molcel.2018.04.016. URL https://doi.org/10.1016/j.molcel.2018.04.016.

159. Fatwa Adikusuma, Sandra Piltz, Mark A Corbett, Michelle Turvey, Shaun R McColl, Karla J Helbig, Michael R Beard, James Hughes, Richard T Pomerantz, and Paul Q Thomas. Large deletions induced by Cas9 cleavage. *Nature*, 560(7717):E8–E9, 2018. ISSN 1476-4687. doi: 10.1038/s41586-018-0380-z. URL https://doi.org/10.1038/s41586-018-0380-z.

160. Andrew Hammond, Paola Pollegioni, Tania Persampieri, Ace North, Roxana Minuz, Alessandro Trusso, Alessandro Bucci, Kyros Kyrou, Ioanna Morianou, Alekos Simoni, Tony Nolan, Ruth Müller, and Andrea Crisanti. Gene-drive suppression of mosquito populations in large cages as a bridge between lab and field. *Nature Communications*, 12(1):4589, 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-24790-6. URL https://doi.org/10.1038/s41467-021-24790-6.

161. Adriana Adolfi, Valentino M Gantz, Nijole Jasinskiene, Hsu-Feng Lee, Kristy Hwang, Gerard Terradas, Emily A Bulger, Arunachalam Ramaiah, Jared B Bennett, J J Emerson, John M Marshall, Ethan Bier, and Anthony A James. Efficient population modification gene-drive rescue system in the malaria mosquito Anopheles stephensi. *Nature Communications*, 11(1):5553, 12 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-19426-0. URL http://biorxiv.org/content/early/2020/08/02/2020.08.02.233056.abstracthttp://www.nature.com/articles/s41467-020-19426-0.

162. Georg Oberhofer, Tobin Ivy, and Bruce A. Hay. Cleave and Rescue, a novel selfish genetic element and general strategy for gene drive. *Proceedings of the National Academy of Sciences*, 116(13):6250–6259, 3 2019. ISSN 0027-8424. doi: 10.1073/pnas.1816928116. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1816928116.

163. Jackson Champer, Emily Yang, Esther Lee, Jingxian Liu, Andrew G. Clark, and Philipp W. Messer. A CRISPR homing gene drive targeting a haplolethal gene removes resistance alleles and successfully spreads through a cage population. *Proceedings of the National Academy of Sciences*, 117(39):24377–24383, 9 2020. ISSN 0027-8424. doi: 10.1073/pnas.2004373117. URL https://www.biorxiv.org/content/10.1101/651737v1http://www.pnas.org/lookup/doi/10.1073/pnas.2004373117.

164. Ming Li, Michelle Bui, Ting Yang, Christian S. Bowman, Bradley J. White, and Omar S. Akbari. Germline Cas9 expression yields highly efficient genome engineering in a major worldwide disease vector, Aedes aegypti. *Proceedings of the National Academy of Sciences*, 114(49):E10540–E10549, 12 2017. ISSN 0027-8424. doi: 10.1073/pnas.1711538114. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1711538114.

165. Dongwoo Chae, Junwon Lee, Nayoung Lee, Kyungsoo Park, Seok Jun Moon, and Hyongbum Henry Kim. Chemical Controllable Gene Drive in Drosophila. *ACS Synthetic Biology*, 9(9):2362–2377, 9 2020. ISSN 2161-5063. doi: 10.1021/acssynbio.0c00117. URL https://pubs.acs.org/doi/10.1021/acssynbio.0c00117.

166. Aaron McKenna, Gregory M. Findlay, James A. Gagnon, Marshall S. Horwitz, Alexander F. Schier, and Jay Shendure. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science*, 353(6298):aaf7907, 7 2016. ISSN 10959203. doi: 10.1126/science.aaf7907. URL http://science.sciencemag.org/content/353/6298/aaf7907.abstract.

167. Shdema Filler-Hayut, Kiril Kniazev, Cathy Melamed-Bessudo, and Avraham A Levy. Targeted Inter-Homologs Recombination in Arabidopsis Euchromatin and Heterochromatin. *International Journal of Molecular Sciences*, 22(22), 2021. doi: https://doi.org/10.3390/ijms222212096.

168. Bernd Giese, Arnim von Gleich, and Johannes L Frieß. Alternative Techniques and Options for Risk Reduction of Gene Drives. In Arnim von Gleich and Winfried Schröder, editors, *Gene Drives at Tipping Points: Precautionary Technology Assessment and Governance of New Approaches to Genetically Modify Animal and Plant Populations*, pages 167–185. Springer International Publishing, Cham, 2020. ISBN 978-3-030-38934-5. doi: 10.1007/978-3-030-38934-5{_}7. URL https://doi.org/10.1007/978-3-030-38934-5_7.

169. Georg Oberhofer, Tobin Ivy, and Bruce A Hay. Split versions of Cleave and Rescue selfish genetic elements for measured self limiting gene drive. *PLOS Genetics*, 17(2):e1009385, 2 2021. ISSN 1553-7404. doi: 10.1371/journal.pgen.1009385. URL https://doi.org/10.1371/journal.pgen.1009385https://dx.plos.org/10.1371/journal.pgen.1009385.

