Title

In Silico Exploration and Analysis of Gene Drive Efficacy

Author

Bennett, Jared B

Publication Date

2021

Peer reviewed|Thesis/dissertation

In Silico Exploration and Analysis of Gene Drive Efficacy

by

Jared B. Bennett

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Biophysics

and the Designated Emphasis

in

Computational Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Assistant Professor in Residence John Marshall, Co-chair
Associate Professor Oskar Hallatschek, Co-chair
Professor Adam Arkin
Professor Michael Boots

Summer 2021

In Silico Exploration and Analysis of Gene Drive Efficacy

Abstract

In Silico Exploration and Analysis of Gene Drive Efficacy

by

Jared B. Bennett

Doctor of Philosophy in Biophysics

and the Designated Emphasis in

Computational Biology

University of California, Berkeley

Assistant Professor in Residence John Marshall, Co-chair

Associate Professor Oskar Hallatschek, Co-chair

Malaria, dengue, Zika, and other mosquito-borne diseases continue to pose a major global health burden through much of the world, despite decades of work combating the mosquito vector and the virus or parasite responsible for each disease. The advent of CRISPR/Cas9-based gene editing greatly simplifies design and implementation of novel genetic-based approaches for vector control. This has reignited interest in gene drive technologies, requiring an understanding of the population and spatial dynamics of the wide range of novel gene drive architectures in development.

This thesis traverses the path from design, to understanding, to prediction of the dynamics for novel gene drive constructs. Chapter 1 develops and extends simulation frameworks for exploring the dynamics of gene drives. The first half enhances the field's genetic and geospatial simulation capabilities, while the second half builds upon that to utilize time-varying properties applied with a continuous-time framework. The outcome of this chapter is two expansive and computationally-efficient software packages freely available in the R ecosystem.

Chapter 2 acknowledges our limited knowledge of factors that impact the success of different gene drive constructs, and in conjunction with the biological design and testing of novel constructs, develops a robust inference methodology to understand and quantify the mechanisms underpinning gene drive behavior. Chapter 2 begins with a standard MCR design targeting a non-essential gene. We then expand our approach for treatment of split-gene

designs that target essential genes with different (pre or post-zygotic) fertility impacts. All of this work is done in a manner consistent with the simulation frameworks from chapter 1.

Finally, chapter 3 applies the models developed in chapter 1, while integrating knowledge gained in chapter 2, to forecast the dynamics of innovative gene drive constructs in realistic settings. The chapter begins with a large-scale exploration of two classic designs, now generated using a CRISPR-based approach, to query the chance of "success", any dangers of "success", and repercussions if "success" was not the correct goal. The chapter finishes with a recent design that seeks to bridge the space between linked and split-gene drive designs, seeking to combine the benefits of both ideas. It is a fitting conclusion to a project that spans math and biology to generate a mechanistic and predictive understanding of modern gene drives.

To Joseph E. Bennett, who saw me in,

and William C. Bradford, who nearly saw me out.

Contents

List of Figures

List of Tables

Acknowledgments

This is the section where I list my family and friends, mentors and colleagues, and thank everyone for their unique and cherished contributions over the last few years. However, it would be embarrassing to forget someone, or even worse, to incorrectly attribute anything, and given my current state of exhaustion ...

Please accept my apologies for not attempting to list everyone by name, and my thanks, for your ideas, your support, and your presence in my life during this journey.

Thank you.

List of Publications

Kaduskar B, Kushwah RB, Auradkar A, Guichard A, Li M, **Bennett JB**, Julio A, Marshall J, Montell C, Bier E. Reversing insecticide resistance with allelic-drive in *Drosophila melanogaster* (*preprint*)

Li M, Yang T, Bui M, Gamez S, Wise T, Kandul NP, Liu J, Alcantara L, Lee H, Edula JR, Raban R, Zhan Y, Wang Y, DeBeaubien N, Chen J, Sanchez HMC, **Bennett JB**, Antoshechkin I, Montell C, Marshall JM, Akbari OS. Eliminating Mosquitoes with Precision Guided Sterile Males (*preprint*)

Wu SL, **Bennett JB**, Sanchez HM, Dolgert AJ, León T, Marshall JM. MGDrivE 2: A simulation framework for gene drive systems incorporating seasonality and epidemiological dynamics

Terradas G, Buchman AB, **Bennett JB**, Shriner I, Marshall JM, Akbari OS, Bier E. Inherently confinable split-drive systems in *Drosophila*

Kandul NP, Liu J, **Bennett JB**, Marshall JM, Akbari OS. A confinable home and rescue gene drive for population modification

Adolfi A, Gantz VM, Jasinskiene N, Lee HF, Hwang K, Bulger EA, Ramaiah A, **Bennett JB**, Terradas G, Emerson JJ, Marshall JM, Bier E, James AA. Efficient population modification gene-drive rescue system in the malaria mosquito *Anopheles stephensi*

Xu XRS, Bulger EA, Gantz VM, Klanseck C, Heimler SR, Auradkar A, **Bennett JB**, Miller LA, Leahy S, Juste SS, Buchman A, Akbari OS, Marshall JM, Bier E. Active Genetic Neutralizing Elements for Halting or Deleting Gene Drives

Sánchez HM, **Bennett JB**, Wu SL, Rašić G, Akbari OS, Marshall JM. Modeling confinement and reversibility of threshold-dependent gene drive systems in spatially-explicit *Aedes aegypti* populations

Li M, Yang T, Kandul NP, Bui M, Gamez S, Raban R, **Bennett JB**, Sánchez HM, Lanzaro GC, Schmidt H, Lee Y, Marshall JM, Akbari OS. Development of a confinable gene drive system in the human disease vector, *Aedes aegypti*

López VD, Bishop AL, Sánchez HM, **Bennett JB**, Feng X, Marshall JM, Bier E, Gantz VM. A transcomplementing gene drive provides a flexible platform for laboratory investigation and potential field deployment

Pham TB, Phong CH, **Bennett JB**, Hwang K, Jasinskiene N, Parker K, Stillinger D, Marshall JM, Carballar-Lejarazú R, James AA. Experimental population modification of the malaria vector mosquito, *Anopheles stephensi*

Sánchez C. HM, Wu SL*, **Bennett JB***, Marshall JM. MGDrivE: A modular simulation framework for the spread of gene drives through spatially explicit mosquito populations

Chapter 1

Novel Genetically and Ecologically Representative Models for the Study of Gene Drive Dynamics

Introduction

The increased interest in gene drives has caused a corresponding increase in the necessity to understand their ecological, spatial, and epidemiological dynamics. This requires models that faithfully capture the genetic inheritance and reflect local environments. However, early models were hyper-specific to location[132], hard-coded for specific gene drives of interest[51, 159, 29], or implemented using a population-frequency approach[138, 137]. We endeavored to expand upon these designs and create a flexible simulation framework capable of modeling novel gene drive designs as they arise and readily adaptable to new locations. Our simulation framework, MGDrivE[177], embodies this ethos of flexibility and ease-of-use for the exploration of gene drive dynamics.

While our first model iteration captures the genetic and spatial complexity inherent to gene drives, it fails to reflect temporally-varying dynamics or the impact of gene drives on disease burden. Following the groundbreaking paper by North *et al*[157], we developed the second iteration of our model, MGDrive 2[207], incorporating time-varying dynamics and two epidemiological models - the basic Ross-Macdonald[174, 130] model for malaria, and the Kermack and McKendrik[109, 110] model for chikungunya, Zika, and other arboviruses. These extensions, along with significant internal updates to the model design, allow us to faithfully reflect fluctuating environmental conditions, as long as seasons or as short as daily temperature cycles, and estimate reduction in disease burden due to the application of novel gene drive technologies.

1.1 MGDrivE: A Modular Simulation Framework for the Spread of Gene Drives Through Spatially Explicit Mosquito Populations[1]

1.1.1 Introduction

The advent of CRISPR/Cas9-based gene editing technology and its application to the engineering of gene drive systems has led to renewed excitement in the use of genetics-based strategies to control mosquito vectors of human diseases and insect agricultural pests[28, 54]. Applications to control mosquito-borne diseases have gained the most attention due to the major global health burden they pose through much of the world and the difficulty of controlling them using currently available tools[198].

The ease of gene editing afforded by CRISPR has also led to significant versatility in terms of the gene drive systems that are now realizable[28, 139]. These include: (a) homing-based systems that cleave a specific target site lacking the drive system and are then copied to this site by serving as a template for homology-directed repair (HDR)[22, 204], (b) remediation systems that could be used to remove effector genes and possibly drive systems from the environment in the event of unwanted consequences[61, 139], and (c) threshold-dependent systems that may permit confineable and reversible releases[4, 21, 141].

Understanding how these systems are expected to behave in real ecosystems requires a flexible modelling framework that can accommodate a range of inheritance patterns, species-specific details, and landscape details where a construct may be released. To this end, we present MGDrivE (Mosquito Gene Drive Explorer): a flexible simulation framework designed to investigate the population dynamics of a variety of gene drive systems and their spread through spatially explicit populations of mosquitoes and other insects.

MGDrivE is unique in its ability to simulate diverse, user-specified inheritance-modifying systems within a single, computationally efficient framework incorporating insect life history and landscape ecology. Other existing frameworks were designed for general purpose simulations and applied to gene drive studies (Table 1.1). For example, Eckhoff[51] used the EMOD malaria model to simulate the spread of homing-based gene drive systems through spatial populations of *Anopheles gambiae*. EMOD is open source and a powerful modelling framework; but significant effort is required to redefine genetic control strategies, life-history parameters and landscape details. Magori et al.[132] created Skeeter Buster by extending the CIMSiM model[56] to incorporate genetic inheritance and spatial structure. Skeeter Buster captures most pertinent mosquito ecology considerations, but is not open source and can only simulate a handful of genetic control strategies[118]. The SLiM genetic simula-

[1] This work has been previously published[177]

tion framework[78] is capable of modelling the spread of user-defined gene drive systems in a metapopulation; however, it does not currently accommodate life-history ecology and overlapping generations.

	Inheritance Patterns	**Life-History Ecology**	**Spatial and Landscape Details**	**Software**
MGDrivE	Very flexible, can be user-specified	Egg-larva-pupa-adult, density-dependence at larval stage, not responsive to environmental variables at present	Populations distributed in space, connected by migration	R package, open source
EMOD[51]	Homing-based gene drive, could be extended with effort	Egg-larva-pupa-adult, density-dependence at larval stage, responsive to environmental variables	Populations arranged ona grid, each representing $1km^2$, connected by migration	Java Script Open Notation (JSON) feeds into executable file, open source
Skeeter Buster[118]	Homing-based gene drive, release of insects carrying a conditional lethal, etc., cannot be user-specified	Egg-larva-pupa-adult, density-dependence at larval stage, responsive to environmental variables	Households and containers modeled explicitly, connected by migration	Executable file, not open source
SLiM[78]	Very flexible, can be user-specified	Discrete generations, no life history at present	Can model either connected populations or cells on a grid	Scripting environment with graphical user interface, open source

Table 1.1: **Comparison of Spatially Explicit Gene Drive Models**

In this paper, we describe the key components of the MGDrivE framework – genetic inheritance, mosquito life history and landscape. We then demonstrate the application of the framework to CRISPR-based homing gene drive systems for: (a) driving a disease-refractory

gene into a population (i.e. population replacement), and (b) disrupting a gene required for female fertility (i.e. population suppression), incorporating homing-resistant alleles. We conclude with a discussion of future applications of genetic simulation packages in the field of gene drive modelling.

1.1.2 Design and Implementation

The MGDrivE framework is a genetic and spatial extension of the lumped age-class model of mosquito ecology by Hancock[88], modified and applied by Deredec[46] to the spread of homing gene drive systems, and by Marshall[142] to population-suppressing homing systems in the presence of resistant alleles. The framework incorporates the egg, larval, pupal and adult life stages, with egg genotypes determined by maternal and paternal genotypes and the allelic inheritance pattern. In MGDrivE, by treating the lumped age-class model equations in a variable-dimension tensor algebraic form, the population dynamic equations can be left unchanged while modifying the dimensionality of the tensor describing inheritance patterns, as required by the number of genotypes associated with the drive system. Spatial dynamics are simulated by a metapopulation structure in which migrants are exchanged between populations with defined probabilities. Full details of this framework are provided in Data S1.

The core framework is developed in R with certain routines in Rcpp for computational speed. By combining the tensor modelling framework with object-oriented programming, the genetic, life history and spatial components of the model are able to be separated into 'modules' to facilitate ease of modification. We now describe the three modules in more detail.

1.1.2.1 Modules

Genetic Inheritance

The fundamental module for gene drive dynamics is that describing genetic inheritance. In MGDrivE, this is embodied by a three-dimensional tensor contained in a drive-specific R file and referred to as an 'inheritance cube' (Figure 1.1). The first and second dimensions of the inheritance cube refer to the maternal and paternal genotypes, respectively, and the third refers to the offspring genotype. Cube entries for each combination of parental genotypes represent the proportion of offspring that are expected to have each genotype, and should sum to one, as fitness and viability are accommodated separately.

The R function that builds the inheritance cube may receive a number of user-defined input parameters. For example, for a homing-based drive system, the list of input parameters should include the homing efficiency, the rate of in-frame resistant allele generation, and the rate of out-of-frame or otherwise costly resistant allele generation[142, 196]. In-frame resistant alleles are those for which the coding frame of the target site is not altered, leading to minimal fitness effects, while out-of-frame resistant alleles disrupt the coding frame, leading to significant fitness effects. Input parameters also include those associated with organisms

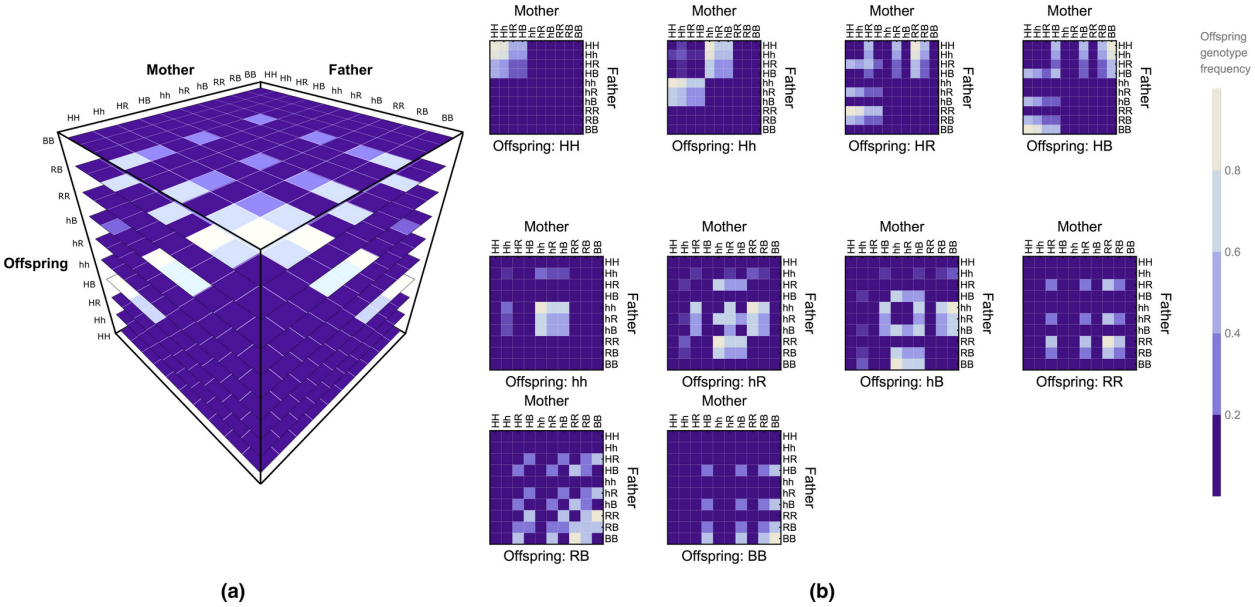

Figure 1.1: **Inheritance Module**
Genetic inheritance is embodied by a three-dimensional tensor referred to as an inheritance cube (left), here depicted for a CRISPR-based homing construct. Maternal and paternal genotypes are depicted on the x and y-axes and offspring genotypes on the z-axis, with slices of the cube pertaining to each offspring genotype shown to the right. The inheritance pattern shown deviates from standard Mendelian inheritance such that, in the germline of Hh parents, the majority of wild-type (h) alleles are converted into homing (H) alleles, while a small proportion are converted into in-frame resistant (R) and out-of-frame resistant alleles (B). For the example pictured, the frequency of accurate homing given cleavage in Hh heterozygotes is 98%, with the remaining 2% of wild-type alleles being converted to either in-frame (1%), or out-of-frame (1%) resistant alleles.

having each genotype – for example, genotype-specific: (a) fertility rates, (b) male mating fitness, (c) sex bias at emergence, (d) adult survival rates, and (e) female and male pupatory success. These parameters feed into the mosquito life-history module, which will be described next. Finally, a 'viability mask' is applied to the offspring genotypes to remove unviable genotypes from the population.

At the time of publication, the MGDrivE package includes inheritance cubes for: (a) standard Mendelian inheritance, (b) homing-based drive intended for population replacement or suppression[62, 80, 142], (c) Medea (a maternal toxin linked to a zygotic antidote)[35], (d) other toxin-antidote-based underdominant systems such as UD^{MEL}[4, 137], (e) reciprocal chromosomal translocations[21], (f) Wolbachia[85], and (g) the RIDL and pgSIT systems[107,

205] (release of insects carrying a dominant lethal gene, and precision-guided sterile insect technique). Details of each of these systems are provided in the online documentation.

Mosquito Life History

The mosquito life-history module follows from the lumped age-class model of Hancock and Godfray[88] adapted by Deredec[46]. In this model (depicted in Figure 1.2), the insect life cycle is divided into four stages – egg (E), larva (L), pupa (P) and adult (F for female and M for male). In MGDrivE, each life stage is associated with a genotype. Adult females mate once and produce batches of eggs from the sperm of the same male, so they obtain a composite genotype upon mating (their own and that of the male they mate with). Egg genotypes are then determined by the parental genotypes and inheritance pattern as provided in the inheritance cube. The adult equilibrium population size, N, in a given habitat patch is used to determine the carrying capacity of that patch for larvae, K, which determines the degree of additional density-dependent mortality at the larval stage in that patch. Following Deredec[46], this is described by an equation of the form: $f(L) = \frac{\alpha}{(\alpha+L)^{\frac{1}{T_L}}}$, where L is the number of larvae in the patch, T_L is the duration of the larval stage, and α is a parameter describing the strength of density dependence. Further details on the mathematical formulation of the lumped-age class model and its generalization to an arbitrary number of genotypes using tensor algebra are provided in Data S1.

The MGDrivE framework currently applies to any species having an egg-larva-pupa-adult life history and for which density-dependent regulation occurs at the larval stage. Switching between species can be achieved by altering parameter values within this module: (a) the number of eggs produced per adult female per day, (b) the durations of the egg, larval and pupal juvenile life stages, (c) the daily mortality risk for the adult life stage, and (d) the daily population growth rate (in the absence of density-dependent mortality). The daily density-independent mortality risks for the juvenile stages are assumed to be identical and are chosen for consistency with the daily population growth rate. Default life-history parameter values are shown in Table 1.2 for three species of interest: (a) *A. gambiae*, the main African malaria vector, (b) *Aedes aegypti*, the main vector of dengue and Zika virus, and (c) *Ceratitis capitata*, a worldwide agricultural crop pest. In some cases, life-history parameters are modified in genotype-specific ways by the gene drive construct. A current limitation of the framework is that equilibrium population size remains constant over time. This will be addressed in the next released version of MGDrivE.