170. Rebecca S. Shapiro, Alejandro Chavez, Caroline B. M. Porter, Meagan Hamblin, Christian S. Kaas, James E. DiCarlo, Guisheng Zeng, Xiaoli Xu, Alexey V. Revtovich, Natalia V. Kirienko, Yue Wang, George M. Church, and James J. Collins. A CRISPR–Cas9-based gene drive platform for genetic interaction analysis in Candida albicans. *Nature Microbiology*, 3(1):73–82, 1 2018. ISSN 2058-5276. doi: 10.1038/s41564-017-0043-0. URL http://www.nature.com/articles/s41564-017-0043-0.

171. Robert L. Unckless, Andrew G. Clark, and Philipp W. Messer. Evolution of Resistance Against CRISPR/Cas9 Gene Drive. *Genetics*, 205(2):827–841, 2 2017. ISSN 1943-2631. doi: 10.1534/genetics.116.197285. URL http://www.genetics.org/lookup/doi/10.1534/genetics.116.197285https://academic.oup.com/genetics/article/205/2/827/6066440.

172. Michelle A E Anderson, Estela Gonzalez, Matthew P Edgington, Joshua Xin De Ang, Deepak-Kumar Purusothaman, Lewis Shackleford, Katherine Nevard, Sebald A N Verkuijl, Tim Harvey-Samuel, Philip T Leftwich, Kevin Esvelt, and Luke Alphey. A multiplexed, confinable CRISPR/Cas9 gene drive propagates in caged Aedes aegypti populations. *bioRxiv*, page 2022.08.12.503466, 1 2022. doi: 10.1101/2022.08.12.503466. URL http://biorxiv.org/content/early/2022/08/12/2022.08.12.503466.abstract.

173. Richard E Grewelle, Javier Perez-Saez, Josh Tycko, Erica K O Namigai, Chloe G Rickards, and Giulio A De Leo. Modeling the efficacy of CRISPR gene drive for snail immunity on schistosomiasis control. *PLOS Neglected Tropical Diseases*, 16(10): e0010894, 10 2022. ISSN 1935-2735. doi: 10.1371/journal.pntd.0010894. URL https://dx.plos.org/10.1371/journal.pntd.0010894.

174. Frederik J H de Haas and Sarah P Otto. Gene Drives Across Engineered Fitness Valleys: Modeling a Design to Prevent Drive Spillover. *bioRxiv*, page 2020.10.29.360404, 1 2020. doi: 10.1101/2020.10.29.360404. URL http://biorxiv.org/content/early/2020/10/29/2020.10.29.360404.abstract.

175. Dieter Egli, Michael V Zuccaro, Michael Kosicki, George M Church, Allan Bradley, and Maria Jasin. Inter-homologue repair in fertilized human eggs? *Nature*, 560(7717):E5–E7, 2018. ISSN 1476-4687. doi: 10.1038/s41586-018-0379-5. URL https://doi.org/10.1038/s41586-018-0379-5.

176. Michael V. Zuccaro, Jia Xu, Carl Mitchell, Diego Marin, Raymond Zimmerman, Bhavini Rana, Everett Weinstein, Rebeca T. King, Katherine L. Palmerola, Morgan E. Smith, Stephen H. Tsang, Robin Goland, Maria Jasin, Rogerio Lobo, Nathan Treff, and Dieter Egli. Allele-Specific Chromosome Removal after Cas9 Cleavage in Human Embryos. *Cell*, 183(6):1650–1664, 12 2020. ISSN 1097-4172 (Electronic). doi: 10.1016/j.cell.2020.10.025. URL http://dx.doi.org/10.1016/j.cell.2020.10.025https://linkinghub.elsevier.com/retrieve/pii/S0092867420313891.

177. Michael R Vella, Christian E Gunning, Alun L Lloyd, and Fred Gould. Evaluating strategies for reversing CRISPR-Cas9 gene drives. *Scientific Reports*, 7(1):11038, 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-10633-2. URL https://doi.org/10.1038/s41598-017-10633-2.

178. R Torres, M C Martin, A Garcia, Juan C Cigudosa, J C Ramirez, and S Rodriguez-Perales. Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR–Cas9 system. *Nature Communications*, 5(1):3964, 2014. ISSN 2041-1723. doi: 10.1038/ncomms4964. URL https://doi.org/10.1038/ncomms4964.

179. Danilo Maddalo, Eusebio Manchado, Carla P Concepcion, Ciro Bonetti, Joana A Vidigal, Yoon-Chi Han, Paul Ogrodowski, Alessandra Crippa, Natasha Rekhtman, Elisa de Stanchina, Scott W Lowe, and Andrea Ventura. In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature*, 516(7531):423–427, 2014. ISSN 1476-4687. doi: 10.1038/nature13902. URL https://doi.org/10.1038/nature13902.

180. Roman Maresch, Sebastian Mueller, Christian Veltkamp, Rupert Öllinger, Mathias Friedrich, Irina Heid, Katja Steiger, Julia Weber, Thomas Engleitner, Maxim Barenboim, Sabine Klein, Sandra Louzada, Ruby Banerjee, Alexander Strong, Teresa Stauber, Nina Gross, Ulf Geumann, Sebastian Lange, Marc Ringelhan, Ignacio Varela, Kristian Unger, Fengtang Yang, Roland M Schmid, George S Vassiliou, Rickmer Braren, Günter Schneider, Mathias Heikenwalder, Allan Bradley, Dieter Saur, and Roland Rad. Multiplexed pancreatic genome engineering and cancer induction by transfection-based CRISPR/Cas9 delivery in mice. *Nature Communications*, 7(1):10770, 2016. ISSN 2041-1723. doi: 10.1038/ncomms10770. URL https://doi.org/10.1038/ncomms10770.

181. Erika Brunet, Deniz Simsek, Mark Tomishima, Russell DeKelver, Vivian M Choi, Philip Gregory, Fyodor Urnov, David M Weinstock, and Maria Jasin. Chromosomal translocations induced at specified loci in human stem cells. *Proceedings of the National Academy of Sciences*, 106(26):10620 LP – 10625, 6 2009. doi: 10.1073/pnas.0902076106. URL http://www.pnas.org/content/106/26/10620.abstract.

182. Irina V Lagutina, Virginia Valentine, Fabrizio Picchione, Frank Harwood, Marcus B Valentine, Barbara Villarejo-Balcells, Jaime J Carvajal, and Gerard C Grosveld. Modeling of the Human Alveolar Rhabdomyosarcoma Pax3-Foxo1 Chromosome Translocation in Mouse Myoblasts Using CRISPR-Cas9 Nuclease. *PLOS Genetics*, 11(2):e1004951, 2 2015. URL https://doi.org/10.1371/journal.pgen.1004951.

183. Yao Yan and Gregory C. Finnigan. Development of a multi-locus CRISPR gene drive system in budding yeast. *Scientific Reports*, 8(1):17277, 12 2018. ISSN 2045-2322. doi: 10.1038/s41598-018-34909-3. URL http://dx.doi.org/10.1038/s41598-018-34909-3http://www.nature.com/articles/s41598-018-34909-3.

184. Arthur R Gorter de Vries, Lucas G F Couwenberg, Marcel van den Broek, Pilar de la Torre Cortés, Jolanda ter Horst, Jack T. Pronk, and Jean-Marc G Daran. Allele-specific genome editing using CRISPR–Cas9 is associated with loss of heterozygosity in diploid yeast. *Nucleic Acids Research*, 47(3):1362–1372, 2 2019. ISSN 0305-1048. doi: 10.1093/nar/gky1216. URL https://academic.oup.com/nar/article/47/3/1362/5230954.

185. Dan Wang, Jia Li, Chun-Qing Song, Karen Tran, Haiwei Mou, Pei-Hsuan Wu, Phillip W L Tai, Craig A Mendonca, Lingzhi Ren, Blake Y Wang, Qin Su, Dominic J Gessler, Phillip D Zamore, Wen Xue, and Guangping Gao. Cas9-mediated allelic exchange repairs compound heterozygous recessive mutations in mice. *Nature Biotechnology*, 36(9):839–842, 2018. ISSN 1546-1696. doi: 10.1038/nbt.4219. URL https://doi.org/10.1038/nbt.4219.

186. Erik Wijnker, Geo Velikkakam James, Jia Ding, Frank Becker, Jonas R Klasen, Vimal Rawat, Beth A Rowan, Daniël F de Jong, C Bastiaan de Snoo, Luis Zapata, Bruno Huettel, Hans de Jong, Stephan Ossowski, Detlef Weigel, Maarten Koornneef, Joost J B Keurentjes, and Korbinian Schneeberger. The genomic landscape of meiotic crossovers and gene conversions in Arabidopsis thaliana. *eLife*, 2:e01426, 2013. ISSN 2050-084X. doi: 10.7554/eLife.01426. URL https://doi.org/10.7554/eLife.01426.

187. Agnès Bergerat, Bernard de Massy, Danielle Gadelle, Paul-Christophe Varoutas, Alain Nicolas, and Patrick Forterre. An atypical topoisomerase II from archaea with implications for meiotic recombination. *Nature*, 386(6623):414–417, 1997. ISSN 1476-4687. doi: 10.1038/386414a0. URL https://doi.org/10.1038/386414a0.

188. S Keeney, C N Giroux, and N Kleckner. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell*, 88(3):375–384, 2 1997. ISSN 0092-8674 (Print). doi: 10.1016/s0092-8674(00)81876-0.

189. Ethan Bier. Gene drives gaining speed. *Nature Reviews Genetics*, 0123456789, 2021. ISSN 1471-0064. doi: 10.1038/s41576-021-00386-0. URL http://dx.doi.org/10.1038/s41576-021-00386-0https://doi.org/10.1038/s41576-021-00386-0.