Landscape

The landscape module describes the distribution of mosquito populations in space, with movement through the resulting network determined by a dispersal kernel. Discrete populations in the resulting metapopulation are randomly mixing populations for which the equations of the lumped age-class model apply. The resolution of the individual populations (in terms of size) should be chosen according to the dispersal properties of the insect species

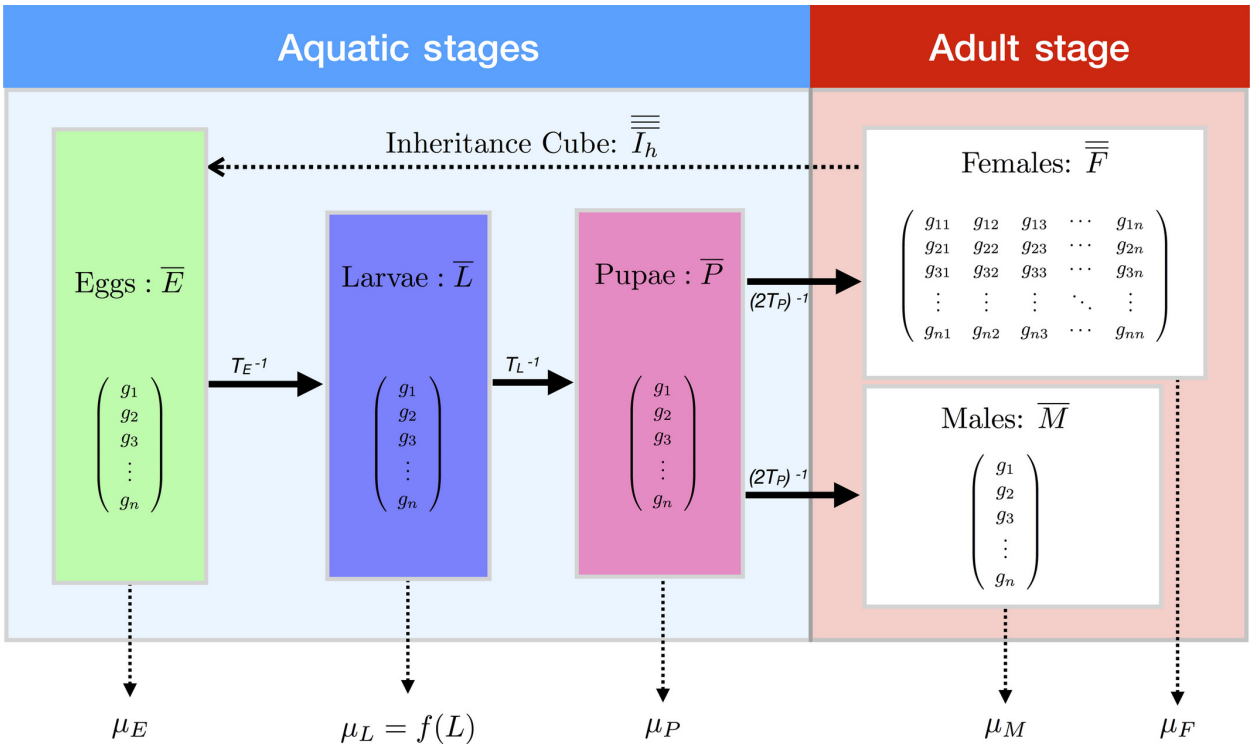

Figure 1.2: **Mosquito Life-History Module**
Life history is modelled according to an egg (E)-larva (L)-pupa (P)-adult (F for female, M for male) life cycle in which density dependence occurs at the larval stage and autonomous mobility occurs at the adult stage. Genotypes are tracked across all life stages, represented by the subscript $i \in \{1, \ldots, g\}$. For example, M_i represents the number of adult males having the i^{th} genotype. Females are modelled as mating once upon emergence and obtain a composite genotype – their own and that of the male they mate with. Egg genotypes are determined by the adult female's composite genotype and the inheritance pattern, which is specific to the gene drive system under consideration.

of interest and the research question being investigated. *A. aegypti* mosquitoes, for instance, are thought to be relatively local dispersers, often remaining in the same household for the duration of their lifespan[179]. For modelling the fine-scale spread of gene drive systems in this species, populations on the scale of households may be appropriate. *A. gambiae* mosquitoes, on the other hand, are thought to display moderate dispersal on the village scale and infrequent inter-village movement[191]. This would suggest villages as an appropriate population unit; however other levels of aggregation are also possible, in both cases, depending on the resolution required and computational power available.

Daily per-capita movement probabilities between populations (nodes in the network) for

Parameter	Symbol	*Aedes aegypti*	*Anopheles gambiae*	*Ceratitis capitata*
Egg production per female (day^{-1})	β	20[162]	32[45]	20[47]
Duration of egg stage (days)	T_E	5[37]	1[45]	2[47]
Duration of larval stage (days)	T_L	6[37]	13[45]	6[47]
Duration of pupa stage (days)	T_P	4[37]	1[45]	10[47]
Daily population growth rate (day^{-1})	r_M	1.175[184]	1.096[148]	1.031[25]
Daily mortality risk of adult stage (day^{-1})	μ_F, μ_M	0.090[193, 55, 96]	0.123[148]	0.100[158]

Table 1.2: **Life-History Module Parameter Values for Three Species of Interest** (at a Temperature of 25°)

these examples were calculated from a zero-inflated exponential kernel, accounting for pairwise distances between nodes. This kernel models movement as a two-stage process, whereby a mosquito first decides whether to leave the current population (governed by a parameter, p_0, representing the daily probability that it remains in the same population), and in the event of movement, selects the destination node from the full set with probabilities based on distance according to an exponential distribution (governed by a parameter, λ, where $\frac{1}{\lambda}$ is approximately equal to the mean dispersal distance in a large landscape). As the simulation only requires a matrix of inter-node movement probabilities, arbitrarily complex kernels that account for barriers, such as roads[179], may be used without altering the model architecture. The matrix of movement probabilities is incorporated in the tensor algebraic model formulation described in Data S1.

Finally, any type of release can be simulated by increasing the number of insects having a given sex and genotype at a specific population and time. As demonstrated in the following use examples, releases are parameterized according to: (a) number of released individuals, (b) number of releases, (c) time of first release, (d) time between releases, (e) population of release, and (f) sex and genotype of released insects.

1.1.2.2 Deterministic Versus Stochastic Simulations

Simulations can be run either in deterministic or stochastic form. Deterministic simulations are faster and less computationally intensive; however, stochastic simulations capture the probabilistic nature of chance events that occur at low population sizes and genotype frequencies. In the stochastic implementation of MGDrivE, daily egg production follows

a Poisson distribution, offspring genotype follows a multinomial distribution informed by parental genotypes and the inheritance cube, mate choice follows a multinomial distribution determined by adult genotype frequencies, and survival and death events follow binomial distributions at the population level. When interpreting stochastic models, many simulations should be run to understand the range of outputs possible for a given model realization.

1.1.3 Results

To demonstrate how the MGDrivE framework can be used to initialize and run a simulation of a gene drive system through a metapopulation, we have provided vignettes with the package, available via instillation from CRAN, and additional examples and information on Github and the package website. The vignettes provide examples of simple simulations and landscape setup. They begin with a deterministic example of Mendelian inheritance, and explore expected genotype frequencies according to Hardy-Weinberg equilibrium, before violating some of these assumptions. Next, they explore the effects of genotype-specific fitness costs on genotype trajectories and population size. The impact of stochasticity on model predictions is then explored through stochastic simulations, with dynamics being compared to those expected from equivalent deterministic simulations.

Here, we describe the application of the package to two homing gene drive strategies: (a) driving a disease-refractory gene into a population[62], and (b) disrupting a gene required for female fertility and hence suppressing a population[80]. In both cases, we consider a population of *A. aegypti* mosquitoes having the bionomic parameters provided in Table 1.2 and distributed through the network landscape depicted in Figure 1.3. To demonstrate the functionality of the MGDrivE package, we model both strategies using the deterministic and stochastic implementations. In both cases, we include the generation of in-frame and out-of-frame or otherwise costly resistant alleles[31] and parameterize the gene drive model based on recently engineered constructs[62, 80].

1.1.3.1 Population Replacement

We begin by modelling a CRISPR-based homing construct similar to that engineered by Gantz[62]. This was the first CRISPR-based homing construct demonstrated in a mosquito disease vector – namely, *Anopheles stephensi*, the main urban malaria vector in India. For this construct, homing and resistant allele generation were shown to occur at different rates in males and females, and there were large fitness reductions associated with having the homing construct. We consider a homing efficiency of 90% in males and 50% in females – i.e. 90% of wild-type (h) alleles are converted to homing (H) alleles in the germline of Hh males, and 50% of h alleles are converted to H alleles in the germline of Hh females. A third of the remaining h alleles in Hh individuals are converted to in-frame resistant alleles (R), and the remainder are converted to out-of-frame or otherwise costly resistant alleles (B) due to error-prone copying during the homing process[31]. Female fecundity and male mating

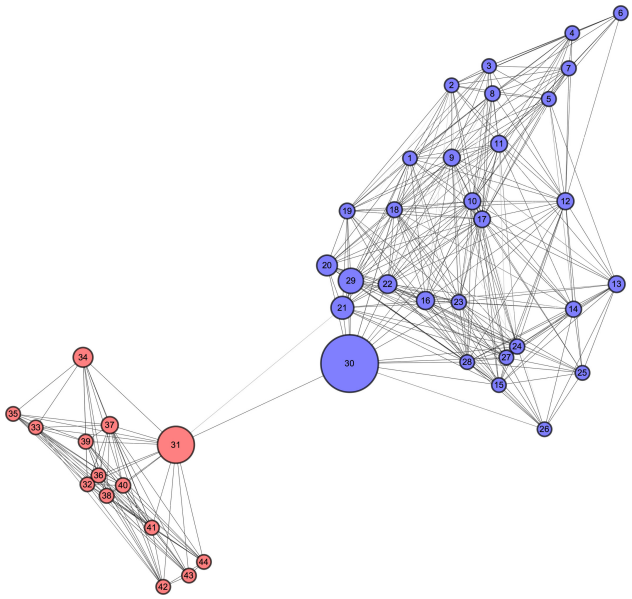

Figure 1.3: **Landscape Module**
Insects are distributed as populations, here depicted by nodes, each having their own coordinates and population size. Movement between populations is derived from a defined dispersal kernel, depicted here by edges between nodes. The example scenario allows both spread within and between communities to be explored. Here, nodes are coloured according to their community (detected by the DBSCAN clustering algorithm[42]), with sizes proportional to their 'betweenness centrality' – a measure of their connectedness to other nodes in the metapopulation[58].

fitness are reduced by 25% per H or R allele and by 50% per B allele.

The general workflow for the simulation is shown in Figure 1.4, with the full code available. We begin by loading the MGDrivE package in R and setting the working and output directories. We then choose between the deterministic and stochastic implementation of the model – in this case, the deterministic version. Next, we specify the bionomic parameters of the species we are modelling – in this case, *A. aegypti*, whose default life-history parameters are provided in Table 1.2. Following this, we define the landscape through which we will model the spread of the drive system. We begin by loading a CSV file containing the coordinates (longitude and latitude) of the populations in Figure 1.3. A function is then applied that computes daily movement rates between each of the populations based on a zero-inflated exponential dispersal kernel, the parameters for which we provide. Equilibrium adult population sizes can be provided for each of the populations; however in this case, we assume these are identical and provide a single population size (Code sample 1.5).

Figure 1.4: **Workflow of an MGDrivE Simulation**

With our life history and landscape modules defined, we now specify the gene drive system and release strategy we intend to model (Code sample 1.6). We use a pre-specified inheritance cube, 'Cube_HomingDrive()', that models the inheritance pattern of a homing-based gene drive system. We specify sex-specific homing rates, resistant allele generation rates and genotype-specific fitness effects based on the construct engineered by Gantz[62]. We then specify the release scheme by generating a list containing: (a) the release size, (b) number of releases, (c) time of first release, and (d) time between releases. This is incorporated into a vector also specifying the inheritance cube and the sex and genotype of the released insects. Finally, the populations in which the release takes place are specified. With the simulation framework now fully specified, the model is ready to run (Code sample 1.7).

```
# LOAD AND SET UP PACKAGES ########################
library(MGDrivE)
## MGDrivE can be set up to run in stochastic/deterministic mode
MGDrivE.Setup(stochasticityON=TRUE)
simulationTime= 5000
## Set to one for the deterministic version
repetitions= 100
# SET UP MOSQUITO LIFE HISTORY ###################
bioParameters=list(
  beta=20, popGrowth=1.175, muAd=.09,
  tEgg=5, tLarva=6, tPupa=4
)
# SET UP LANDSCAPE ################################
distancesMatrix=as.matrix(
  read.csv(
    "./GeoLandscapes/ATaleOfTwoCities_Distances.csv",
    sep=",", header=FALSE
  )
)
lifespanNonMigratoryProbability=.90
movementKernel=calcHurdleExpKernel(
  distancesMatrix,
  kernels$exp_rat,
  calcZeroInflation(
    lifespanNonMigratoryProbability,
    bioParameters$muAd
  )
)
patchPops=rep(50,sitesNumber)
```

Figure 1.5: **Code Sample 1**
Loading the package and setting up the life history and landscape modules.

```
# SET UP INHERITANCE / GENE DRIVE #################
### A. Replacement Drive
sH=sR=.25 # Fitness cost (lifetime reduction) of H & R allele
sB=.50 # Fitness cost (lifetime reduction) of B allele
eM=0.9 # Rate of accurate homology-directed repair in males
eF=0.5 # Rate of accurate homology-directed repair in females
driveCube=cubeHomingDrive(
  eM=eM,eF=eF,
  rM=(1/3)*(1-eM), bM=(2/3)*(1-eM),
  rF=(1/3)*(1-eF), bF=(1/3)*(1-eF),
  s=c(
    "WW"=1, "WH"=1-sH, "WR"=1-sR, "WB"=1-sB,
    "HH"=1-2*sH, "HR"=1-sH-sR, "HB"=1-sH-sB,
    "RR"=1-2*sR, "RB"=1-sR-sB,
    "BB"=1-2*sB
  ),
  eta=c(
    "WW"=1, "WH"=1-sH, "WR"=1-sR, "WB"=1-sB,
    "HH"=1-2*sH, "HR"=1-sH-sR, "HB"=1-sH-sB,
    "RR"=1-2*sR, "RB"=1-sR-sB,
    "BB"=1-2*sB
  )
)
### B. Suppression Drive
sHet=.9 # Fitness cost (lifetime reduction) in heterozygotes for H/B
eM=eF=0.999 # Rate of accurate homology-directed repair in males & females
driveCube=cubeHomingDrive(
  eM=eM,eF=eF,
  rM=(1/3)*(1-eM), bM=(2/3)*(1-eM),
  rF=(1/3)*(1-eF), bF=(1/3)*(1-eF),
  s=c(
    "WW"=1,"WH"=1-sHet,"WR"=1,"WB"=1-sHet,
    "HH"=0,"HR"=1-sHet,"HB"=0,
    "RR"=1,"RB"=1-sHet,
    "BB"=0
  )
)
# SET UP RELEASES ################################
patchReleases=replicate(
  n=sitesNumber,
  expr={list(maleReleases=NULL,femaleReleases=NULL)},
  simplify=FALSE
)
releasesParameters=list(
  releasesStart=100, releasesNumber=5,
  releasesInterval=2*(
    bioParameters$tEgg+bioParameters$tLarva+bioParameters$tPupa
  ),
  releaseProportion=2*round(mean(patchPops))
)
maleReleasesVector=generateReleaseVector(
  driveCube=driveCube,
  releasesParameters=releasesParameters,
  sex="M"
)
for(i in 6:6){patchReleases[[i]]$maleReleases=maleReleasesVector}
```

Figure 1.6: **Code Sample 2**
Setting up the inheritance/gene drive module and defining the release scheme. Here, code is shown for both: **A**) homing-based replacement drive, and **B**) suppression drive. Only one of these should be selected when running the simulation.

```
# PREPARE THE FOLDERS ############################
folderNames=list()
for(i in 1:repetitions){
  folderName=paste0(outputDirectory,str_pad(i,4,"left","0"))
  dir.create(folderName)
  folderNames=c(folderNames,folderName)
}
# RUN THE MODEL ##################################
for(i in 1:repetitions){
  outputFolder=folderNames[[i]]
  netPar=parameterizeMGDrivE(
    runID=i, simTime=simulationTime,
    nPatch=sitesNumber, beta=bioParameters$beta,
    muAd=bioParameters$muAd, popGrowth=bioParameters$popGrowth,
    tEgg=bioParameters$tEgg, tLarva=bioParameters$tLarva,
    tPupa=bioParameters$tPupa, AdPopEQ=patchPops
  )
  network=Network$new(
    networkParameters=netPar, driveCube=driveCube,
    patchReleases=patchReleases, migrationMale=movementKernel,
    migrationFemale=movementKernel, directory=outputFolder
  )
  network$oneRun()
  network$reset()
}
```

Figure 1.7: **Code Sample 3**
Preparing output folders and running the model. It is recommended to store simulation files for each run in its own separate folder.

1.1.3.2 Population Suppression

As a second example, we demonstrate the application of the MGDrivE package to a population suppression homing construct similar to that engineered by Hammond[80]. For this construct, the homing system targets a gene required for female fertility, causing females lacking the gene (those having the genotypes HH, HB and BB) to be infertile, and inducing a large fecundity reduction of 90% in females only having one functioning copy of the gene (those having the genotypes Hh, HR, hB and RB). The homing efficiency is very high – 99.9% in both males and females – with a third of the remaining h alleles in Hh individuals being converted R alleles and the remainder being converted to B alleles. This is similar to the first CRISPR-based homing construct demonstrated in *A. gambiae*, although with a higher homing efficiency that could be achieved through guide RNA multiplexing[142]. Lines of code that differ for this system are shown in Code sample 1.6. While the same inheritance cube applies, specific parameters differ – namely, homing and resistant allele generation rates, and genotype-specific fitness effects.

1.1.4 Output Analysis

In the current version of MGDrivE, complete simulation results are output as CSV files, two basic plotting functions are provided in R, and several functions are provided for aggregating the data – by population, genotype, or some combination thereof – as required by the question of interest.

In Figure 1.8, we display a potential visualization scheme produced in Mathematica for the simulations described above (additional videos are provided in the Supplementary Information: S1 Video and S2 Video). We depict allele count on the y-axis in the Figure 1.8a and b and allele frequency (depicted as colour density) in Figure 1.8c and d, with time on the horizontal axis. For population replacement (Figure 1.8a and c), we see the gene drive allele (H) spread through the population, and the in-frame resistant allele (R) accumulate to a small extent. This occurs because the R allele has neither a fitness cost nor benefit relative to the H allele once it has saturated the population, while the B allele is selected against due to its inherent selective disadvantage. Stochasticity slows these dynamics, on average, and introduces variability around the mean (Figure 1.8a).