190. Christine R Preston, Carlos C Flores, and William R Engels. Differential Usage of Alternative Pathways of Double-Strand Break Repair in Drosophila. *Genetics*, 172(2):1055–1068, 2 2006. ISSN 1943-2631. doi: 10.1534/genetics.105.050138. URL https://doi.org/10.1534/genetics.105.050138.

191. Alena L Bishop, Víctor López Del Amo, Emily M Okamoto, Zsolt Bodai, Alexis C Komor, and Valentino M Gantz. Double-tap gene drive uses iterative genome targeting to help overcome resistance alleles. *Nature Communications*, 13(1):2595, 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-29868-3. URL https://doi.org/10.1038/s41467-022-29868-3.

192. Johannes L Frieß, Carina R Lalyer, Bernd Giese, Samson Simon, and Mathias Otto. Review of gene drive modelling and implications for risk assessment of gene drive organisms. *Ecological Modelling*, 478:110285, 2023. ISSN 0304-3800. doi: https://doi.org/10.1016/j.ecolmodel.2023.110285. URL https://www.sciencedirect.com/science/article/pii/S0304380023000133.

193. Charleston Noble, Ben Adlam, George M Church, Kevin M Esvelt, and Martin A Nowak. Current CRISPR gene drive systems are likely to be highly invasive in wild populations. *eLife*, 7:e33423, 6 2018. ISSN 2050-084X. doi: 10.7554/eLife.33423. URL https://doi.org/10.7554/eLife.33423.

194. Chun-Chieh Lin and Christopher J Potter. Non-Mendelian Dominant Maternal Effects Caused by CRISPR/Cas9 Transgenic Components in Drosophila melanogaster. *G3 Genes|Genomes|Genetics*, 6(11):3685–3691, 11 2016. ISSN 2160-1836. doi: 10.1534/g3.116.034884. URL https://doi.org/10.1534/g3.116.034884.

195. Shdema Filler Hayut, Cathy Melamed Bessudo, and Avraham A. Levy. Targeted recombination between homologous chromosomes for precise breeding in tomato. *Nature Communications*, 8(1):15605, 8 2017. ISSN 2041-1723. doi: 10.1038/ncomms15605. URL https://doi.org/10.1038/ncomms15605http://dx.doi.org/10.1038/ncomms15605http://www.nature.com/articles/ncomms15605.

196. Svenia D. Heinze, Tea Kohlbrenner, Domenica Ippolito, Angela Meccariello, Alexa Burger, Christian Mosimann, Giuseppe Saccone, and Daniel Bopp. CRISPR-Cas9 targeted disruption of the yellow ortholog in the housefly identifies the brown body locus. *Scientific Reports*, 7(1):1–9, 2017. ISSN 20452322. doi: 10.1038/s41598-017-04686-6.

197. Dena M Johnson-Schlitz, Carlos Flores, and William R Engels. Multiple-Pathway Analysis of Double-Strand Break Repair Mutations in Drosophila. *PLOS Genetics*, 3(4):e50, 4 2007. URL https://doi.org/10.1371/journal.pgen.0030050.

198. Dena Johnson-Schlitz and William R Engels. Template disruptions and failure of double Holliday junction dissolution during double-strand break repair in Drosophila BLM mutants. *Proceedings of the National Academy of Sciences*, 103(45):16840–16845, 11 2006. doi: 10.1073/pnas.0607904103. URL https://doi.org/10.1073/pnas.0607904103.

199. Yikang S. Rong and Kent G. Golic. The Homologous Chromosome Is an Effective Template for the Repair of Mitotic DNA Double-Strand Breaks in Drosophila. *Genetics*, 165(4):1831–1842, 12 2003. ISSN 1943-2631. doi: 10.1093/genetics/165.4.1831. URL https://academic.oup.com/genetics/article/165/4/1831/6050473.

200. Sitara Roy, Sara Sanz Juste, Marketta Sneider, Ankush Auradkar, Carissa Klanseck, Zhiqian Li, Alison Henrique Ferreira Julio, Victor Lopez del Amo, Ethan Bier, and Annabel Guichard. Cas9/Nickase-induced allelic conversion by homologous chromosome-templated repair in Drosophila somatic cells. *Science Advances*, 8(26):eabo0721, 7 2022. ISSN 2375-2548. doi: 10.1126/sciadv.abo0721. URL https://doi.org/10.1126/sciadv.abo0721https://www.science.org/doi/10.1126/sciadv.abo0721.

201. Bing Wu, Liqun Luo, and Xiaojing J. Gao. Cas9-triggered chain ablation of cas9 as a gene drive brake. *Nature Biotechnology*, 34(2):137–138, 2 2016. ISSN 1087-0156. doi: 10.1038/nbt.3444. URL http://dx.doi.org/10.1038/nbt.3444http://www.nature.com/articles/nbt.3444.

202. Chris Stokel-Walker and Richard Van Noorden. What ChatGPT and generative AI mean for science. *Nature*, 614(7947):214–216, 2 2023. ISSN 0028-0836. doi: 10.1038/d41586-023-00340-6. URL https://www.nature.com/articles/d41586-023-00340-6.

203. R Core Team. R: A Language and Environment for Statistical Computing, 2021. URL https://www.r-project.org/.

204. Wolfgang Viechtbauer. Conducting meta-analyses in {R} with the {metafor} package. *Journal of Statistical Software*, 36(3): 1–48, 2010. doi: 10.18637/jss.v036.i03.

205. Shinichi Nakagawa and Eduardo S A Santos. Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology*, 26:1253–1274, 2012.

206. Shinichi Nakagawa, Malgorzata Lagisz, Rose E O'Dea, Joanna Rutkowska, Yefeng Yang, Daniel W A Noble, and Alistair M Senior. The orchard plot: Cultivating a forest plot for use in ecology, evolution, and beyond. *Research Synthesis Methods*, 12: 4–12, 2021. doi: DOI:10.1002/jrsm.1424.

207. Russell V Lenth. emmeans: Estimated Marginal Means, aka Least-Squares Means, 2023. URL https://cran.r-project.org/package=emmeans.

208. Georg Oberhofer, Tobin Ivy, and Bruce A Hay. Gene drive and resilience through renewal with next generation Cleave and Rescue selfish genetic elements. *Proceedings of the National Academy of Sciences*, 117(16):9013–9021, 4 2020. ISSN 0027-8424. doi: 10.1073/pnas.1921698117. URL http://www.pnas.org/content/117/16/9013.abstracthttp://www.pnas.org/lookup/doi/10.1073/pnas.1921698117.

209. W A Hickey and G B Jr Craig. Genetic distortion of sex ratio in a mosquito, Aedes aegypti. *Genetics*, 53(6):1177–1196, 6 1966. ISSN 0016-6731 (Print). doi: 10.1093/genetics/53.6.1177.

210. T L Sweeny and A R Barr. Sex Ratio Distortion Caused by Meiotic Drive in a Mosquito, Culex pipiens L. *Genetics*, 88(3): 427–446, 3 1978. ISSN 0016-6731 (Print). doi: 10.1093/genetics/88.3.427.

211. Kalle Magnusson, Antonio M. Mendes, Nikolai Windbichler, Philippos-Aris Papathanos, Tony Nolan, Tania Dottorini, Ermanno Rizzi, George K. Christophides, and Andrea Crisanti. Transcription Regulation of Sex-Biased Genes during Ontogeny in the Malaria Vector Anopheles gambiae. *PLoS ONE*, 6(6):e21572, 6 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0021572. URL https://dx.plos.org/10.1371/journal.pone.0021572.

212. Vanessa L. Bauer DuMont, Heather A. Flores, Mark H. Wright, and Charles F. Aquadro. Recurrent Positive Selection at Bgcn, a Key Determinant of Germ Line Differentiation, Does Not Appear to be Driven by Simple Coevolution with Its Partner Protein Bam. *Molecular Biology and Evolution*, 24(1):182–191, 1 2007. ISSN 1537-1719. doi: 10.1093/molbev/msl141. URL https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msl141.

213. Azadeh Aryan, Michelle A. E. Anderson, James K. Biedler, Yumin Qi, Justin M. Overcash, Anastasia N. Naumenko, Maria V. Sharakhova, Chunhong Mao, Zach N. Adelman, and Zhijian Tu. Nix alone is sufficient to convert female Aedes aegypti into fertile males and myo-sex is needed for male flight. *Proceedings of the National Academy of Sciences*, 117(30): 17702–17709, 7 2020. ISSN 0027-8424. doi: 10.1073/pnas.2001132117. URL https://www.pnas.org/content/117/30/17702http://www.pnas.org/lookup/doi/10.1073/pnas.2001132117.

214. Fatwa Adikusuma, Nicole Williams, Frank Grutzner, James Hughes, and Paul Thomas. Targeted Deletion of an Entire Chromosome Using CRISPR/Cas9. *Molecular Therapy*, 25(8):1736–1738, 8 2017. ISSN 15250016. doi: 10.1016/j.ymthe.2017.05.021. URL https://linkinghub.elsevier.com/retrieve/pii/S1525001617302575.

215. Erwei Zuo, Xiaona Huo, Xuan Yao, Xinde Hu, Yidi Sun, Jianhang Yin, Bingbing He, Xing Wang, Linyu Shi, Jie Ping, Yu Wei, Wenqin Ying, Wei Wei, Wenjia Liu, Cheng Tang, Yixue Li, Jiazhi Hu, and Hui Yang. CRISPR/Cas9-mediated targeted chromosome elimination. *Genome Biology*, 18(1):224, 12 2017. ISSN 1474-760X. doi: 10.1186/s13059-017-1354-4. URL https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1354-4.