For population suppression (Figure 1.8b and d), we see the gene drive system (H) spread through the population at the same time as it induces suppression due to its impact on female fertility. Eventually, we see an in-frame resistant allele (R) emerge and spread into the population due to its selective advantage over both the wild-type and homing alleles. In the deterministic model output, the in-frame resistant allele spreads to fixation; however in the stochastic model output, the homing allele is often lost from the population and, as a result, the selective advantage of the in-frame resistant allele is lost, causing it to equilibrate at a lower population frequency than in the deterministic simulation (in which it is never

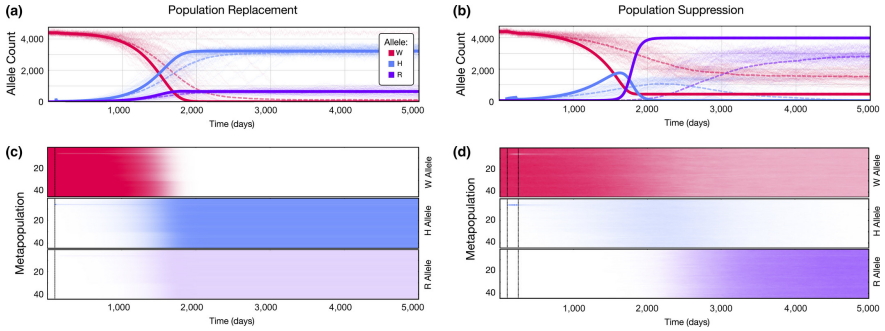

Figure 1.8: **Example MGDrivE Simulations for CRISPR-Based Homing Constructs**
In both cases, an Aedes aegypti population is simulated having the bionomic parameters in Table 1.2 and distributed through the landscape depicted in Figure 1.3. Deterministic simulations are denoted by solid lines in panels (**A**) and (**B**), while stochastic simulations are denoted by thin lines, each corresponding to the output of a single simulation, and dotted lines, corresponding to the mean of 100 stochastic simulations. (**A**) A population replacement homing construct that drives a disease-refractory gene into the population is simulated having a homing efficiency of 90% in males and 50% in females. Wild-type (h) alleles that are not converted to homing (H) alleles in the germline of Hh heterozygotes are cleaved and converted to either in-frame (R) or out-of-frame (B) resistant alleles. Female fecundity and male mating fitness are reduced by 25% per H or R allele and by 50% per B allele. A single release of 100 HH females at node 6 is modeled. As the homing allele (blue) is driven into the population, the wild-type allele (red) is eliminated, and the in-frame resistant allele (purple) accumulates to a population frequency of 17%. (**B**) A population suppression homing construct that interferes with a gene required for female fertility is simulated having a homing efficiency of 99.9% in both females and males. Wild-type alleles that are not converted to homing alleles in the germline of Hh heterozygotes are cleaved and converted to either in-frame or out-of-frame resistant alleles. Females without a copy of the h or R allele are infertile, while females having only one copy of the h or R allele have a 90% fecundity reduction. Five releases of 100 HH females at node 6 are modeled. As the homing allele (blue) is driven into the population, it suppresses the population due to its impact on female fertility. Eventually, an in-frame resistant allele (purple) emerges and leads the population to rebound due to its selective advantage over both wild-type and homing alleles. (**C**, **D**) Population frequencies of the wild-type, homing and in-frame resistant alleles are shown in each population over time for a deterministic model of the population replacement construct (panel **C**) and a stochastic simulation of the population suppression construct (panel **D**). Out-of-frame resistant alleles are omitted due to their low frequencies in both simulations. Dashed vertical lines represent the beginning and end of the releases.

lost). Stochasticity also significantly slows the mean allele frequency trajectories, as well as introducing variability around the mean. Mathematica and Python files to generate Figure 1.8 are provided.

1.1.5 Future Directions

We are continuing development of the MGDrivE software package, and welcome suggestions and requests from the research community regarding future directions. The field of gene drive has been moving very quickly, especially since the discovery of CRISPR-based gene editing, and we intend the MGDrivE package to provide a flexible tool to model novel inheritance-modifying constructs as they are proposed and become available. Future functionality will include: (a) 'shadow drive', in which the Cas9 enzyme is passed on to the offspring even if the gene expressing it is not[31], (b) life-history models incorporating a range of density-dependence relationships, and encompassing a more diverse range of insect disease vectors and agricultural pests, and (c) populations that vary in size seasonally or in response to environmental drivers such as temperature and rainfall. Incorporation of environmental drivers will allow both seasonal trends and short-term fluctuations to be accommodated within the same framework.

Additionally, we are developing a corresponding individual-based model that is capable of modelling multi-locus systems for which the number of possible genotypes exceeds the number of individuals in the population. This will enable us to efficiently model confineable systems such as daisy-drive involving several loci[156], and multiplexing schemes in which a single gene is targeted at multiple locations to reduce the rate of resistant allele formation[167].

1.2 MGDrivE 2: A Simulation Framework for Gene Drive Systems Incorporating Seasonality and epidemiological Dynamics[2]

1.2.1 Introduction

Interest in gene drive technology has continued to grow in recent years as a range of promising new constructs have been developed in the lab and discussions have moved towards implementing field trials in some cases. Recently developed systems include a CRISPR-based homing system intended for population suppression targeting the *doublesex* gene in *Anopheles gambiae*, the main African malaria vector[114], a split gene drive system intended for confinable and transient population replacement in *Aedes aegypti*, the main vector of dengue, chikungunya and Zika viruses[123], and CRISPR-based homing systems intended for population replacement in *An. gambiae*[24] and *Anopheles stephensi*, the main malaria vector in urban India[2].

As the technology advances and potential field trials are discussed[101], models are needed that incorporate additional ecological detail, including parameters that change over time in response to environmental variables such as temperature and rainfall, as well as models linking entomological and epidemiological outcomes[102]. Many insects, including mosquitoes, display a high degree of seasonality in their population dynamics, as development time from one life stage to another, and mortality rates associated with each life stage, vary with temperature and other environmental variables[150]. For *An. gambiae* and several other mosquito disease vectors, population size varies largely in response to recent rainfall, which creates pools of standing water and hence enhanced carrying capacity of the environment for mosquito larvae[202]. Seasonal changes in temperature and rainfall thus lead to seasonal changes in mosquito population density and consequent disease transmission, which must be accounted for in disease control strategies.

Models of disease transmission are also becoming increasingly relevant to models of gene drive dynamics, as: i) the readiness of a gene drive system for field trials will be determined in part by its expected (i.e. modeled) epidemiological impact, and ii) initial field trials are expected to have a measured entomological outcome alongside a modeled epidemiological outcome[101]. Given the potential for a non-localized gene drive system to spread broadly, it has been acknowledged that such constructs at the trial stage should be expected to cause a significant reduction in disease transmission, as even a confined trial could lead to wide-scale spread for an effective system[101]. Therefore, readiness for field trials should be determined by alignment with a target product profile (TPP) and/or list of preferred product characteristics (PPCs) that include expected impact on disease transmission[102]. Models that

[2]This work has been previously published[207]

incorporate both gene drive and epidemiological dynamics can account for local malaria or arboviral transmission dynamics and specify gene drive construct parameters that achieve the desired level of epidemiological control.

Previously, we developed the MGDrivE 1 modeling framework to model the population dynamics of a variety of genetics-based and biological control systems and their spread through spatially-explicit populations of mosquitoes, or insects having a similar life history[177]. Here, we present MGDrivE 2, which significantly improves upon the capabilities of MGDrivE 1 by addressing the above-mentioned considerations, namely: i) the ability of parameter values to change over time, and hence to model seasonal population dynamics, and ii) the incorporation of an epidemiology module that can accommodate pathogen transmission between humans and mosquitoes. Minor additional improvements have been made to the inheritance, life history and landscape modules of the framework to reflect advances in these fields; for instance, a more resolved understanding of maternal deposition of Cas protein for CRISPR-based gene drive systems has been incorporated[33]. Models in MGDrivE 2 are represented as a stochastic Petri net (SPN), which has both computational and architectural benefits: model specification is separate from simulation, models can be efficiently stored and updated in memory, and a wealth of fast simulation algorithms from other fields can be used[69].

In this paper, we describe the key developments implemented in MGDrivE 2. We then demonstrate the application of the framework to the disease control impact of a CRISPR-based split gene drive system intended to drive a disease-refractory gene into a population in a confinable and reversible manner, and conclude with a discussion of future needs and applications for simulation packages in the field of gene drive modeling.

1.2.2 Design and Implementation

MGDrivE 2 is a significant extension of and development from MGDrivE 1, a model for the spread of gene drive systems through spatially-explicit mosquito populations. The MGDrivE 2 model incorporates: i) an "inheritance module" that describes the distribution of offspring genotypes for given maternal and paternal genotypes, ii) a "life history module" that describes the development of mosquitoes from egg to larva to pupa to adult, iii) a "landscape module" that describes the distribution and movement of mosquitoes through a metapopulation, and iv) an "epidemiology module" that describes pathogen transmission between mosquitoes and humans (Fig. 1.9). The framework is formulated as a SPN that can be mapped to a continuous-time Markov process in which model parameters may vary over time. It can also be implemented as a deterministic model via mean-field approximation of the stochastic model[18].

The core framework is developed in R (https://www.r-project.org/). The SPN framework enables separation of model components, allowing users to modify code on a component-

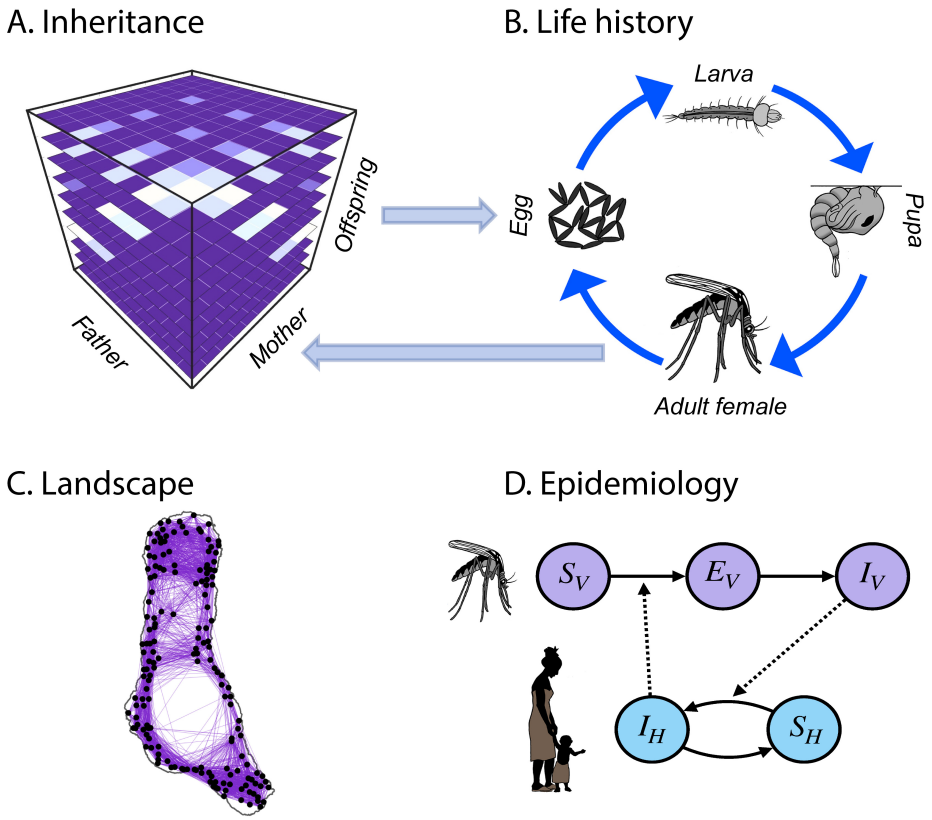

Figure 1.9: **Modules in the MGDrivE 2 Framework**
(**A**) Genetic inheritance is embodied by a three-dimensional tensor referred to as an "inheritance cube." Maternal and paternal genotypes are depicted on the x and y-axes and offspring genotypes on the z-axis. (**B**) Mosquito life history is modeled according to an egg-larva-pupa-adult (female and male) life cycle in which density dependence occurs at the larval stage, and life cycle parameters may vary as a function of environmental variables over time. Genotypes are tracked across all life stages, and females obtain a composite genotype upon mating - their own and that of the male they mate with. Egg genotypes are determined by the inheritance cube. (**C**) The landscape represents a metapopulation in which mosquitoes are distributed across population nodes and move between them according to a dispersal kernel. Population sizes and movement rates may vary as a function of environmental variables. (**D**) The epidemiology module describes reciprocal transmission of a vector-borne pathogen between mosquitoes and humans. This requires modeling human as well as mosquito populations, and the number of individuals having each infectious state. Epidemiological parameters may vary as a function of environmental variables.

by-component basis as needed for model specification or computational speed. We now describe the model extensions and developments from MGDrivE 1 to 2 in more detail. Full

details of the MGDrivE 2 model framework are provided in the S1 Text.

1.2.2.1 Time-Dependent Parameters and Seasonality

The incorporation of time-dependent parameters represents a significant improvement of the MGDrivE 2 modeling framework. In MGDrivE 1, the mosquito life history module follows the lumped age-class model of Hancock and Godfray as adapted by Deredec *et al.*[46], which describes development from egg to larva to pupa to adult using delay-difference equations. The delay framework allows development times to be modeled as fixed rather than exponentially-distributed; however, it is not compatible with time-varying parameters as these could vary during the delay. In MGDrivE 2, the discrete-time, fixed-delay framework of MGDrivE 1 is replaced by a continuous-time implementation in which each life stage is divided into a series of substages. For a single substage, the development time is exponentially-distributed; but as the number of substages increases, the distribution of development times becomes concentrated around the mean. Specifically, if a life stage with a mean development time of $\frac{1}{d}$ is divided into a series of n substages, the new development times are Erlang-distributed with mean, $\frac{1}{d}$, and variance, $\frac{1}{d \cdot n^2}$, or equivalently, with shape parameter, n, and rate parameter, $\frac{d}{n}$. The development time, $d(t)$, may also vary over time, t; however the number of substages, n, and hence the mean-variance relationship for development times, must remain constant within a simulation.

Most importantly, the new model implementation allows any model parameter to vary with time, enabling the framework to account for seasonal variation in development times and mortality rates due to environmental dependencies. Temperature is known to strongly influence development times for juvenile mosquito stages, and mortality rates for all mosquito life stages[150, 13], and rainfall is known to influence the carrying capacity of the environment for larvae, and therefore density-dependent larval mortality rates[202, 152]. The new model formulation allows these parameters to vary in continuous time in response to environmental data, and hence for seasonal variations in temperature and rainfall to drive seasonal variations in mosquito population density.

Parameters defining other modules of the model - inheritance, landscape and epidemiology - are also able to vary over time within the new model formulation. For instance, gene drive systems under the control of temperature-dependent promoters[209, 48] may have time-varying homing efficiencies, mosquito movement rates may vary seasonally in response to temperature and other environmental factors[116], and epidemiological parameters such as the extrinsic incubation period (EIP) and pathogen transmission probabilities from human-to-mosquito and mosquito-to-human are all known to display seasonal variation through temperature dependence[150, 13].

1.2.2.2 Epidemiology Module

The epidemiology module describes reciprocal transmission of a vector-borne pathogen between mosquitoes and humans. This requires modeling of both vector and human populations, as well as an attribute describing the number of individuals in the vector and human populations having each infectious state (Fig. 1.10). To model malaria, the Ross-Macdonald model is included, which has susceptible (S_V), exposed/latently infected (E_V), and infectious (I_V) states for mosquitoes, and susceptible (S_H), and infected/infectious (I_H) states for humans[175, 130]. Malaria infection in humans is described by an SIS model, in which humans become infected at a per-capita rate equal to the "force of infection" in humans, λ_H, and recover at a rate, r. Malaria infection in mosquitoes is described by an SEI model, in which adult mosquitoes emerge from pupae in the susceptible state, become exposed and latently infected at a per-capita rate equal to the force of infection in mosquitoes, λ_V, and progress to infectiousness at a rate equal to γ_V. The force of infection in humans, λ_H, is proportional to the number of mosquitoes that are infectious, I_V and the force of infection in mosquitoes, λ_V, is proportional to the fraction of humans that are infectious, $\frac{I_H}{N_H}$, where N_H is the human population size. Since an exponentially-distributed EIP leads to some mosquitoes having unrealistically brief incubation periods, we divide the E_V state into a series of n sub-states, as described in section 1.2.2.1, leading to the EIP being Erlang-distributed with shape parameter, n, and rate parameter, $\frac{\gamma_V}{n}$[186]. Finally, transmission parameters may be tied to specific mosquito genotypes - for instance, an antimalarial effector gene may be associated with a human-to-mosquito or mosquito-to-human transmission probability of zero.

To model arboviruses such as chikungunya, Zika and single serotypes of dengue virus, we include an SEIR model for human transmission, in which the human states are: susceptible (S_H), exposed/latently infected (E_H), infectious (I_H), and removed/recovered (R_H)[171, 113]. The E_H and R_H states are included because arboviruses are generally thought to be immunizing, and have latent periods that tend to be on a similar timescale to the duration of infectiousness. Humans become latently infected at a per-capita rate equal to λ_H, progress to infectiousness at rate, γ_H, and recover at rate, r. For mosquito transmission, the SEI model with an Erlang-distributed EIP is used again. Further details of the mathematical formulation of both the malaria and arbovirus models are provided in the S1 Text. The extensibility of the SPN framework means that more complex epidemiological models can be developed and implemented by users.

Modeling vector-borne disease transmission within a metapopulation framework generally requires each population node in the network to have both a defined mosquito and human population size. Since the mosquito vectors we are interested in are anthropophilic, they tend to coexist with humans, so human population sizes and state distributions can be attributed to the same nodes at which mosquito populations are defined; however, MGDrivE 2 also includes the possibility of human-only and mosquito-only nodes. Mosquito-only nodes could represent sites with only non-human vertebrates from which mosquitoes bloodfeed, while

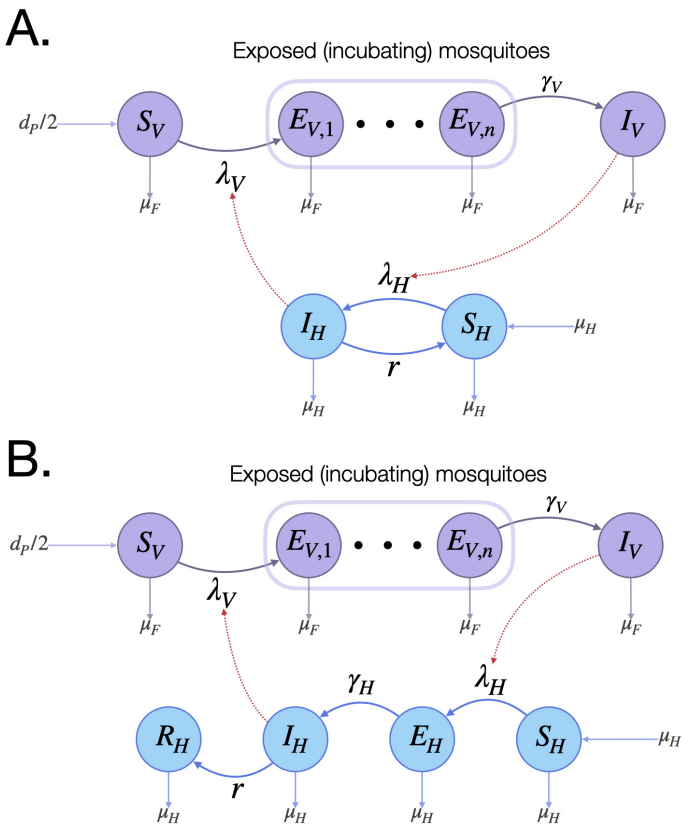

Figure 1.10: **Epidemiology Module**
MGDrivE 2 includes two basic models for reciprocal pathogen transmission between mosquitoes and humans - one for malaria (**A**), and one for arboviruses (**B**). In both cases, female mosquitoes emerge from pupae at a rate equal to $\frac{d_P}{2}$ as susceptible adults (S_V), become exposed/latently infected ($E_{V,1}$) at a rate equal to the force of infection in mosquitoes, λ_V, and progress to infectiousness (I_V) through the extrinsic incubation period (EIP $= \frac{1}{\gamma_V}$), which is divided into n bins to give an Erlang-distributed dwell time. The mortality rate, μ_F, is the same for female mosquitoes in each of these states. For malaria (**A**), susceptible humans (S_H) become infected/infectious (I_H) at a rate equal to the force of infection in humans, λ_H, and recover at rate r, becoming susceptible again. For arboviruses (**B**), susceptible humans (S_H) become exposed/latently infected (E_H) at a rate equal to λ_H, progress to infectiousness (I_H) at rate equal to γ_H, and recover (R_H) at rate, r. Infection dynamics couple the mosquito and human systems via the force of infection terms; λ_V is a function of I_H, and λ_H is a function of I_V, shown via red edges.

human-only nodes could represent locations unsuitable for mosquitoes. As mosquitoes are able to move between nodes in the metapopulation, so can humans. This is an important

factor to include, as human movement has been shown to drive the spatial transmission of mosquito-borne diseases such as dengue virus[189].