216. Barbara Fasulo, Angela Meccariello, Maya Morgan, Carl Borufka, Philippos Aris Papathanos, and Nikolai Windbichler. A fly model establishes distinct mechanisms for synthetic CRISPR/Cas9 sex distorters. *PLOS Genetics*, 16(3):e1008647, 3 2020. ISSN 1553-7404. doi: 10.1371/journal.pgen.1008647. URL http://dx.doi.org/10.1371/journal.pgen.1008647https://dx.plos.org/10.1371/journal.pgen.1008647.

217. Hui Xu, Mingzhe Han, Shiyi Zhou, Bing-zhi Li, Yi Wu, and Ying-jin Yuan. Chromosome drives via CRISPR-Cas9 in yeast. *Nature Communications*, 11(1):4344, 12 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-18222-0. URL http://www.nature.com/articles/s41467-020-18222-0.

218. Perran A Ross, Nancy M. Endersby-Harshman, and Ary A Hoffmann. A comprehensive assessment of inbreeding and laboratory adaptation in Aedes aegypti mosquitoes. *Evolutionary Applications*, 12(3):572–586, 3 2019. ISSN 1752-4571. doi: 10.1111/eva.12740. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/eva.12740.

219. Andrea Gloria-Soria, John Soghigian, David Kellner, and Jeffrey R Powell. Genetic diversity of laboratory strains and implications for research: The case of Aedes aegypti. *PLOS Neglected Tropical Diseases*, 13(12):e0007930, 12 2019. ISSN 1935-2735. doi: 10.1371/journal.pntd.0007930. URL https://doi.org/10.1371/journal.pntd.0007930https://dx.plos.org/10.1371/journal.pntd.0007930.

220. Benjamin J. Matthews, Olga Dudchenko, Sarah B. Kingan, Sergey Koren, Igor Antoshechkin, Jacob E. Crawford, William J. Glassford, Margaret Herre, Seth N. Redmond, Noah H. Rose, Gareth D. Weedall, Yang Wu, Sanjit S. Batra, Carlos A. Brito-Sierra, Steven D. Buckingham, Corey L. Campbell, Saki Chan, Eric Cox, Benjamin R. Evans, Thanyalak Fansiri, Igor Filipović, Albin Fontaine, Andrea Gloria-Soria, Richard Hall, Vinita S. Joardar, Andrew K. Jones, Raissa G. G. Kay, Vamsi K. Kodali, Joyce Lee, Gareth J. Lycett, Sara N. Mitchell, Jill Muehling, Michael R. Murphy, Arina D. Omer, Frederick A. Partridge, Paul Peluso, Aviva Presser Aiden, Vidya Ramasamy, Gordana Rašić, Sourav Roy, Karla Saavedra-Rodriguez, Shruti Sharan, Atashi Sharma, Melissa Laird Smith, Joe Turner, Allison M. Weakley, Zhilei Zhao, Omar S. Akbari, William C. Black, Han Cao, Alistair C. Darby, Catherine A. Hill, J. Spencer Johnston, Terence D. Murphy, Alexander S. Raikhel, David B. Sattelle, Igor V. Sharakhov, Bradley J. White, Li Zhao, Erez Lieberman Aiden, Richard S. Mann, Louis Lambrechts, Jeffrey R. Powell, Maria V. Sharakhova, Zhijian Tu, Hugh M. Robertson, Carolyn S. McBride, Alex R. Hastie, Jonas Korlach, Daniel E. Neafsey, Adam M. Phillippy, and Leslie B. Vosshall. Improved reference genome of Aedes aegypti informs arbovirus vector control. *Nature*, 563(7732):501–507, 11 2018. ISSN 0028-0836. doi: 10.1038/s41586-018-0692-z. URL http://www.nature.com/articles/s41586-018-0692-z.

221. Gloria I. Giraldo-Calderón, Scott J. Emrich, Robert M. MacCallum, Gareth Maslen, Emmanuel Dialynas, Pantelis Topalis, Nicholas Ho, Sandra Gesing, Gregory Madey, Frank H. Collins, and Daniel Lawson. VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. *Nucleic Acids Research*, 43(D1):D707–D713, 1 2015. ISSN 1362-4962. doi: 10.1093/nar/gku1117. URL https://academic.oup.com/nar/article/43/D1/D707/2437683.

222. Andrew Marc Hammond. *The development of gene drives for genetic control of the malaria mosquito*. PhD thesis, Imperial College London, 2016.

223. E Krzywinska, V Kokoza, M Morris, E de la Casa-Esperon, A S Raikhel, and J Krzywinski. The sex locus is tightly linked to factors conferring sex-specific lethal effects in the mosquito Aedes aegypti. *Heredity*, 117(6):408–416, 12 2016. ISSN 1365-2540 (Electronic). doi: 10.1038/hdy.2016.57.

224. Dina Simkin, Vasileios Papakis, Bernabe I Bustos, Christina M Ambrosi, Steven J Ryan, Valeriya Baru, Luis A Williams, Graham T Dempsey, Owen B McManus, John E Landers, Steven J Lubbe, Alfred L George, and Evangelos Kiskinis. Homozygous might be hemizygous: CRISPR/Cas9 editing in iPSCs results in detrimental on-target defects that escape standard quality controls. *Stem Cell Reports*, 17(4):993–1008, 2022. ISSN 2213-6711. doi: https://doi.org/10.1016/j.stemcr.2022.02.008. URL https://www.sciencedirect.com/science/article/pii/S2213671122000972.