1.2.2.3 Other Extensions to Inheritance, Life History and Landscape Modules

Additional functionality has been included in the inheritance and life history modules of the MGDrivE framework since publication of version 1.0. The inheritance module is unchanged, and inheritance "cubes", describing the distribution of offspring genotypes given maternal and paternal genotypes for a given genetic element, are usable in both versions. Several new inheritance cubes have been made available, including: a) homing-based remediation systems, including ERACR (Element for Reversing the Autocatalytic Chain Reaction) and e-CHACR (Erasing Construct Hitchhiking on the Autocatalytic Chain Reaction)[60, 208], and b) newly proposed drive systems capable of regional population replacement, including CleaveR (Cleave and Rescue)[159] and TARE (Toxin-Antidote Recessive Embryo) drive[29].

In the life history module, we have provided two alternative parameterizations of a quadratic density-dependent larval mortality rate function corresponding to logistic and Lotka-Volterra ecological models. For mosquito vectors such as *Ae. aegypti* and *An. gambiae*, density-dependence is thought to act at the larval stage due to increased resource competition at higher larval densities[202, 152]. The adult population size, N, is used to determine the value of K, the larval density at which the larval mortality rate is twice the density-independent mortality rate at a given patch, which produces the appropriate equilibrium population size. For the logistic model, the per-capita larval mortality rate is given by $\mu_L \cdot \frac{1+L(t)}{K}$, where μ_L is the density-independent larval mortality rate, and $L(t)$ is the total larval population size for the patch at time t. For the Lotka-Volterra model, the per-capita larval mortality rate is given by $\mu_L + \alpha \cdot L(t)$, where α is the density-dependent term. While related by the expression, $\alpha = \frac{\mu_L}{K}$, these two models provide an example of how different functional forms can be used for rates in MGDrivE 2, and may serve as a template for incorporating more elaborate density-dependent functions.

In the landscape module, movement through the network of population nodes is again determined by a dispersal kernel; however, due to the continuous-time nature of MGDrivE 2, movement between patches is described by a rate rather than a probability. MGDrivE 2 provides functions to map transition probability matrices from MGDrivE 1, such as the zero-inflated exponential or lognormal dispersal kernels, to continuous-time transition rate matrices for MGDrivE 2. These mapping functions may also be applied to transition probability matrices derived from empirical or simulated data. The mathematical mapping between the rate matrix of MGDrivE 2 and the transition probability matrix of MGDrivE 1 is provided in the S1 Text.

1.2.2.4 Stochastic Petri Net Formulation

The most fundamental change from MGDrivE 1 to 2 is restructuring the model as a SPN[76]. Adopting a SPN framework has several benefits. First, SPNs allow the mathematical specification of a model to be decoupled from its algorithmic implementation, allowing users to leverage extensive sampling algorithms from the physical and chemical simulation communities for efficient computation[69, 65]. Second, SPNs have a well-established and consistent formalism, allowing them to be readily understood and modified by anyone familiar with this[73]. And third, SPNs are isomorphic to continuous-time Markov chains (CTMCs), meaning that model parameters can be time-varying, including Erlang-distributed aquatic stage durations and the pathogen's EIP.

A Petri net is a bipartite graph consisting of a set of places, P, and a set of transitions, T. Directed edges or "arcs" lead from places to transitions (input arcs) and from transitions to places (output arcs). The set of arcs that connect places to transitions and transitions to places can be denoted by two matrices whose entries are non-negative integers describing the weight of each arc. The places define the allowable state space of the model; however, in order to describe any particular state of the model, the Petri net must be given a marking, M, which is defined by associating each place with a non-negative integer number of tokens. In the language of CTMCs, M is referred to as a "state". When a transition occurs, it induces a state change by "consuming" tokens in M given by the set of input arcs, and "producing" tokens in M according to the set of output arcs[203]. Each transition has a "clock process," parameterized by a "hazard function" which defines that event's current rate of occurrence. In MGDrivE 2, tokens represent an integer number of mosquitoes or humans, and the distribution of tokens (mosquitoes or humans) across states at time t defines a marking, $M(t)$. A graphical representation of a Petri net for the mosquito life history module of MGDrivE 2 is depicted in Fig. 1.11A, with a full description of the mathematical formalism provided in the S1 Text.

The code that generates the Petri net is independent of the code that simulates trajectories from it. Once the Petri net is stored as a set of sparse matrices, it is passed to a simulation application program interface (API) which allows trajectories to be simulated as ordinary differential equations (ODEs), stochastic differential equations (SDEs), or CTMCs (Fig 1.11B). Each of these are referred to as "step" functions, but are not limited to discrete time steps; these functions are responsible for updating the model between time points where the user requests output to be recorded. The ODE step function provides a deterministic approximation and interfaces with the numerical routines provided in the "deSolve" R package[187]. Three stochastic numerical routines are provided that treat the model as a continuous-time Markov process and provide different levels of approximation. The most straightforward method to sample trajectories is Gillespie's direct method, which samples each event individually[64]. While statistically exact, this is prohibitively slow for medium-to-large population sizes. Two approximate stochastic methods are provided

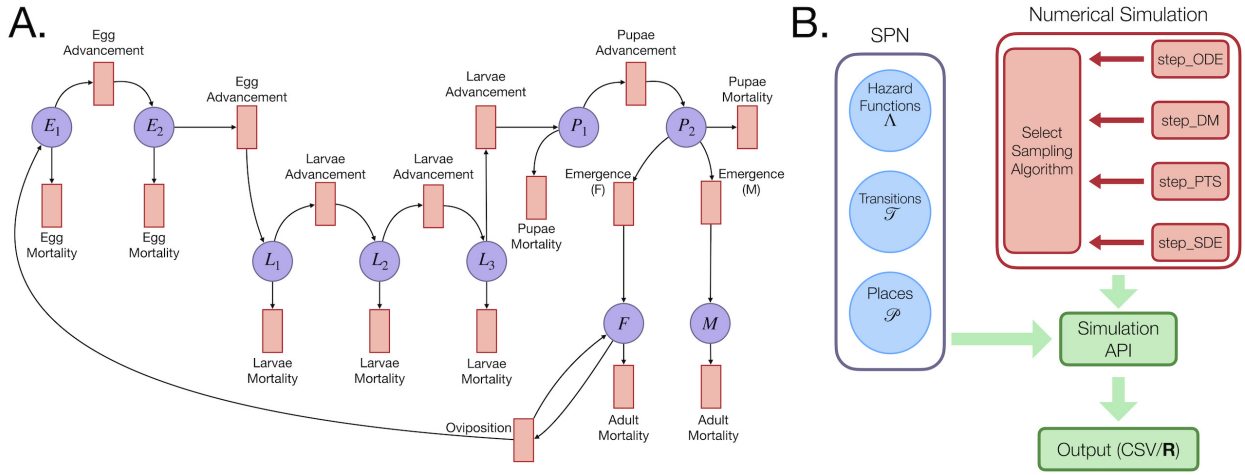

Figure 1.11: **Stochastic Petri Net (SPN) Implementation of MGDrivE 2** (**A**) Petri net representation of the life history module. The set of purple circles corresponds to places, P, and red rectangles to transitions, T. This Petri net shows a model in which development times for the egg stage are Erlang-distributed with shape parameter $n = 2$, and for the larval stage are Erlang-distributed with shape parameter $n = 3$. Population dynamics are derived directly from this graph. E.g. The transition corresponding to oviposition has one edge beginning at F, meaning at least one female mosquito must be present for oviposition to occur. When oviposition occurs, a token is added to E_1 (new eggs are laid) and a token is returned to F. (**B**) Conceptual representation of the SPN software architecture showing the separation between the model representation (blue circles) and set of sampling algorithms (red rectangles). These two components of the codebase meet at the simulation API, enabling users to match models and simulation algorithms interchangeably. Output may be returned as an array in R for exploratory work, or written to CSV files for large simulations.

that have been widely used in the chemical physics literature: i) a second order continuous SDE approximation known as the chemical Langevin equation[66], and ii) a fixed-step tau-leaping method[63]. Both methods achieve substantial gains in computational speed at the expense of statistical accuracy. While the SDE approximation is often faster, tau-leaping retains the discrete character of the process it approximates and is usually the preferred technique. A full description of each of the numerical routines is provided in the S1 Text. In addition, we demonstrate how a user can write a custom simulation algorithm and incorporate it within the MGDrivE 2 codebase in the "Advanced Topics" vignette available at https://marshalllab.github.io/MGDrivE/docs_v2/articles/advanced_topics.html.

1.2.3 Results

To demonstrate how the MGDrivE 2 framework can be used to initialize and run a simulation of the spread of a gene drive system through a metapopulation with time-varying model parameters, including its implications for vector-borne pathogen transmission, we have provided vignettes with the package, available via installation from CRAN at https://CRAN.R-project.org/package=MGDrivE2 and additional examples and information on GitHub at https://marshalllab.github.io/MGDrivE/docs_v2/index.html. The vignettes provide extensive examples of how to use the software, including advanced features such as implementing custom time-varying rates and numerical simulation algorithms. They consist of a set of five "core" manuals that describe how to simulate population genetics and dynamics for a mosquito-only population and metapopulation, then how to incorporate SEI-SIS Ross-Macdonald malaria transmission dynamics in a population with humans included, and finally how to incorporate SEI-SEIR arbovirus transmission dynamics. Following these are three "advanced" manuals that introduce: i) how to process and analyze output from simulations that write to CSV files, ii) how users can write custom time-varying hazard functions, and iii) how a user might implement their own numerical simulation routine, using an explicit Euler method for ODEs as an example.

Here, we describe the application of the package to model the release of a split gene drive system designed to drive a malaria-refractory gene into an *An. gambiae* mosquito population with seasonal population dynamics and transmission intensity calibrated to a setting resembling the island of Grand Comore, Union of the Comoros. The split gene drive system resembles one engineered in *Ae. aegypti*[123], which was chosen as the only published split drive system in a mosquito vector to date. Split drive designs are well-suited to initial field trials of gene drive systems as they display transient drive activity before being eliminated by virtue of a fitness cost. The spatial spread of these systems is limited by the distance the host organism disperses while the drive system persists. In the split drive system explored here, two components - the Cas9 and guide RNA (gRNA) - are present at separate, unlinked loci, and a disease-refractory gene is linked to the gRNA. We assume that only one copy of the disease-refractory allele is required for it to block pathogen transmission. Four alleles are considered at the gRNA locus: an intact gRNA/refractory allele (denoted by "H"), a wild-type allele (denoted by "W"), a functional, cost-free resistant allele (denoted by "R"), and a non-functional or otherwise costly resistant allele (denoted by "B"). At the Cas9 locus, two alleles are considered: an intact Cas9 allele (denoted by "C"), and a wild-type allele (denoted by "W"). Full details of the inheritance dynamics are provided in Li *et al.*[123] and model parameters are summarized in S1 Table.

The life history module is parameterized with typical bionomic parameter values for *An. gambiae* (S1 Table), including mean-variance relationships describing the development times of juvenile life stages[11]. The carrying capacity of the environment for larvae is a function of recent rainfall, and the adult mortality rate is a function of temperature. Re-

motely sensed rainfall data for Grand Comore was obtained from the ERA5 dataset (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5), and a mathematical relationship adapted from White *et al.*[202] was used to translate this to larval carrying capacity, assuming that half of the island's carrying capacity was provided by permanent breeding sites (e.g. large cisterns) and half was provided by recent rainfall. Temperature data for Grand Comore was also obtained from the ERA5 dataset, and adult mortality was derived using methods described by Mordecai *et al.*[150]. Both climatological time series covered the six-year period beginning January 1, 2010. For the purpose of this demonstration, Grand Comore was treated as a single randomly mixing population, although simulations involving a more detailed landscape module are included in the vignettes.

The epidemiology module is parameterized with typical parameter values for *Plasmodium falciparum* transmission (S1 Table), human population size and life expectancy parameters from the National Institute of Statistics and Demographic Studies, Comoros[98], and is calibrated to local malaria prevalence estimates from the Malaria Atlas Project[164]. This calibration was achieved by multiplying the carrying capacity time series by a constant such that the average adult female mosquito population over a year sustained malaria transmission in the human population at the estimated local prevalence. Finally, we caution that these simulations are merely intended to demonstrate the software's capabilities and that, while the simulations are parameterized with data from Grand Comore, they are not intended to provide an accurate forecast of local gene drive mosquito dynamics, or to imply approval of the intervention by the local population and regulatory agencies.

1.2.3.1 Simulation Workflow

The code for this simulation is available at https://github.com/MarshallLab/MGDrivE/tree/master/Examples/SoftwarePaper2. We begin by loading the MGDrivE 2 package in R, as well as the package for the original MGDrivE simulation, which provides the inheritance cubes required for simulation of genetically-stratified mosquito populations. Next, we define model parameters, including the bionomic parameters of *An. gambiae* s.l., and demographic and epidemiological parameters specific to Grande Comore. To parameterize time-varying adult mosquito mortality (hourly) and larval carrying capacity (daily), we load CSV files containing those data as time series for the ten-year simulation period. We then use the base "stepFun()" function in R to create an interpolating function of those time-series data that will return a value for any time within the simulation period, which is required for calculation of hazard functions. More sophisticated interpolating functions, such as splines, may also be used. We also specify the inheritance cube at this point, as the number of modeled genotypes and distribution of offspring genotypes for given parental genotypes will be used to build the Petri net.

Next, we use functions from MGDrivE 2 to create the "places" and "transitions" of the Petri net, which are stored as lists in R and then converted into a sparse matrix representation

used in the simulation code. Epidemiological dynamics and states are coded automatically by calling the functions that create the Petri net. In this case, "spn_P_epiSIS_node()" and "spn_T_epiSIS_node()" will generate the places and transitions for a single node model with SEI-SIS mosquito and human malaria transmission dynamics. Each transition has a tag that specifies the hazard function it requires. Following that, we write custom time-varying hazard functions for adult mosquito mortality and larval mortality (a function of carrying capacity). We provide a guided walkthrough of how a new user might write their own time-varying hazard function in the vignette "Simulation of Time-inhomogeneous Stochastic Processes." Once the vector of hazard functions has been stored (as a list), we create the data frame that stores the times, genotypes, sex, and size of each release event.

With the construction of all model components necessary for the simulation, we call the simulation API which handles the details of simulating trajectories from the model. In this case, we chose the tau-leaping algorithm to sample stochastic trajectories, and to record output on a daily basis. MGDrivE 2 allows users to choose how model output is reported back - for exploratory or smaller simulations, users may return output directly to R as an array; however for larger simulations, it is often preferable to write directly to CSV files due to memory considerations, and MGDrivE 2 has sophisticated functions to both specify CSV output and process completed simulations.

1.2.3.2 Entomological Population Dynamics

In Figure 1.12, we display a potential visualization scheme produced in Python for the simulations described above. The code to produce this visualization is available at https://github.com/Chipdelmal/MoNeT/tree/master/DataAnalysis/v2 (note that MGDrivE 2 code does not depend on Python). Figure 1.12A displays the climatological time-series data — temperature in magenta and rainfall in blue — which were used to calculate time-varying adult mosquito mortality rate and larval carrying capacity, respectively. The total adult female population size averaged over 100 stochastic runs is shown in green. This is relatively consistent throughout the year due to moderate seasonal changes in temperature in the tropical climate of the Comoros and the presence of permanent breeding sites such as cisterns throughout the island; however population spikes are observed after significant rainfall. Figure 1.12B displays allele frequencies for adult female mosquitoes over the simulation period. After eight consecutive weekly releases of 50,000 male mosquitoes homozygous for both the Cas9 (C) and gRNA/refractory (H) alleles three years into the simulation, we see the C and H alleles accumulate to high post-release frequencies, and the H allele continue to spread to a higher frequency over the subsequent ~ 6 months while the H and C alleles regularly co-occur enabling drive to occur at the gRNA locus. The wild-type allele (W) at the gRNA locus is almost completely lost over this period, and both in-frame and out-of-frame resistant alleles (R and B, respectively) accumulate to a small yet significant extent. The C allele slowly declines in frequency following the releases due to a fitness cost; and beginning about 1 year after the releases, the H allele gradually declines in frequency as its fitness cost

begins to outweigh its inheritance bias. The declines in C and H allele frequencies continue beyond the simulated timeframe, although not before the H allele has a chance to interfere with disease transmission.

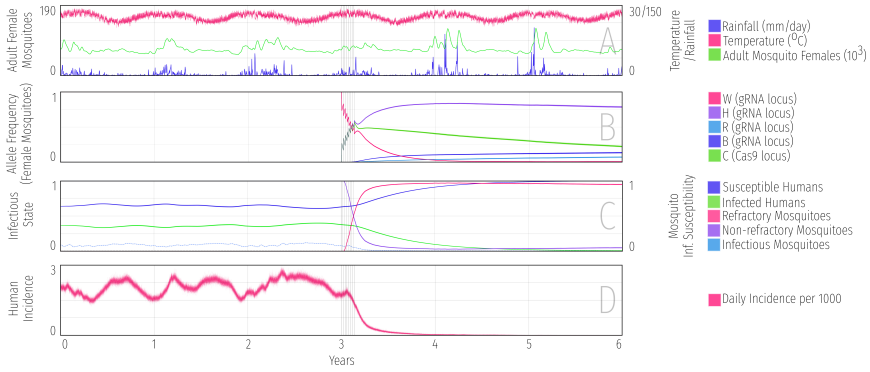

Figure 1.12: **Example MGDrivE 2 Simulations**
Simulations for a split gene drive system designed to drive a malaria-refractory gene in a confinable and reversible manner into an *An. gambiae* s.l. mosquito population with seasonal population dynamics and transmission intensity calibrated to a setting resembling the island of Grand Comore, Union of the Comoros. The split drive system resembles one recently engineered in *Ae. aegypti* [123] – the only split drive system in a mosquito vector to date. In the modeled system, two components – the Cas9 and guide RNA (gRNA) – are present at separate, unlinked loci, and a disease-refractory gene is linked to the gRNA. Four alleles are considered at the gRNA locus: an intact gRNA/refractory allele (denoted by "H"), a wild-type allele (denoted by "W"), a functional, cost-free resistant allele (denoted by "R"), and a non-functional or otherwise costly resistant allele (denoted by "B"). At the Cas9 locus, two alleles are considered: an intact Cas9 allele (denoted by "C"), and a wild-type allele (denoted by "W"). Model parameters describing the construct, mosquito bionomics and malaria transmission are summarized in S1 Table. (**A**) Climatological time-series data - temperature in red and rainfall in blue - that were used to calculate time-varying adult mosquito mortality rate and larval carrying capacity, respectively. The resulting adult female population size is shown in green. (**B**) Allele frequencies for adult female mosquitoes over the simulation period. Grey vertical bars beginning at year three denote eight consecutive weekly releases of 50,000 male mosquitoes homozygous for both the gRNA and Cas9 alleles (H and C, respectively). (**C**) Spread of the malaria-refractory trait through the female mosquito population, and consequences for mosquito and human infection status. Following releases of the drive system at year three, the proportion of refractory female mosquitoes (solid red line) increases and the proportion of infectious mosquitoes (dotted light blue line) declines. As humans recover from infection and less develop new infections, the *P. falciparum* parasite rate (solid green line) declines until it reaches near undetectable levels by year five. (**D**) Human malaria incidence is halted by the beginning of year four.

1.2.3.3 Epidemiological Dynamics

The split drive system we consider includes a malaria-refractory gene that results in complete inability of mosquitoes to become infected with the malaria parasite, whether present in either one or two allele copies. In Figure 1.12C, we depict the spread of the malaria-refractory trait through the female mosquito population, and the consequences this has for mosquito and human infection status. Prior to the release, we see that infection prevalence in humans (*P. falciparum* parasite rate, PfPR) is mildly seasonal, with the proportion of infected humans (solid green line) waxing and waning in response to the fluctuating mosquito population size (green line in Fig. 1.12A). The proportion of infectious female mosquitoes (dotted light blue line) oscillates in synchrony with the proportion of infected humans; but at a much lower proportion due to the short mosquito lifespan and the fact that most mosquitoes die before the parasite completes its EIP. Following releases of the split drive system and refractory gene at year three, the proportion of refractory female mosquitoes (red line) increases and, consequently, the proportion of infectious mosquitoes declines. As humans recover from infection and less develop new infections, the PfPR declines until it reaches near undetectable levels by year five. Lastly, Figure 1.12D depicts human malaria incidence, measured as the number of new infections per 1,000 humans per day. Stochastic variation in this model output is more pronounced due to the small number of incident cases relative to the total population. Incidence is halted by the beginning of year four, but PfPR takes almost a year longer to approach zero as infected humans clear parasites.

1.2.4 Availability and Future Directions

MGDrivE 2 is available at `https://CRAN.R-project.org/package=MGDrivE2`. The source code is under the GPL3 License and is free for other groups to modify and extend as needed. Mathematical details of the model formulation are available in the S1 Text, and documentation for all MGDrivE 2 functions, including vignettes, are available at the project's website at `https://marshalllab.github.io/MGDrivE/docs_v2/index.html`. To run the software, we recommend using R version 2.10 or higher.

We are continuing development of the MGDrivE 2 software package and welcome suggestions and requests from the research community regarding future directions. The field of gene drive research is moving quickly, and we intend the MGDrivE 2 framework to serve as a flexible tool to address exploratory, logistical and operational questions regarding genetics-based control systems for mosquito disease vectors. This includes exploratory modeling of novel genetic constructs, assessment of candidate constructs against TPPs and PPCs, and field trial planning as constructs progress through the development pipeline. Current functionality presents a new opportunity to explore modeling-based research topics such as the invasiveness of threshold-dependent drive systems in the presence of climate fluctuations, seasonal source-sink dynamics and evolution towards smaller fitness costs. Future functionality that we are planning includes: i) modeling of mosquito traps to address questions related to

monitoring and surveillance, and ii) more detailed epidemiological models addressing phenomena important to malaria and arbovirus transmission - for instance, dengue models that incorporate multiple serotypes with temporary cross-protective immunity and complications related to antibody-dependent enhancement[201], and malaria models that incorporate age-structure, immunity, asymptomatic infection and superinfection[72].

Additionally, we are exploring numerical sampling algorithms that can increase computational efficiency and speed, facilitated by separation of model specification and simulation in the software. The complexity of models that can be developed in MGDrivE 2 means that sensitivity analyses can become extremely computationally intensive, and the ability of the SPN framework to leverage efficient algorithms in these circumstances will be highly valuable. We also continue to be interested in developing a corresponding individual-based model capable of efficient modeling when the number of possible states exceeds the number of individuals in the population - for instance, for multi-locus systems such as daisy-drive[155] and multiplexing schemes in which a single gene is targeted at multiple locations to reduce the rate of resistance allele formation[167], and for epidemiological models in which age structure, immunity and mosquito biting heterogeneity become prohibitive for population models[72].

As gene drive technology matures, potential species of interest are not limited to arthropod vectors of disease. In addition to public health applications, gene drive has been proposed as a technique to help address problems in agriculture and conservation, with target species including insect agricultural pests[20] and invasive rodents that predate native birds[119]. While MGDrivE 1 was not designed with non-arthropod species in mind, it has been adapted for application to invasive rodents on islands[115]. We expect MGDrivE 2 to be easier to adapt to other insect or mammalian species due to the separation of model specification from simulation, meaning there is no need to adapt code to numerically simulate trajectories. To adapt model specification, the set of places and transitions for the SPN will need to be updated according to the life stages and development and mortality rates for the species of interest. New SPN transitions may be needed for behaviors not currently included, such as multiple mating in adults, while redundant transitions may be removed.

Conclusion

Computationally efficient and biologically representative models are one cornerstone to the complete study of gene drives and their potential to reduce disease burden. Here, we provide a highly-efficient model accurately reflecting the complex genetic and landscape components necessary to describe gene drive dynamics, MGDrive (section 1.1), and then extend that model to account for variable ecological dynamics and directly explore the impact on disease burden. However, a model's validity is only partially based on the structural implementation, the rest comes from accurate and experimentally informed parameterization. Towards that end, we adapt our models to determine biologically-informed parameters from experimental results.

Chapter 2

The Principled Extraction of Bionomic Parameters Material to the Success of Gene Drives

Introduction

The previous chapter introduced novel mathematical models to simulate the dynamics of gene drives. However, there are several steps along the path from development to application of novel, and potentially dangerous, technologies [102]. New designs are first tested in small-cage trials within a lab[165], followed by larger trials in out-door, but enclosed, cages[79], before finally being applied to isolated field trials[135]. At each stage of testing, it is important to distill all possible information to instruct subsequent trials and forecast possible outcomes if applied more generally.

Towards this end, we have collaborated with experimentalists developing novel gene drive constructs to extract relevant information from small cage trials for accurate design and parameterization of our models. We have adapted our original software, MGDrivE[177], to reflect the discrete-generation, small-population size cages that are used for early experimentation. This approach benefits from the flexibility of genetic constructs built into our model and provides consistency from parameter estimation to full-scale simulation.

In this chapter, we extend previous work from us[21] and others[114] by applying our reduced model within a Bayesian framework for statistically robust estimation of properties impacting gene drive performance.

2.1 Efficient Population Modification Gene-Drive Rescue System in the Malaria Mosquito *Anopheles stephensi*[1]

2.1.1 Introduction

Challenges faced by current programs to eliminate malaria from high-endemic areas[198] have fostered the development of novel control strategies including those based on genetically modified mosquitoes. These genetic approaches are supported by the development of Cas9/guide RNA-(Cas9/gRNA) based gene-drives[60, 131, 61], with pioneering studies demonstrating effective mosquito population suppression[80, 114, 183] and modification[62, 165, 24], the latter aimed at impairing the ability of adult females to transmit the *Plasmodium* parasites causing the disease.

One population modification approach is to express multiple anti-parasite effector molecules in the form of single-chain antibodies directed against sexual stages of *Plasmodium falciparum* in relevant mosquito tissues[62, 99, 100]. This strategy relies on efficient copying of a drive system from one chromosome to its homolog in the germline by homology-directed repair (HDR) following the induction of double-stranded DNA breaks by the Cas9 endonuclease. This efficient repair process results in highly biased transmission of the drive element and the rapid spread of the parasite-refractory traits throughout a naive population. However, competing end-joining (EJ) repair mechanisms can result in insertions or deletions (INDELs) that may be resistant to subsequent Cas9-mediated cleavage and may impede the copying process[165, 82]. These resistant alleles can be generated at different developmental stages and tissues, including the germline and somatic tissues during early embryogenesis, due to parental deposition of Cas9/gRNA complexes or somatic expression from "leaky" promoters[62, 165, 12, 31, 33]. INDELs at the cut site of a coding region of a gene often disrupt subsequent protein function (nonfunctional EJ), but mutations that maintain protein function (functional EJ) also can occur. Both outcomes can affect the introduction dynamics of the drive element, and in the absence of strategies that prevent or mitigate EJ evens[33, 142, 81], gene-drives can cease to spread and be eliminated from a population if they carry a fitness cost[165, 82].

One resistance-mitigation strategy for population modification approaches is to design a drive that targets an essential gene while concomitantly providing a recoded rescue sequence[54]. Under these conditions, EJ resistant alleles that lack the rescue sequence are eliminated from the population as the result of the fitness load associated with the loss-of-function of the targeted gene. The feasibility of "rescue systems", such as Cleave and Rescue (CleaveR) and toxin-antidote, was demonstrated in the fruit fly, *Drosophila melanogaster*, by target-

[1] This work has been previously published[2]

ing essential genes using alternative combinations of *trans*- or *cis*-acting drive elements and rescue sequences[159, 29, 160]. These systems are independent of HDR copying of the drive and result in threshold-dependent drive invasion dynamics[159, 29]. When arranged in a configuration where the Cas9 and gRNAs are not located in the same transgenic construct, these systems are predicted to remain confined locally[29, 160]. Alternative designs based on HDR-mediated spread of rescue sequences display non-confinable invasion dynamics in model-based simulation and *D. melanogaster* cage experiments[156, 27].

Here, we report an HDR-based autonomous gene-drive rescue system that eliminates non-functional EJ events in *Anopheles* mosquitoes and provide experimental evidence of sustainable long-term population modification in cages. This efficient performance is based on a phenomenon referred to as lethal/sterile mosaicism, recently reported in *D. melanogaster*[75], in which recessive nonfunctional resistant alleles act as dominant deleterious mutations and are eliminated from the mosquito population as they arise. The drive system, designated "Recoded" (Rec), was developed in the Indo-Pakistan malaria mosquito, *Anopheles stephensi*, the major vector in urban areas of India that also recently invaded the Horn of Africa[185, 182]. The drive system is inserted into the autosomal gene, *kynurenine hydroxylase* (*kh*), involved in the tryptophan metabolism pathway[93], and carries a partial recoded *kh* sequence that restores full activity of the *kh* gene (Rec*kh*). Homozygous loss-of-function mutations in *An. stephensi kh* result in a pleiotropic phenotype that includes loss of the black eye pigment (white eyes) and reduced female survival, fertility, and fecundity following blood feeding[62, 165]. The recoded *kh* sequence carried by the drive construct supports normal survival and reproductive capacity in females, while females failing to inherit the Rec*kh* construct from their mothers and carrying nonfunctional mutated copies of *kh* are culled from the population. This elimination does not rely solely on the standard negative selection of load-conferring recessive alleles but is driven by lethal/sterile mosaicism resulting from maternally deposited Cas9/gRNA complexes mutating the wild-type (WT) paternal allele. If this lethal/sterile mosaic process occurs in a sufficient number of cells, it can act dominantly to eliminate females that inherit nonfunctional EJ alleles from the reproductive pool. The combination of lethal/sterile mosaicism and negative selection results in robust drive outcomes demonstrable in population cage experiments with initial Rec*kh*-to-WT male seeding ratios of 1:1, 1:3, and 1:9. We also report a Cas9/gRNA-mediated cassette exchange[71] in mosquitoes called "Swap" that permits the rapid and flexible replacement of specific sequences within genomically integrated transgenes without the need for docking sites.

2.1.2 Results

2.1.2.1 Generation of Rec*kh* Gene-Drive Mosquitoes by Swap

The first *An. stephensi* Cas9/gRNA-based gene-drive system, AsMCRkh2 (referred hereon as "non-recoded", nRec), was inserted into the coding region of the *kh* locus[62]. Homozygous (kh^{nRec}/kh^{nRec}) or heteroallelic (kh^{nRec}/kh^{-}) loss-of-function *kh* allelic combinations

cause a white-eye phenotype and impairment of blood-fed female survival and reproduction, resulting in a suppression drive that limits its spread in caged populations[165].

We modified this loss-of-function prototype drive system by inserting a recoded portion of the *kh* complementary DNA (cDNA) precisely at the 3'-end junction of the gRNA cleavage site producing an in-frame chimeric functional *kh* gene that restores endogenous gene activity (kh^{Rec+})(Fig. 2.1a). Swap facilitated integration of the recoded cDNA through the coordinated action of two gRNAs in the presence of a donor template carrying homology arms matching the flanking regions of each of the cut sites. A mixture comprising a donor plasmid marked with green fluorescent protein (GFP) (pRec*kh*) and two plasmids, each encoding one of the two gRNAs, was injected into 504 embryos of the DsRed-marked nRec gene-drive line, which carries *vasa*-Cas9 and U6Akh2-gRNA transgenes targeting the *kh* locus in the germline[62]. Successful cassette replacement was visualized by loss of the DsRed marker and concomitant acquisition of the GFP-marked Rec*kh* cassette. Two independent transformation events were recovered from the 184 G_0 adults surviving microinjection, one in a female and one in a male founder line, totaling 96 transformants among the 25,293 G_1 larvae screened (Supplementary Table 1).

The recoded *kh* sequence carried by the Rec*kh* cassette is integrated at the junction where insertion of the original nRec line disrupted the *kh* gene and restores the coding sequence and gene function (Supplementary Fig. 1). As a result, individuals from these GFP-marked mosquito lines exhibit WT eye color while retaining the core nRec autonomous gene-drive components (Fig. 2.1b). Therefore, Rec*kh* should sustain efficient copying onto unmodified WT *kh* alleles (kh^{+}) while producing a fully functional *kh* allele (kh^{Rec+}) encoding a protein product identical to that produced by the endogenous WT locus (Fig. 2.1c).

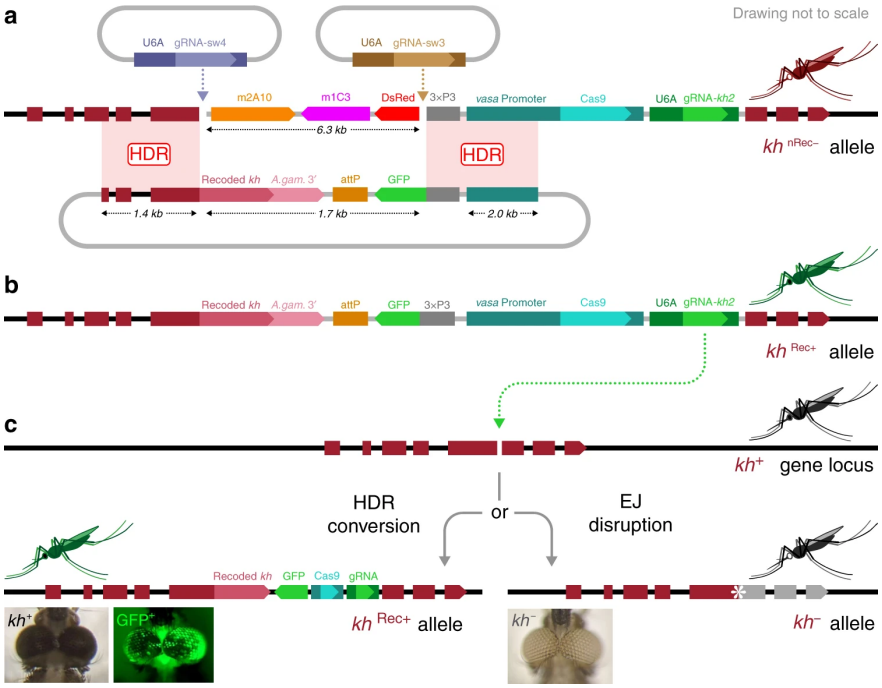

Figure 2.1: **The Rec*kh* Gene Drive**
(**A**) Swap strategy for Cas9/gRNA-mediated cassette exchange. Two plasmid-encoded gRNAs (top) guide cleavage in the genome of the white-eyed nRec mosquito line (kh^{nRec-})[62] (middle), leading to the excision of a fragment including the DsRed eye (3xP3) marker and the two antimalarial effectors m2A10 and m1C3 [100]. The HDR template plasmid (bottom) carries homology arms flanking either cut site, promoting the insertion of a GFP-marked donor template that carries a recoded portion of the *kh* gene followed by the 3'-end sequence of the *An. gambiae kh* gene including the 3'UTR (A.gam.3') to minimize homology. (**B**) The insertion of this unit restores *kh* gene function while creating a sequence (kh^{Rec+}) that is uncleavable by the endogenous drive components. (**C**) The Rec*kh* gene-drive includes an *An. stephensi* codon-optimized Cas9 driven by the germline-specific *vasa* promoter from *An. stephensi* and a gRNA (gRNA-kh2) directed to the fifth exon of the unmodified kh^{+} gene (top) regulated by the ubiquitous promoter of the *An. stephensi* U6A gene [62]. The cut in the *kh* gene of the Rec*kh* mosquito germline can be repaired by drive integration via HDR (homology-directed repair) or by the less desirable EJ (end-joining) pathway (bottom). HDR results in the integration of the drive cassette that maintains *kh* gene function at the integration site (kh^{Rec+}), while EJ usually causes the formation of loss-of-function alleles (kh^{-}). When function is lost in both copies of the gene, individuals with white eyes are produced. *kh, kynurenine hydroxylase* gene; attP, $\phi C31$ recombination site; U6A, RNA polymerase-III promoter; gRNA, guide RNAs; Cas9, Cas9 open reading frame; *vasa*, *vasa* promoter; 3xP3, eye-marker promoter; GFP, green fluorescent protein; dominant marker gene. The horizontal dimension of the mosquito heads at the eyes in the images is $\sim 1mm$.

2.1.2.2 Primary Transmission of the Rec*kh* Element is Reduced by Maternal Deposition of Cas9/gRNA Complexes

Drive performance of the Rec*kh* element was evaluated when inherited either through male or female lineages by sequentially outcrossing heterozygous Rec*kh* individuals to WT mosquitoes and assessing the percentage of individuals inheriting the drive allele (i.e., transmission) and the percentage of *kh* alleles converted to Rec*kh* by HDR copying (i.e., gene conversion or HDR rate).

As reported previously for the nRec element[62], near-complete drive transmission ($99.8\% \pm 0.15\% SEM, n = 6$) and gene conversion ($99.5\% \pm 0.29\% SEM, n = 6$) were observed in progeny whose drive-bearing parents were males (Fig. 2.2; Supplementary Table 2). Similar experiments with Rec*kh* females showed reduced transmission ($57\% \pm 2.2\% SEM, n = 6$) and gene conversion ($14\% \pm 4.4\% SEM, n = 6$)(Fig. 2.2; Supplementary Table 2), also consistent with data reported for nRec[62]. We attribute this reduced level of drive through females to result primarily from maternal deposition of Cas9/gRNA complexes in the unfertilized egg, which accumulate to impactful levels when using the *vasa* promoter to express the Cas9 transgene[80, 62, 81]. A large fraction of mosquitoes that failed to inherit the drive element through the female lineage displayed a white-eye phenotype, which most likely results from early somatic mutagenesis of the WT paternal allele in eggs that inherited a nonfunctional EJ (kh^-) allele from drive mothers. The combination of an inherited kh^- maternal allele and a newly mutated paternal allele creates a mosaic mosquito with large sectors of homozygous or heteroallelic (kh^-/kh^-) mutant cells that then give rise to the eyes and other tissues.

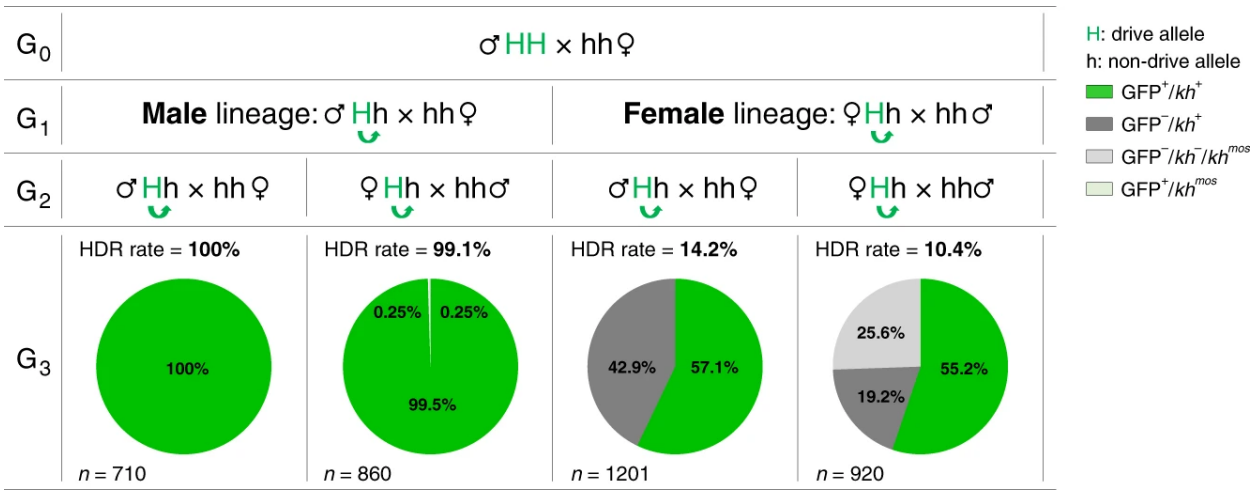

Figure 2.2: **Inheritance of Rec*kh* Through Paternal and Maternal Lineages** Charts represent the proportion of individuals inheriting the Rec*kh* drive element (GFP^+, in green) from heterozygous parents originating from drive males or drive females. The proportions of individuals that have not inherited the drive (GFP^-) element and have WT black eyes (kh^+) (dark grey) and those with white (kh^-) or mosaic (kh^{mos}) eyes (light grey) are also shown. Rare ($n = 2$) drive individuals with mosaic eyes (GFP^+/kh^{mos}) are depicted in light green. "H" and "h" refer to the mosquito genome at the *kh* locus, where "H" is the Rec*kh* drive allele and "h" is a non-drive allele. The green arrows show the potential for conversion of the h allele in the germline. The corresponding HDR rate, i.e., the proportion of h alleles converted to H alleles is reported. Each cross was performed en masse (30 females and 15 males) in triplicate cages using drive individuals mated to WT and by screening a representative subset of individuals (n) generated in the progeny. The numbers reported are pooled from the three replicate cages. Raw data for these crosses are reported in Supplementary Table 2. Data on transmission and HDR rates reported in section 2.1.2 refer to the averages of progeny from both mothers and fathers originating from a male or a female drive individual.