225. Rebeca Carballar-Lejarazú, Christian Ogaugwu, Taylor Tushar, Adam Kelsey, Thai Binh Pham, Jazmin Murphy, Hanno Schmidt, Yoosook Lee, Gregory C. Lanzaro, and Anthony A. James. Next-generation gene drive for population modification of the malaria vector mosquito, Anopheles gambiae. *Proceedings of the National Academy of Sciences*, 117(37):22805–22814, 9 2020. ISSN 0027-8424. doi: 10.1073/pnas.2010214117. URL http://www.pnas.org/lookup/doi/10.1073/pnas.2010214117.

226. Ying Dang, Gengxiang Jia, Jennie Choi, Hongming Ma, Edgar Anaya, Chunting Ye, Premlata Shankar, and Haoquan Wu. Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biology*, 16(1):280, 2015. ISSN 1474-760X. doi: 10.1186/s13059-015-0846-3. URL https://doi.org/10.1186/s13059-015-0846-3.

227. Xuechun Feng, Víctor López Del Amo, Enzo Mameli, Megan Lee, Alena L Bishop, Norbert Perrimon, and Valentino M Gantz. Optimized CRISPR tools and site-directed transgenesis towards gene drive development in Culex quinquefasciatus mosquitoes. *Nature Communications*, 12(1):2960, 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-23239-0. URL https://doi.org/10.1038/s41467-021-23239-0.

228. Bernd Pulverer. When things go wrong: correcting the scientific record., 10 2015. ISSN 1460-2075 (Electronic).

229. Peter S Choi and Matthew Meyerson. Targeted genomic rearrangements using CRISPR/Cas technology. *Nature Communications*, 5(1):3728, 2014. ISSN 2041-1723. doi: 10.1038/ncomms4728. URL https://doi.org/10.1038/ncomms4728.

230. Emma Haapaniemi, Sandeep Botla, Jenna Persson, Bernhard Schmierer, and Jussi Taipale. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nature Medicine*, 24(7):927–930, 2018. ISSN 1546-170X. doi: 10.1038/s41591-018-0049-z. URL https://doi.org/10.1038/s41591-018-0049-z.

231. Long Jiang, Katrine Ingelshed, Yunbing Shen, Sanjaykumar V Boddul, Vaishnavi Srinivasan Iyer, Zsolt Kasza, Saikiran Sedimbi, David P Lane, and Fredrik Wermeling. CRISPR/Cas9-Induced DNA Damage Enriches for Mutations in a p53-Linked Interactome: Implications for CRISPR-Based Therapies. *Cancer Research*, 82(1):36 LP – 45, 1 2022. doi: 10.1158/0008-5472.CAN-21-1692. URL http://cancerres.aacrjournals.org/content/82/1/36.abstract.

232. Lucie A Bergeron, Søren Besenbacher, Jiao Zheng, Panyi Li, Mads Frost Bertelsen, Benoit Quintard, Joseph I Hoffman, Zhipeng Li, Judy St. Leger, Changwei Shao, Josefin Stiller, M Thomas P Gilbert, Mikkel H Schierup, and Guojie Zhang. Evolution of the germline mutation rate across vertebrates. *Nature*, 615(7951):285–291, 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-05752-y. URL https://doi.org/10.1038/s41586-023-05752-y.

233. Angela F Harris, Andrew R McKemey, Derric Nimmo, Zoe Curtis, Isaac Black, Siân A Morgan, Marco Neira Oviedo, Renaud Lacroix, Neil Naish, Neil I Morrison, Amandine Collado, Jessica Stevenson, Sarah Scaife, Tarig Dafa'alla, Guoliang Fu, Caroline Phillips, Andrea Miles, Norzahira Raduan, Nick Kelly, Camilla Beech, Christl A Donnelly, William D Petrie, and Luke Alphey. Successful suppression of a field mosquito population by sustained release of engineered male mosquitoes. *Nature Biotechnology*, 30(9):828–830, 2012. ISSN 1546-1696. doi: 10.1038/nbt.2350. URL https://doi.org/10.1038/nbt.2350.

234. Renaud Lacroix, Andrew R McKemey, Norzahira Raduan, Lim Kwee Wee, Wong Hong Ming, Teoh Guat Ney, Siti Rahidah A.A., Sawaluddin Salman, Selvi Subramaniam, Oreenaiza Nordin, Norhaida Hanum A.T., Chandru Angamuthu, Suria Marlina Mansor, Rosemary S Lees, Neil Naish, Sarah Scaife, Pam Gray, Geneviève Labbé, Camilla Beech, Derric Nimmo, Luke Alphey, Seshadri S Vasan, Lee Han Lim, Nazni Wasi A., and Shahnaz Murad. Open Field Release of Genetically Engineered Sterile Male Aedes aegypti in Malaysia. *PLOS ONE*, 7(8):e42771, 8 2012. URL https://doi.org/10.1371/journal.pone.0042771.

235. Danilo O Carvalho, Andrew R McKemey, Luiza Garziera, Renaud Lacroix, Christl A Donnelly, Luke Alphey, Aldo Malavasi, and Margareth L Capurro. Suppression of a Field Population of Aedes aegypti in Brazil by Sustained Release of Transgenic Male Mosquitoes. *PLOS Neglected Tropical Diseases*, 9(7):1–15, 2015. doi: 10.1371/journal.pntd.0003864. URL https://doi.org/10.1371/journal.pntd.0003864.

236. Adi Utarini, Citra Indriani, Riris A. Ahmad, Warsito Tantowijoyo, Eggi Arguni, M. Ridwan Ansari, Endah Supriyati, D. Satria Wardana, Yeti Meitika, Inggrid Ernesia, Indah Nurhayati, Equatori Prabowo, Bekti Andari, Benjamin R. Green, Lauren Hodgson, Zoe Cutcher, Edwige Rancès, Peter A. Ryan, Scott L. O'Neill, Suzanne M. Dufault, Stephanie K. Tanamas, Nicholas P. Jewell, Katherine L. Anders, and Cameron P. Simmons. Efficacy of Wolbachia-Infected Mosquito Deployments for the Control of Dengue. *New England Journal of Medicine*, 384(23):2177–2186, 6 2021. ISSN 0028-4793. doi: 10.1056/NEJMoa2030243. URL http://www.nejm.org/doi/10.1056/NEJMoa2030243.

237. S Bhatt, D J Weiss, E Cameron, D Bisanzio, B Mappin, U Dalrymple, K E Battle, C L Moyes, A Henry, P A Eckhoff, E A Wenger, O Briët, M A Penny, T A Smith, A Bennett, J Yukich, T P Eisele, J T Griffin, C A Fergus, M Lynch, F Lindgren, J M Cohen, C L J Murray, D L Smith, S I Hay, R E Cibulskis, and P W Gething. The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature*, 526(7572):207–211, 2015. ISSN 1476-4687. doi: 10.1038/nature15535. URL https://doi.org/10.1038/nature15535.

238. Adeline E Williams, Alexander W E Franz, William R Reid, and Ken E Olson. Antiviral Effectors and Gene Drive Strategies for Mosquito Population Suppression or Replacement to Mitigate Arbovirus Transmission by Aedes aegypti. *Insects*, 11(1), 2020. ISSN 2075-4450. doi: 10.3390/insects11010052. URL https://www.mdpi.com/2075-4450/11/1/52.

239. John M. Marshall, Robyn R. Raban, Nikolay P. Kandul, Jyotheeswara R. Edula, Tomás M. León, and Omar S. Akbari. Winning the Tug-of-War Between Effector Gene Design and Pathogen Evolution in Vector Population Replacement Strategies. *Frontiers in Genetics*, 10(OCT):1–12, 10 2019. ISSN 1664-8021. doi: 10.3389/fgene.2019.01072. URL https://www.frontiersin.org/article/10.3389/fgene.2019.01072/full.

240. William R Reid, Ken E Olson, and Alexander W E Franz. Current Effector and Gene-Drive Developments to Engineer Arbovirus-Resistant Aedes aegypti (Diptera: Culicidae) for a Sustainable Population Replacement Strategy in the Field. *Journal of Medical Entomology*, 58(5):1987–1996, 2021. ISSN 0022-2585. doi: 10.1093/jme/tjab030. URL https://doi.org/10.1093/jme/tjab030.

241. Scott L O'Neill. The Use of Wolbachia by the World Mosquito Program to Interrupt Transmission of Aedes aegypti Transmitted Viruses BT - Dengue and Zika: Control and Antiviral Treatment Strategies. pages 355–360. Springer Singapore, Singapore, 2018. ISBN 978-981-10-8727-1. doi: 10.1007/978-981-10-8727-1{_}24. URL https://doi.org/10.1007/978-981-10-8727-1_24.

242. Michael S. Jones, Jason A. Delborne, Johanna Elsensohn, Paul D. Mitchell, and Zachary S. Brown. Does the U.S. Public support using gene drives in agriculture? And what do they want to know? *Science Advances*, 5(9), 2019. ISSN 23752548. doi: 10.1126/sciadv.aau8462.

243. IAEA. DIR-SIT Database. URL https://nucleus.iaea.org/sites/naipc/dirsit/SitePages/AllFacilities.aspx.

244. Sierra M Brooks and Hal S Alper. Applications, challenges, and needs for employing synthetic biology beyond the lab. *Nature Communications*, 12(1):1390, 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-21740-0. URL https://doi.org/10.1038/s41467-021-21740-0.

245. A G Parker, W Mamai, and H Maiga. Mass-rearing for the sterile insect technique. In *Sterile insect technique*, pages 283–316. CRC Press, 2021. ISBN 1003035574.