2.1.2.3 Rec*kh* Drives Efficiently in Caged Mosquito Populations

Because loss-of-function kh^- alleles are recessive, a single copy of the Rec*kh* element producing a WT protein sequence under native transcriptional control is predicted to rescue *kh* activity. Consistent with this expectation, fitness assessments indicate that Rec*kh* drive females are comparable to WT in their ability to reproduce (Supplementary Tables 3-4) and males carrying a drive allele have an equal ability to contribute to the subsequent generation as WT male counterparts (Supplementary Tables 3-5). In contrast, homozygous EJ events resulting in nonfunctional kh^- alleles reduce viability, fecundity, and fertility in blood-fed females relative to WT, heterozygous, or homozygous Rec*kh* females (Supplementary Tables 3-4). These potentially drive-resistant nonfunctional EJ alleles are expected to be eliminated from populations in cage experiments by a combination of two processes. The first mechanism, lethal/sterile mosaicism, relies on the maternally deposited Cas9/gRNA complexes acting somatically on the WT kh^+ paternal allele to mutate it by EJ-induced INDEL formation rendering part or all of the embryonic cells homozygous or heteroallelic for *kh* loss-of-functional alleles. The second process is standard Mendelian inheritance and negative selection whereby mating of two heterozygous kh^- carriers results in 25% of the progeny carrying homozygous or heteroallelic mutant kh^-/kh^- combinations. Females with these genotypes are unable to contribute effectively to the following generation. We hypothesize that the combination of fast-acting lethal/sterile mosaicism and slower-acting Mendelian processes efficiently culls kh^- alleles from the population resulting in the drive achieving full introduction (i.e., all individuals carrying at least one drive allele).

The performance of Rec*kh* in small laboratory populations was tested by assessing the drive dynamics in three sets of triplicate competition experiments (**A**-**C**) with differing drive seeding ratios. Each of these cage trials was initiated with specific release ratios of heterozygous Rec*kh* to WT males, 1:1 (initial drive allele frequency = 25%), 1:3 (initial drive allele frequency =12.5%), or 1:9 (initial drive allele frequency = 5%) and an equal total number of WT females (1:1 sex ratio) (Supplementary Fig. 2). The proportions of eye-phenotype combinations (GFP^+ or GFP^- fluorescence, and black, white, or mosaic color) were scored in a randomly selected sample of ~ 500 individuals for 18-20 discrete (nonoverlapping) generations (Fig. 2.3; Supplementary Tables 6-8).

Cages seeded with 1:1 Rec*kh*:WT males reached $\geq 95\%$ introduction within 5-7 generations (Fig. 2.3a). Interestingly, cages $1:3_A$ and $1:3_C$ achieved $\geq 95\%$ introduction within a comparable time-frame, generations 6 and 7, while $1:3_B$ fluctuated below 30% for 7 generations before increasing with dynamics similar to that of the other two cages and reaching $> 90\%$ introduction by generation 11 (Fig. 2.3b). The drive dynamics in $1:3_B$ were evaluated further and are reported in the next section. Finally, cages seeded with the lowest release ratio (1:9) reached $\geq 95\%$ introduction by generations 10 and 11 (Fig. 2.3c), consistent with an expected $\sim$ 3-generation delay relative to cages seeded at 1:1. Estimated population levels fluctuated in all cages, but did not show any obvious decline associated with the increasing

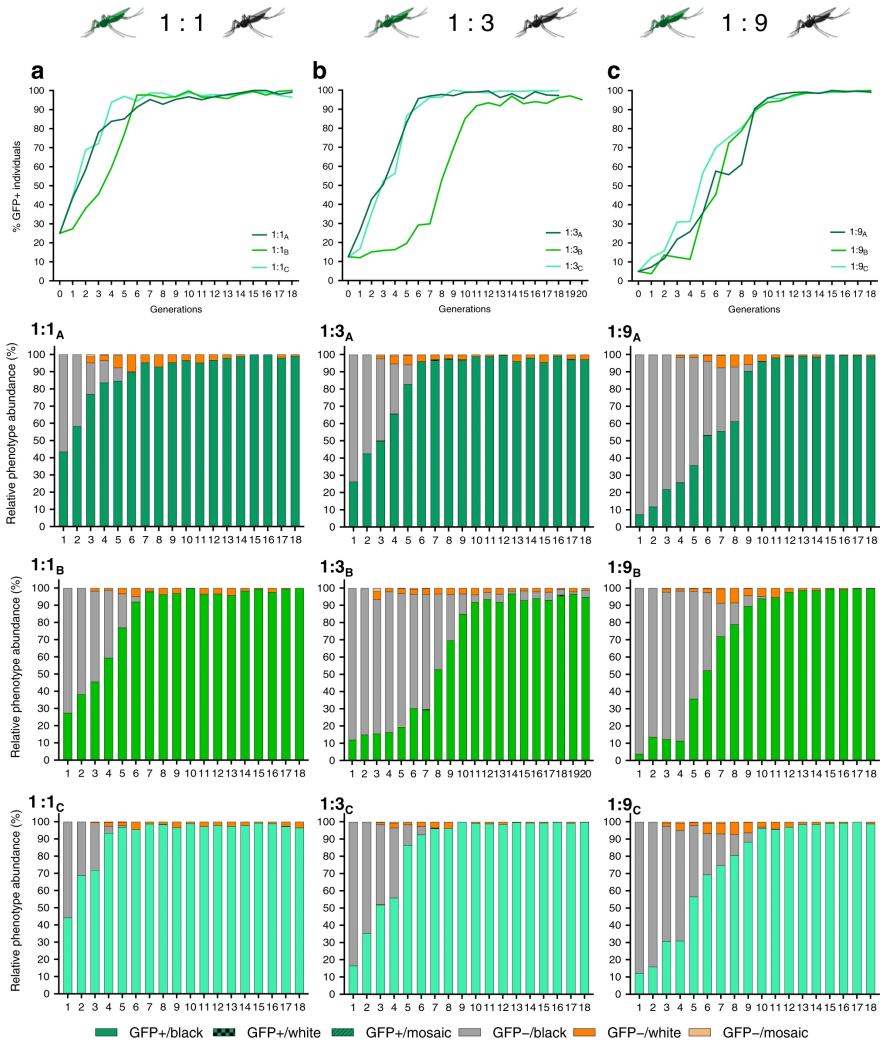

Figure 2.3: **Dynamics of Rec*kh* over 18-20 Discrete Generations in Caged Populations Seeded with Three Release Ratios of Rec*kh*:WT Males**
Top row: drive efficiency shown as percentage of GFP^+ individuals (Y-axes) at each generation (X-axes) in triplicate cages seeded with 1:1 (**A**), 1:3 (**B**), and 1:9 (**C**) Rec*kh*:WT male ratios. Bottom three rows: relative proportion of eye phenotypes (Y-axes) observed in a sample of $\sim$ 500 individuals reported for each generation (X-axes) for all cages. Individuals containing the drive are shown in green, those with WT phenotype in grey, and non-drive individuals with white or mosaic eyes in dark and light orange, respectively. A schematic of the protocol used is reported in Supplementary Fig. 2 and raw data for each cage in Supplementary Tables 6-8.

prevalence of the drive (Supplementary Fig. 3).

Potential drive-resistant individuals (GFP$^-$/white) could be scored first in generation 3 mosquitoes. Sequencing their *kh* loci confirmed that the phenotype was due to both gene copies being inactivated by out-of-frame or in-frame nonfunctional heteroallelic or homozygous mutations (Supplementary Table 9). As expected from the load observed in adult kh^-/kh^- females, the proportion of GFP$^-$/white individuals decreased progressively over subsequent generations (Fig. 2.3). However, as hypothesized, the rate of decrease exceeded that expected for the elimination of homozygous recessive individuals solely by negative selection and Mendelian inheritance. We propose that this is drive by lethal/sterile mosaicism (Fig. 2.4a-c) which accelerates the elimination of kh^- mutations inherited through females. According to this hypothesis, perduring Cas9/gRNA complexes mutagenize WT paternal kh^+ alleles somatically rendering EJ-derived loss-of-function mutations functionally dominant when transmitted by females. Because inheriting a copy of the kh^{Rec+} allele restores WT *kh* activity, including eye color and female survival and reproductive capacity, Rec*kh* females are protected from the deleterious effects of mosaicism (Fig. 2.4b). In contrast, females carrying a nonfunctional EJ-derived kh^- allele often have somatic tissues comprised of double-mutant cells and therefore fail to contribute significantly to the following generation (Fig. 2.4c). Consistent with this interpretation, the frequency of EJ-induced kh^- alleles (scored as individuals with white eyes) was highest at the steepest phase in the drive curves reflecting the generation of mosaic individuals (Fig. 2.3; Supplementary Fig. 4). Evidence of somatic mutagenesis affecting small patches of tissue in the eyes was observed by the transient appearance of non-drive individuals with mosaic eyes (GFP$^-$/mosaic) in generations 3-9 (Fig. 2.3; Supplementary Tables 6-8).

Rare (20/79,425; 0.025%) white-eyed drive (GFP$^+$/white) individuals also were recovered from six cages (Supplementary Tables 6-8). Sequencing of the *kh* locus in a subset of these individuals showed that they carried drive alleles with an out-of-frame INDEL at the insertional site junction that disrupts *kh* recoding (kh^{Rec-}) in combination with a kh^- allele (Supplementary Table 10). Such kh^{Rec-} alleles may be generated by inaccurate HDR or EJ events resulting from the drive allele being targeted at low frequency by the endogenous gRNA. While such alleles may still retain drive capacity, they also suffer from the load observed in white-eyed females, consistent with these genotypes being rare and not accumulating in any of the cages. The rarity of this class of events and their failure to thrive provide clear experimental support for robust HDR-mediated rescue of endogenous gene function and validation of the recoding strategy.

When comparing the performance of Rec*kh* in population cages to previous studies conducted on the nRec drive[165] (Supplementary Fig. 5), we found little difference in the dynamics of the initial exponential phase of the drive curves at all seeding ratios. However, drive trajectories diverged significantly at later generations during the middle and final phases of the drive process. All Rec*kh* cages reached maximum introduction and the drive system was maintained stably for the remaining observed generations. In contrast, only two of the three

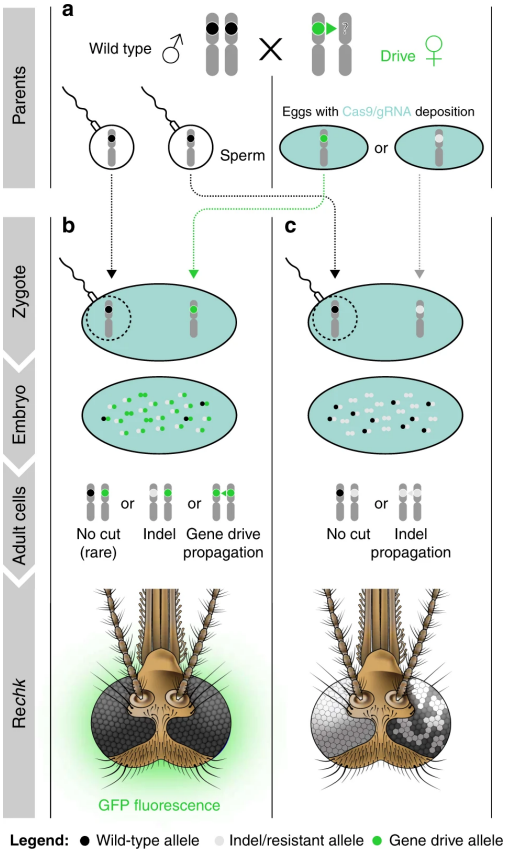

Figure 2.4: **Effects of Lethal/Sterile Mosaicism on the Rec*kh* Gene-Drive System** (**A**) A female heterozygous for the drive can produce eggs carrying a copy of the drive (green circle, kh^{Rec+}) or eggs carrying an EJ-induced nonfunctional resistant allele (white circle, kh^{-}). Both types of eggs carry maternally deposited cytoplasmic Cas9/gRNA complexes (light blue filling) that can act on the incoming WT paternal allele (black circle, kh^{+}). (**B**) The soma of individuals inheriting a copy of the drive from their mothers is a mosaic of cells with varying proportions of genotypes kh^{Rec+}/kh^{-}, kh^{Rec+}/kh^{+}, and kh^{Rec+}/kh^{Rec+}. Rec*kh* individuals emerging from such embryos have at least one functional copy of *kh* provided by the drive system (kh^{Rec+}), therefore have GFP^{+}/black eyes and females are fit for reproduction. (**C**) The soma of individuals inheriting an EJ nonfunctional mutation from their drive mothers is a mosaic of cells with genotypes kh^{-}/kh^{-} or kh^{-}/kh^{+}. The ability of females emerging from such embryos to survive and reproduce depends on the proportion of somatic cells with genotype kh^{-}/kh^{-}. These individuals may display mosaic or white-eye phenotype if mutations affect the cells forming the eyes. Diploid cells in (**B**) and (**C**) that become germline progenitors also may be affected by mosaicism, which can affect drive capabilities.

1:1 and 1:3 nRec cages, and none of the 1:9 replicates, reached > 95% introduction before either the mosquito population crashed due to fixation of the *kh* double-mutant genotypes or the drive was selected out of the population. A delay in mid-stage growth in the nRec 1:3 cages was observed, mostly likely resulting from the elimination through lethal/sterile mosaicism of female progeny receiving a kh^{nRec-} or a kh^{-} allele from drive mothers[165].

2.1.2.4 A Functional Resistant Allele Does Not Prevent Drive Introduction in Caged Populations

Cage replicate $1{:}3_B$ displayed anomalous drive dynamics where the drive stalled for the first 7 generations before increasing and reaching full introduction (Fig. 2.3). We examined the potential basis for this outlier cage by assessing whether the presence and relative abundance of specific Cas9-induced mutations might have accounted for the observed drive delay. We deep-sequenced amplicons of the genomic DNA surrounding the cut site in non-drive alleles from pooled individuals from generations G_0, G_8, and G_{14} (Supplementary Table 11).

A total of 98% of the non-drive *kh* alleles found in the G_0 were identical to the unmutated WT sequence and the remaining were mostly INDELs adjacent to the gRNA-directed cut site. These kh^{-} mutations were likely carried by the heterozygous males selected at random to seed the first cage from a founder cage that had been intercrossed for several generations. Single-nucleotide substitutions (0.03-0.2%) are consistent with the presence of rare polymorphisms around the target site in our colony cage. By the time the drive reached 50% introduction in generation G_8, 83% of the non-drive *kh* alleles were still unmodified WT kh^{+} alleles. The most prevalent mutated non-drive alleles were INDELs of various lengths (1-3% each) and two in-frame substitutions (1-2% each) causing non-synonymous amino acid changes (TACG>CGAT:Y328R-G329W and CAG>GCA:Q330A). WT *kh* alleles were rare ($< 0.1\%$) in generation G_{14} and the drive reached > 95% introduction. The Y328R-G329W substitution was absent, suggesting it disrupted protein function, while the Q330A substitution had increased in frequency to 12% of the total non-drive alleles.

Further investigation of the CAG>GCA:Q330A mutation confirmed that it maintains *kh* function, as individuals homozygous for the mutation or trans-heterozygous for the mutation and a kh^{-} allele isolated from generation G_{16} had black eyes (Supplementary Table 12). This functional mutation affects the PAM site creating a *kh* allele that is largely (or totally) resistant to further Cas9 cleavage (kh^{+R}), as demonstrated by the 3:1 Mendelian segregation of the GFP marker and the absence of white-eyed or mosaic progeny in multigenerational competition cage experiments seeded with equal proportions of kh^{Rec+}/kh^{+R} males and females (described in more detail below and in Fig. 2.5).

Resistant EJ events that preserve gene activity could be (1) under negative selection relative to the kh^{Rec+} if they affect protein function, (2) positively selected if they cause a relative increase in fitness compared to drive individuals, or (3) neutral if the fitness of the two

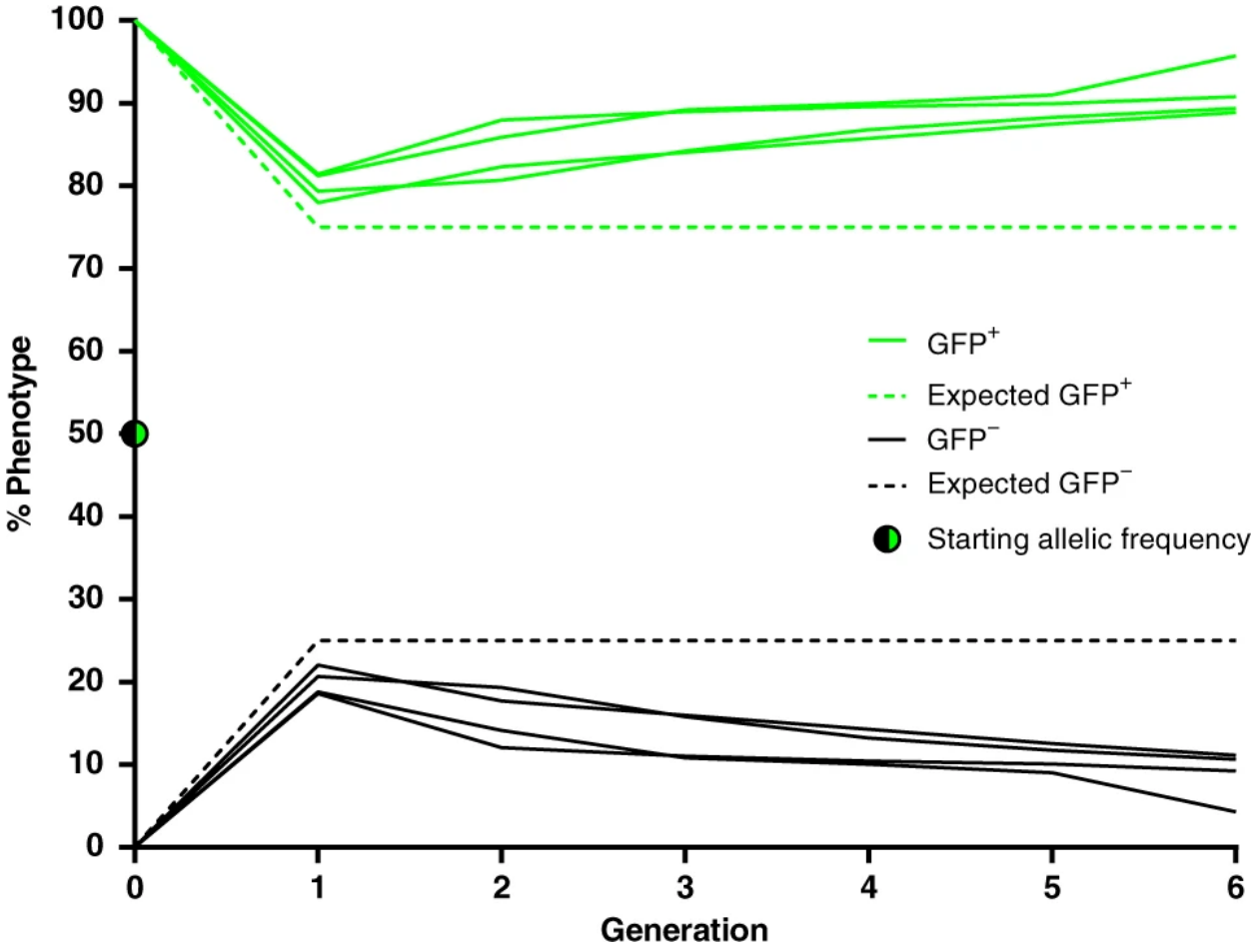

Figure 2.5: **Competition Between the Drive kh^{Rec+} and the Resistant Functional kh^{+R} Alleles in Caged Populations**
Four independent cages were set up using 100 male and 100 female kh^{Rec+}/kh^{+R} mosquitoes each, which corresponds to an initial allelic frequency of 50%, marked using a half green-half black dot at generation G_0. Eye fluorescence (GFP^+ or GFP^-) and color (black, white, or mosaic) frequencies associated with each modified kh allele were scored at every generation for six consecutive nonoverlapping generations. The proportion of GFP^+ individuals (genotype kh^{Rec+}/kh^{Rec+} or kh^{Rec+}/kh^{+R}) in the replicate cages is depicted by green lines and its expected frequency in the presence of equal competition between the two alleles (75%) as a dashed green line. The proportion of GFP^- individuals (genotype kh^{+R}/kh^{+R}) is depicted as black lines and its expected frequency in the presence of equal competition between the two alleles (25%) as a dashed black line. All individuals screened exhibited fully WT black eyes. Raw data for these crosses are reported in Supplementary Table 13.

modified *kh* alleles is comparable. Population analysis of the late stages in the drive process did not support either a positive or negative selection model for the kh^{+R} mutation, since the frequency of GFP$^-$ black-eyed individuals remained steady (1-6%) over 10 generations (G_{11}-G_{20}), consistent with a negligible fitness difference between the kh^{+R} EJ allele and the recoded kh^{Rec+} drive allele (Supplementary Table 7).

We further tested the relative fitness of the drive and the functional resistant alleles in a multigenerational cage experiment where kh^{Rec+} and kh^{+R} could compete. To do so, we monitored eye fluorescence and pigmentation phenotypes in triplicate cages seeded at equal (1:1) ratios of the two alleles where kh^{Rec+}/kh^{+R} individuals could mate. Assuming random mating in these conditions, phenotypes should stabilize at a 3:1 ratio of GFP$^+$ to GFP$^-$ individuals if the two alleles are comparably competitive. The proportion of GFP$^-$/black-eyed kh^{+R}/kh^{+R} did not exceed the 25% expected for Mendelian inheritance, but in fact decreased over time (Fig. 2.5; Supplementary Table 13), consistent with a fitness load associated with this mutation relative to the drive allele. We found no evidence in these experiments for the resistant allele being cut at low frequency by Cas9 since we did not recover any individual with mosaic or white eyes. In light of these observations, we conclude that the appearance of the kh^{+R} resistant functional allele in the $1{:}3_B$ cage is unlikely to have contributed to the pause in drive dynamics observed, and moreover that the kh^{Rec+} drive allele is at least as fit, if not more so, than the functional kh^{+R} EJ allele.

2.1.2.5 Modeling Rec*kh* Gene Drive

We performed mathematical modeling to assess whether the observed experimental drive dynamics in the cages conformed with predictions based on the copying efficiency of Rec*kh* in single-generation crosses and with genotype-specific loads in females. A model of autosomal Cas9/gRNA-based gene drive, similar to one used for nRec[165], was fitted to the observed cage data and includes two alternate mutated resistant alleles, functional (kh^{+R}) and non-functional (kh^-), maternal deposition of Cas9/gRNA complexes, and genotype-specific loads.

Model fitting was consistent with high ($> 99\%$) HDR efficiencies in males and females in the absence of maternal Cas9/gRNA deposition with 17% (95% CrI: 16-18%) of the remaining EJ alleles ($< 1\%$) being functional, and the remainder being nonfunctional. Maternal deposition was inferred to result in cleavage of embryonic WT alleles with a frequency of 93.7% (95% CrI: 99.2-100%), while the drive allele is associated with a negligible load, consistent with laboratory observations. The trajectories of GFP$^+$ individuals align well with experimental observations at all release ratios and are consistent with a highly efficient gene-drive system (Fig. 2.6, Supplementary Fig. 6).

Functional resistant alleles are generated at a low rate, and although they can persist in the population due to their resistance to cleavage (GFP$^-$/black in Supplementary Fig. 6), they were not a significant obstacle to the drive at the release ratios analyzed. Nonfunctional

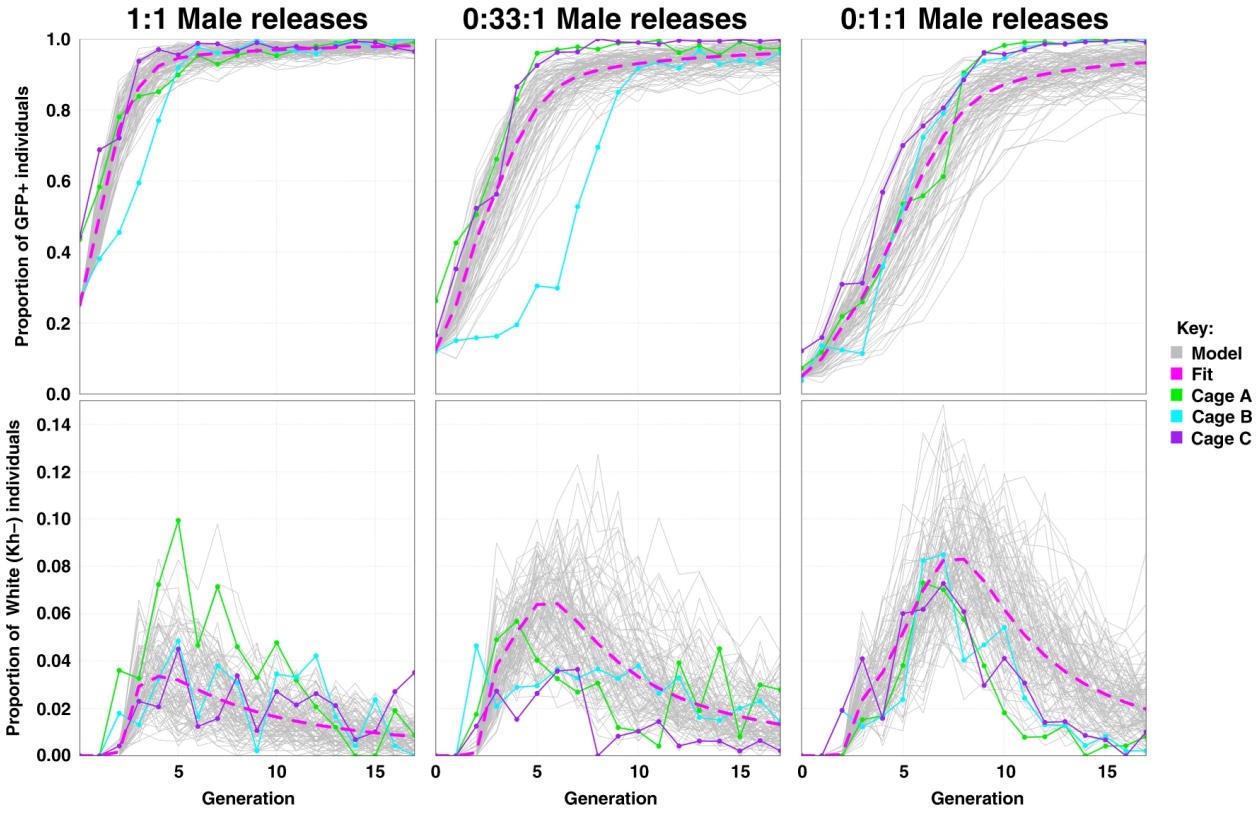

Figure 2.6: **Observed and Model-Predicted Dynamics of GFP^+ and kh^- Phenotypes in the Rec*kh* Cage Experiments**
Solid green, blue, and purple lines represent the experimental data over 18 generations observed in 3 replicates (Cages A-C) with release ratios of Rec*kh*:WT males of 1:1, 1:3, and 1:9, respectively. Dotted pink lines represent the fitted deterministic model (Fit), and grey lines are 100 stochastic realizations of the fitted model for each release ratio (Model). X-axes report the generation number after release and Y-axes the proportion of each eye phenotype. The GFP^+ phenotype results from having at least one copy of the drive allele and hence reflects the spread of the gene-drive system to full or near full introduction for all experiments. The kh^- phenotype is associated with having no copies of the WT, Rec*kh*, or functional resistant alleles (i.e., having two copies of the nonfunctional resistant allele) and reflects the low-level emergence and gradual elimination of this allele from the population due to its load in homozygous females. The stochastic model captures the variability inherent in the experimental process and reflects some of the variability observed in the early stages of the spread of the gene-drive allele.

resistant alleles also were generated at a low rate and were strongly selected against in the progeny of females that generated them, while their subsequent elimination was gradual

(GFP$^-$/white in Supplementary Fig. 6) due to their viability in males and heterozygotes of both sexes.

Finally, a stochastic model captured the potential role of chance events such as mate choice (multinomial distributed), egg production (Poisson distributed), progeny genotype (multinomial distributed), and finite sampling of the next generation (multivariate hypergeometric distributed) (Fig. 2.6). Stochastic model trajectories reflect some of the variability observed in the early stages of spread of the gene-drive allele, with additional transient delays observed in cages 1:1$_B$ and 1:3$_B$.

2.1.3 Discussion

Population modification strategies employing gene-drive systems to spread anti-parasite effector molecules through populations of *Anopheles* mosquitoes are gaining momentum in the fight against malaria[23, 24, 136]. However, the creation of mutated target sequences resistant to the drive, especially those preserving gene function, can limit transgene introduction, particularly if a load is associated with the presence of the drive system[165, 82].

Here, we provide experimental evidence for a readily generalizable gene-drive rescue system for efficient population modification in *Anopheles* mosquitoes that actively eliminates nonfunctional resistant alleles as they arise and rapidly attains $> 95\%$ introduction in caged populations, despite the creation of rare functional EJ variants. As a result, the Rec*kh* system converts a population suppression gene-drive (nRec) to an efficient population modification system in *An. stephensi*. Furthermore, while proof-of-concept for CRISPR/Cas9-based rescue systems that employ recoded sequences has been produced in *D. melanogaster*[159, 29, 27], this work reports the application of such a system to mosquitoes. Assessments of the long-term dynamics of this system in caged mosquito populations show that Rec*kh* spreads quickly (5-11 generations depending on release ratio) and efficiently even when a small percentage of males are released. After reaching maximum introduction, the Rec*kh* drive persists without an evident impact on the population size.

Rec*kh* targets the haplosufficient gene *kh*, required for adult female survival and reproduction in *An. stephensi* following a blood meal[165], and provides a recoded portion of the gene that rescues its function. In doing so, individuals that carry a copy of the drive are functionally protected and comparably as fit as their WT counterparts, while nonfunctional resistant alleles are eliminated owing to the reduced survival and impaired reproductive capacity in white-eyed homozygous and mosaic heterozygous females. We propose that such elimination is driven initially by the active process of lethal/sterile mosaicism[75], during which recessive nonfunctional kh^- mutations function as dominant alleles during the drive process, therefore acting as an autocatalytic mechanism that eliminates female progeny of drive mothers from the breeding pool. Mosaicism-mediated elimination is complemented by negative selection of the residual kh^- recessive alleles transmitted through male progeny, which acts over many

generations to eliminate these costly alleles from the population. Overall, the accelerated elimination of kh^- alleles results in an increased apparent HDR-mediated conversion frequency.

Modeling predictions based on transmission frequencies and loads of different allele combinations are consistent with the highly efficient gene-drive outcomes observed in the cage experiments. Simulations are consistent with nonfunctional kh^- resistant alleles, which are generated primarily in females, being eliminated rapidly by lethal/sterile mosaicism and then gradually as standard recessive alleles due to the load associated with bi-allelic *kh* loss-of-function. Consistent with lethal/mosaicism taking place in offspring of gene-drive females, the presence of white-eyed individuals carrying *kh* homozygous mutations in the cage experiments followed a general trend in which they were most abundant at the steepest phase in the drive curve, while the remaining kh^- alleles propagating through males were slowly eliminated through negative selection over time. Potential loads in white-eyed males, such as impaired vision, also could have contributed to the elimination of nonfunctional resistant alleles. Future recoded drive systems inserted into genes essential for viability of both sexes or required for both male and female fertility could drive even more rapidly.

The lag in drive invasion observed in one of the 1:3 release ratio cages, which deviates from the consistent trends observed in the other eight cages, remains difficult to explain since it is not fully supported by the predictions of the stochastic model. It is not likely to derive from a sampling effect, as a delay in the drive increase was not observed in any other cage, including those seeded with the lowest release ratio where such effect, if present, would be more prominent. Also, our analysis does not support involvement of resistant functional alleles in this observation. Further analysis of these trajectories will be required to ascertain the underlying mechanisms responsible for such outlier events. Nevertheless, the transient delay in drive invasion observed in this cage did not prevent eventual maximum introduction of the drive.

A major concern associated with the persistence of gene-drives is the generation of cleavage-resistant sequences that preserve gene function. In population suppression systems, these alleles are positively selected over the drive allele if the latter causes loss-of-function of the targeted gene resulting in a fitness disadvantage[165, 82]. Therefore, choosing a functionally constrained target site should alleviate the issue of resistance and this proved successful in suppression and sex-distorting strategies[114, 183]. This approach also can mitigate the effect of resistance in population modification systems, where the effect of functional resistant target sequences on the drive dynamics is dependent on their fitness relative to that of the drive. Data from the cage experiments and modeling show that functional resistant alleles in the Rec*kh* system are generated at a low rate and do not prevent the drive from reaching full introduction. We attribute this outcome to the high functional constraint of the chosen target site within the *kh* gene that is likely to cause total or partial protein malfunction when mutated. This hypothesis is supported by evidence from the crystal structure

of the KH enzyme from *Saccharomyces cerevisiae* where the P326-F327-Y328-G329-Q330 loop, which encompasses the gRNA site in our experiments, plays an important role in accommodating the rearrangements of the active site of the enzyme during binding[8]. Indeed, this region shows high primary amino acid sequence conservation from insects to humans[93, 83]. Furthermore, we found examples of in-frame mutations in individuals with white eyes, consistent with previous findings from similar cage experiments conducted using the same gRNA where the strong contribution of the Y328 and G329 residues to protein function was highlighted[165]. Here we report evidence for the contribution of the Q330 residue, whose codon forms part of the PAM site. While individuals homozygous for the Q330A substitution display a black WT eye phenotype and the mutation is resistant to Cas9 cleavage, we observed only a modest accumulation of this mutation amongst the non-drive alleles that did not prevent drive invasion. We explored this further in cage experiments were the mutated resistant allele was in direct competition with the drive and found that the mutation was outcompeted by the drive. We therefore conclude that the Q330A mutation likely only partially restores WT levels of KH enzyme and thus does not pose a major obstacle to the drive process. Given the documented functional constraints of multiple amino acid residues at the *kh* target site and evidence that the kh^{Rec+} allele does not carry an obvious fitness burden, it is likely that this drive will outcompete a number of Cas9-induced functional resistant mutations. Similarly, while we currently do not have access to genome data on field populations of *An. stephensi*, we would not expect the targeted conserved coding sequence to show high levels of natural polymorphism. Nevertheless, analysis of the *kh* locus in field-caught *An. stephensi* natural populations is needed to assess this question.

Field performance of population modification drives is not only dependent on the drive copying process as such but also on effects on mosquito fitness produced by sequences encoded in the cargo (e.g., the antimalarial effectors). The Rec*kh* construct currently lacks antimalarial molecules as the two effectors m2A10 and m1C3 present in the original prototype were excised due to molecular constraints to allow for partial *kh* recoding and simultaneous fluorescent marker exchange. However, we do not anticipate major fitness impacts in our drive mosquitoes as our population modification strategy is based on the use of synthetic single-chain antibodies (scFv) specifically directed against *Plasmodium* parasite antigens. Due to their target specificity and regulated blood meal-induced expression, combinations of these molecules have been shown to have little (if any) impact on mosquito fitness[100]. In contrast, expression of multiple toxins and synthetic molecules with broader activity can exert undesired impacts on crucial physiological processes or the gut microbiota[49]. Nonetheless, comprehensive fitness assessments in insectary cages and contained field trials are needed to validate these hypotheses. Furthermore, we predict that the addition of genes coding for antimalarial effectors and their regulatory regions will not greatly affect drive copying since the initial exponential growth of Rec*kh* and nRec systems were similar, supporting the conclusion that Rec*kh* will tolerate an increase in cargo size (i.e., the nRec construct is $\sim$ 4.6 kb larger than Rec*kh*).

We also report the application to mosquitoes of the highly efficient Swap transgenesis technology, a Cas9/gRNA-driven cassette exchange system[71], to update sequences of previously integrated gene-drive systems. Swap is a flexible tool that efficiently edits existing gene-drive mosquito lines using small constructs. Swap also can enable broader genome engineering efforts in mosquitoes. Since it does not require the presence of recombination sites such as those needed for $\phi C31$ recombinase-mediated cassette exchange, updating sequences could be inserted anywhere suitable gRNA sites are available. For example, the system could be used to exploit endogenous cis-acting elements[153, 94] through the seamless integration of desired coding regions. For population modification purposes, a drive line can be envisioned that carries strategically placed gRNA targets to replace antiparasitic molecule combinations and test their blocking efficacy. Indeed, our next steps include the addition of combinations of antimalarial scFvs that block parasite development using Swap. While the Rec*kh* line can be used as a classic docking line via the $\phi C31$ *attP* site introduced along with the recoded *kh* sequence, integration via this site would significantly increase the size of the sequence between the homology arms as well as bring in additional sequences that might impact drive performance. Nevertheless, this efficient integration strategy can be used to assess blocking capabilities of alternative combinations of antimalarial effectors expressed at the *kh* genomic locus. Finally, while the Rec*kh* system described here is already highly efficient, substituting the current *vasa* promoter with more tightly regulated control sequences, such as those of the *nanos*[24] or *zpg*[81] genes employed for gene-drive designs in *An. gambiae*, could improve its performance by reducing the fraction of EJ alleles persisting during the drive process. However, we believe that lethal-mosaicism provides a general solution to the problem of females generating EJ alleles and should allow for the design of efficient gene drives at many different target loci and insect species where identification of "ideal" promoters may not be straightforward or transferable.

Overall, the laboratory assessments conducted so far, carried out in line with the recommended phased pathway for testing gene-drive mosquitoes[15, 101, 102], show that the characteristics of the Rec*kh* gene-drive are likely to conform with those defined as part of a proposed Target Product Profile for population modification of mosquito strains[23] once the addition of the anti-malaria effectors is proven effective under this configuration. The highly efficient performance of the Rec*kh* drive system makes it an excellent candidate for genetic control of an important malaria mosquito, *An. stephensi*, and this technology should be readily adaptable to other mosquito species as well as other insect disease vectors.

2.1.4 Methods

2.1.4.1 Mosquitoes

An. stephensi Indian strain (gift from M. Jacobs-Lorena, Johns Hopkins University) were maintained in insectary conditions (27°C and 77% humidity) with a photoperiod of 12-h light/dark and 30min of dawn/dusk. Sucrose solutions (10% wt/vol) were provided ad

libitum and blood meals consisting of defibrinated calf blood (Colorado Serum Co., Denver) were offered to 3-7-day-old adults through the Hemotek® membrane feeding system. Larval stages were reared in distilled water and fed TetraMin® fish food mixed with yeast powder. Gene-drive mosquitoes were contained in ACL-2 insectary facilities at the University of California, Irvine and handled according to recommended safety procedures[3, 1, 14].

2.1.4.2 Plasmids for the Swap Strategy

The Swap strategy employed to convert nRec to Rec*kh* uses the three plasmids shown in Fig. 2.1a: (1) pVG362_Aste-U6A-Swap3-gRNA to express the gRNA-sw3, (2) pVG363_Aste-U6A-Swap4-gRNA to express the gRNA-sw4, and (3) pVG344_Aste_kh2-MCRv3-vasa-Cas9 to provide the HDR template containing the recoded-*kh* coding fragment and the GFP marker.

To generate plasmids pVG362 and pVG363, a pair of oligonucleotides were synthesized (Integrated DNA Technologies) for each plasmid with 19 (pVG362) or 20 (pVG363) bases of the target sequence chosen for the strategy. These were annealed and ligated with T4 ligase (New England Biolabs) into the pVG145-Aste-U6A-Bbs1 plasmid[62] linearized with *Bbs*I. The cloning strategy was adapted from the work of Port et al.[166]. The oligonucleotides used to construct pVG362 were 1288_Aste-Swap3-Target_F (CTTGTTCTTGGAGGAGCGCACCA) and 1289_Aste-Swap3-Target_R (AAACTGGTGCGCTCCTCCAAGAA). The oligonucleotides used to construct pVG363 were 1290_Aste-Swap4-Target_F (CTTGTTACGttaattaaCGTAGAA) and 1291_Aste-Swap4-Target_R (AAACTTCTACGttaattaaCGTAA).

The pVG344 plasmid was cloned using the NEBuilder HIFI DNA Assembly Cloning Kit (New England Biolabs) to assemble four amplified fragments. Fragment 1 was generated by amplification of the backbone region of plasmid pVG163_pAsMCRkh2[62], fragment 2 was generated by amplification of the *kh* recoded rescue fragment from a plasmid synthesized by GenScript Inc., fragment 3 was amplified from a plasmid containing a 3× P3-GFP cassette commonly used for insect transgenesis, and fragment 4 also was amplified using pVG163_pAsMCRkh2 as a template. Primer pairs used to amplify each fragment were: Fragment 1: 494_pUC19_Backbone_F (GGTATCAGCTCACTCAAAGGCGGTAATACGG) and 1227_As-MCR2_GA_backbone_R (CGTAGAACGGAACCATCGCGTG),
Fragment 2: 1231_As-MCR2_GA_RecodedFrag_F (CGCGATGGTTCCGTTCTACGG) and 1232_As-MCR2_GA_RecodedFrag_R (CTACGCCCC,,CAACTGAGAGAACTC), Fragment 3: 1230_As-MCR2_GA_GFP_F (TCTCTCAGTTGGGGGCGTAGCGTACGCGTATCGATAAGCTTTAAGATAC) and 1229_As-MCR2_GA_GFP_R (CACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCAC), Fragment 4: 1228_As-MCR_GA_backbone_F (CACCATGGTGGCGACCGGTGGATC) and 1241_As_MCR2_GA_HA2_R (CGCCTTTGAGTGAGCTGATACCGTGAGCAAAAGGAGACGG).

2.1.4.3 Microinjections and Establishment of Rec*kh*

Embryos were obtained from heterozygous females of the *An. stephensi* AsMCRkh2 (nRec) gene-drive line[62]. Microinjection procedures[10] were performed using a plasmid mix containing 600 ng/μL of pRec-*kh* donor (pVG344) and 200 ng/μL of each gRNA-sw3 (pVG362) and gRNA-sw4 (pVG363) plasmids. Surviving G_0 adults were sorted in pools of 2-4 males and 7-10 females and outcrossed to 10x WT females and 1× WT males, respectively. G_1 progeny were screened as larvae for the inheritance of the GFP eye marker and kept as separate lines according to their male (male-4) or female (female-4) founder lineage. The two lines were screened routinely as larvae for the inheritance of the GFP eye marker and as pupae for the eye-color phenotype (black [WT], white, or mosaic) and maintained by intercrossing GFP^+ black-eyed individuals. A homozygous drive line was established from the female-4 intercrossed line.

Molecular confirmation of HDR-mediated target site integration of the Rec*kh* cargo was performed on genomic DNA extracted from single GFP^+ black-eyed individuals using the Wizard® genomic DNA purification kit (Promega). Primers Kh1-ext-fw (CACTGTTGGCACTCCATCTG) and Rec-kh-rv2 (GGGCTTCAACAACTGAAAAG) were used to amplify a 2190 bp region spanning the cut site of gRNA-sw4, while primers eGFP-fw (AAGTCGTGCTGCTTCATGTG) and Vasa-rv (GTAAAAGCCGCATTTTCCAA) were used to amplify a 2303 bp region across the cut site of gRNA-sw3. Gene amplification reactions were performed using Phusion® High-Fidelity PCR Master Mix (New England Biolabs). Sanger sequencing (Genewiz, San Diego) with primers Rec-kh-rv2 and eGFP-fw was used to confirm the sequence of the integration sites.

2.1.4.4 Primary Drive Transmission Assessments

Drive transmission and HDR conversion rates through the male and female lineages were assessed in sequential en masse outcrosses of Rec*kh* individuals to WT. Each cross comprised 30 females and 15 males and was performed in three replicate cages. A representative subset of the progeny of each cross was scored for the presence of the GFP fluorescent marker and the eye color phenotype (black, white, or mosaic) in adults. A schematic of the crossing performed is reported in Fig. 2.2.

Drive transmission is defined as the percentage of individuals inheriting the Rec*kh* element. Gene conversion or HDR rate is defined as the percentage of *kh* alleles converted to Rec*kh* by HDR copying and is calculated using the formula $\frac{2 \cdot X - 0.5 \cdot n}{n}$ (X is the number of GFP^+ individuals and n the total number of mosquito counted)[62].

2.1.4.5 Female Fecundity and Fertility

Homozygous Rec*kh* (kh^{Rec+}/kh^{Rec+}), heterozygous Rec*kh* carrying a copy of the drive and a kh^- allele (kh^{Rec+}/kh^-), WT (kh^+/kh^+), and white-eye (kh^-/kh^-)[165] females were

included in this analysis. Adult females 5-7 day old were offered a blood meal for 45 min over 2 consecutive days and unfed females removed. After 3 days, single females were set up to oviposit in 16 oz ($\sim 454cm^3$) paper cups containing a plastic oviposition cup lined with damp filter paper. Eggs were counted the next day using a stereomicroscope and transferred to water cups lined with filter paper for hatching. Larvae emerging from single egg batches were counted at the first or second instar (L_1-L_2). Fecundity refers to the number of eggs laid by a single female and fertility to the proportion of larvae hatching from these individual egg batches. A one-way ANOVA with Tukey's multiple comparison post hoc test was used to assess significant differences ($p > 0.05$) in the performance of females from the four groups tested.

2.1.4.6 Male Contribution to the Following Generation

Triplicate cages were seeded with 75 Rec*kh* homozygous males, 75 WT males, and 150 WT females. All individuals were added to the cage as 3-7-day-old adults and females were offered a blood meal over two consecutive days. Approximately, 2000-2500 L_4 larvae were selected randomly from the progeny of each replicate cage and scored for the presence of GFP. A two-tail binomial test was used to compare the observed and expected distributions and test for significant ($p > 0.05$) deviation from equal frequency (50%) of GFP^+ and GFP^- individuals.

2.1.4.7 Cage Trial Set Up and Maintenance

A schematic representation of the cage trial protocol implemented is shown in Supplementary Fig. 2. The trial consisted of 18 nonoverlapping generations and was conducted in 5000 cm^3 cages essentially as described by Pham et al.[165]. Triplicate cages (A–C) were seeded with three single release ratios, 1:1, 1:3, 1:9, of 3-5 day old age-matched Rec*kh* heterozygous to WT (Rec*kh* :WT) male adults (100 in total), and 100 WT adult females were added to reach an equal sex ratio for a total of 200 individuals per cage. The number of Rec*kh* males was 50 in the 1:1 release cages, 25 in the 1:3 cages, and 10 in the 1:9 cages. Adults 5-7 days old were offered a blood meal over two consecutive days. Three days later, dead adults were removed from each cage and an egg cup was provided for two days. Of the hatching larvae: 200 L_1-L_2 were selected at random and reared to adulthood to establish the following generation, 500 L_1-L_2 were selected randomly to assess the progression of the drive by screening the eye phenotype in L_4 larvae and pupae (see section 2.1.4.8), and the remainder reared to L_4 and stored in ethanol for population counts and molecular analysis. Due to the small initial number of transgenics, the only exception to the random selection of the 200 L_1s to seed the following generation was that individuals from generation 1 of the 1:9 cages were all screened for their eye phenotype and new cages were seeded with the same proportion of drive individuals found. Finally, the screening of cage $1{:}3_B$ was carried out for additional two generations for a total of 20 generations.

2.1.4.8 Screening of the Eye Phenotype

A sample of 500 randomly selected L_4 larvae were scored at each generation for the presence of GFP fluorescence (GFP^+ and GFP^-) and separated in two corresponding larval trays; pupae emerging from each tray were screened for eye color (black, white, or mosaic). The phenotypes were reported as follows: GFP^+/kh^+ (drive individuals with black eyes); GFP^+/kh^- (drive individuals with white eyes); GFP^+/kh^{mos} (drive individuals with mosaic eyes); GFP^-/kh^+ (non-drive individuals with black eyes); GFP^-/kh^- (non-drive individuals with white eyes); GFP^-/kh^{mos} (non-drive individuals with mosaic eyes). Among these, individuals with phenotypes GFP^+/kh^- and GFP^-/kh^- were stored as adults at -20°C for sequencing.

2.1.4.9 Population Count

L_4 larvae collected throughout the experiments were stored in 50 mL conical centrifuge tubes filled with ethanol. Before counting, ethanol was rinsed off and larvae were re-suspended in a fixed volume of deionized water and placed onto a shaker moving at a constant speed. Larvae were collected using a fixed-volume scoop and counted before returning them to the shaker. A total of 6-9 measurements were taken per cage every two generations. The estimated population size was calculated by averaging the number of larvae from replicate measurements and multiplying by the conversion factor (volume of water/scoop volume). The only exception to this method of counting was generation 1 of the 1:9 cages where the whole L_4 population was counted.

2.1.4.10 Sanger Sequencing on Single Mosquitoes

Genomic DNA was extracted from whole single adult mosquitoes using either the Wizard Genomic DNA Purification Kit (Promega) or the DNeasy Blood & Tissue Kit (Qiagen). All gene amplification reactions were performed using the Phusion High Fidelity PCR Master Mix with HF Buffer (New England Biolabs). To analyze the non-drive allele, primers KhE5-4 (GACGGTGACACTGTTCATGC) and KhE5-3 (CAGATGGCATGTGCATCCTC) were used to generate a 372 bp amplicon spanning the gRNA-directed cut site in the *kh* gene. Sanger sequencing (Genewiz, San Diego) of the non-drive amplicons was performed using primer KhE5-4. To analyze the Rec*kh* drive allele, primers KhE4 (CGTTCGAGTAGCACGTTG) and Agam3 rv (CAGGTGTAGAAGAAAACACGTTG) were used to produce a 1287 bp amplicon. Sanger sequencing (Genewiz, San Diego) of the Rec*kh* amplicon was performed using primer KhE4. Sequencing results from mixed traces were resolved using CRISP-ID (http://crispid.gbiomed.kuleuven.be)[44].

2.1.4.11 DNA Extraction and Amplification from Pooled Mosquitoes

Genomic DNA from individuals used to seed the $1{:}3_B$ cage (generation 0) was extracted from pools of 20 adults (total of ~ 140); while DNA from individuals from the same cage at gen-

erations 8 and 14 was extracted from pools of 50 larvae (total of 300 each). Extractions were performed using the DNeasy Blood & Tissue Kit (Qiagen) according to manufacturer's protocol with an overnight initial lysis step. An equal volume of genomic DNA was pooled from each replicate extraction and used as template for amplification. Gene amplification was performed using the Phusion High Fidelity PCR Master Mix with HF Buffer (New England Biolabs) and primers KhE5-4 (GACGGTGACACTGTTCATGC) and KhE5-3 (CAGATGGCATGTGCATCCTC). Generated amplicons were purified from 1% agarose gels using the Zymoclean Gel DNA Recovery Kit (Zymo Research) before library preparation.

2.1.4.12 Library Preparation and Sequencing

Illumina libraries were prepared for each of three samples (G_0, G_8, and G_{16} from cage $1{:}3_B$) using the NEXTFLEX PCR-free library preparation kit and NEXTFLEX Unique Dual Index Barcodes (BIOO Scientific) following the manufacturer's instructions. The input amount of DNA was 500 ng. The ends of the DNA were repaired and adenylated. The reaction was cleaned using AMPure XP magnetic beads and Illumina barcoded adapters were ligated onto the blunt-end adenylated product. The adapter-ligated product was cleaned using AMPure XP beads. DNA quantity was measured by Qubit DNA HS assay and the fragment size assessed by Agilent Bioanalyzer 2100 DNA HS chip assay at the genomics facility of the University of Utah (GNomEx) where the libraries were sequenced on the Illumina NovaSeq with the SP flowcell 2x250 paired end.

2.1.4.13 Sequencing Data Analysis

The raw paired-end Illumina reads from the amplified genomic region were cleaned for low quality and trimmed for the presence of adapters using Trimmomatic v0.35[17]. High-quality reads were mapped against the amplicon sequence using BWA-MEM v0.7.8[122] and the alignments sorted using SAMtools v1.9[121]. Mapped paired-end reads were extracted using Picard Tools v1.96 (http://broadinstitute.github.io/picard/), and then joined to reconstruct the complete amplicon sequence using PEAR v0.9.8[210]. Identical amplicon sequences were clustered using module fastx_collapser in FASTX-ToolKit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/). Clustered sequences were aligned to the reference amplicon with MAFFT v7[108] under FFT-NS-2 tree-based progressive method with 1PAM/K = 2 scoring matrix, if they were represented by $\geq$ 3 paired-end reads in each dataset. After alignment with the amplicon, the analysis was focused on a 34 bp target sequence including the 23 bp of the gRNA to identify single-nucleotide polymorphisms and/or INDELs in this region. The final quantification of mutations at the target site was measured as relative frequency of paired reads in sequence variants represented in at least 100 reads.

2.1.4.14 Functional Resistant Allele Assessments

Individuals carrying a mutated functional *kh* allele (kh^{+R}) due to the presence of a CAG>GCA-Q330A substitution affecting the PAM site were isolated from non-drive black-eyed (GFP^-/black)

individuals from cage 1:3$_B$ at generation G_{16}. Resistance of the kh^{+R} allele to Cas9-induced cleavage was assessed in the progeny of the cross between males heterozygous for a copy of the Rec*kh* drive allele and the kh^{+R} allele (kh^{Rec+}/kh^{+R}) to WT females by scoring the frequencies of GFP^+ and GFP^- mosquitoes (expected to be $\sim 50\%$ in case of resistance to cleavage).

Allele competition experiments were conducted in four replicate cages (A–D) each seeded with 200 individuals heterozygous for a copy of the Rec*kh* drive allele and a copy of the *kh* functional resistant allele (kh^{Rec+}/kh^{+R}) with a 1:1 sex ratio. Allele competition was inferred from the eye phenotype of the progeny of these crosses by scoring for the presence of the GFP fluorescent marker (GFP^+ or GFP^-) and the eye color (black, white, or mosaics) in adults. This was carried out for six discrete (nonoverlapping) generations by screening a representative sample of $\sim$ 300 adults at each generation and seeding new cages with 200 randomly picked individuals, as described for the gene-drive cage trials. In this set-up, assuming random mating, equally competitive modified *kh* alleles are expected to maintain a phenotypic ratio of 3:1 GFP^+:GFP^- individuals in each generation; deviations from this ratio would signify unequal competitiveness.

2.1.4.15 Modeling of Cage Population Dynamics

Empirical data from the nonoverlapping gene-drive experiments were used to parameterize a model of Cas9/gRNA-based homing gene-drive including resistant allele formation, and a stochastic implementation of the fitted model was used to qualitatively compare the time series of observed genotype frequencies to model-predicted ones. Model fitting was carried out for all nine gene-drive cage experiments using Markov chain Monte Carlo methods in which estimated parameters related to loads, resistant allele generation, and the consequences of maternal deposition of Cas9/gRNA complexes were used.

We considered discrete generations, random mixing, and Mendelian inheritance rules at the gene-drive locus, with the exception that for adults heterozygous for the homing allele (denoted by "H") and WT allele (denoted by "W"), a proportion, c, of the W alleles are cleaved, while a proportion, $1-c$, remain as W alleles. Of those that are cleaved, a proportion, p_{HDR}, are subject to accurate HDR and become H alleles, while a proportion, $(1-p_{HDR})$, become resistant alleles. Of those that become resistant alleles, a proportion, p_{RES}, become in-frame, functional, cost-free resistant alleles (denoted by "R"), while the remainder, $(1-p_{RES})$, become out-of-frame, nonfunctional, or otherwise costly resistant alleles (denoted by "B"). The value of p_{HDR} is allowed to vary depending on whether the HW individual is female or male, and values for female- and male-specific HDR parameters were estimated based on G_0 crosses that provided direct information on them.

The effects of maternal deposition of Cas9/gRNA complexes were accommodated after computing the gene-drive-modified Mendelian inheritance rules. If offspring having a W allele had a mother having the H allele, then this would lead to Cas9/gRNA complexes being

deposited in the embryo by the mother, possibly resulting in cleavage of the W allele. We considered cleavage to occur in a proportion, p_{MC}, of these embryos, with a proportion, p_{MR}, of the cleaved W alleles becoming R alleles, and the remainder, $(1-p_{MR})$, becoming B alleles.

These considerations allow us to calculate expected genotype frequencies in the next generation, and to explore the impacts of loads and maternal deposition parameters that maximize the likelihood of the experimental data. Estimated parameters include loads in females associated with having one or two copies of the H allele or the BB genotype, and p_{RES}, p_{MC}, and p_{MR}, as defined earlier. A stochastic version of the fitted model was implemented using a discrete generation version of the Mosquito Gene-drive Explorer (MGDrivE) model[177] with an adult population size of 200. The complete modeling framework is described in the S1 Text of Pham et al.[165].

2.1.4.16 Statistical Analysis

Statistical tests were performed as detailed in the relevant method sections using GraphPad Prism 8.4.2.

2.2 Inherently Confinable Split-Drive Systems in *Drosophila*[2]

2.2.1 Introduction

CRISPR-based gene-drive systems offer new potential strategies for controlling pest populations[145] or disease-transmitting vectors[2, 80, 131, 123]. So-called efficient "low-threshold" CRISPR gene-drive systems (capable of spreading from low-seeding frequencies) rely on a cut-and-repair mechanism, wherein Cas9 produces a double-stranded break (DSB) in the DNA followed by homology-directed repair (HDR) using the homologous chromosome as a template and copying of the genetic element into the break on the Cas9-cleaved chromosome[61, 22]. However, the competing non-homologous end-joining (NHEJ) pathway can generate alleles resistant to Cas9 cleavage[31, 82] that interrupt spread of the desired genetic element[165], particularly if the cleavage-resistant allele carries less of a fitness cost than the driving allele. NHEJ-induced alleles consisting primarily of short insertions or deletions (indels) result in amino acid substitutions or frameshifts that modify the gRNA target sequence, preventing further Cas9 cleavage[34]. Non-functional drive-resistant loss-of-function (LOF) mutations, typically the most abundant NHEJ products, can delay complete drive introduction since they are eliminated only gradually by negative selection when homozygous[165]. Another important consideration is to generate as few, if any, *in frame* functional indels, since such cleavage resistant alleles could compete with the drive element. This latter concern is particularly relevant to suppression drive systems that they would rapidly outcompete, but also could impede the spread of modification systems if such functional alleles were numerous and/or were more fit than the drive allele.

Several strategies have been proposed to reduce the incidence or effects of cleavage-resistant alleles. One approach is to employ tightly-regulated germline-specific promoters to avoid somatic expression of Cas9 that can lead to NHEJ-induced repair events[82, 165, 33]. Some germline promoters are quite specific, at least in certain genomic contexts (e.g., *nanos*[24] or *zpg*[114]), while others are leaky (e.g., *vasa*[62, 106]). Another tactic is to identify highly-conserved genomic targets as cleavage sites such that all NHEJ mutants suffer severe fitness costs[114, 105, 183, 27].

Recently, it has been suggested that drives targeting conserved genes essential for survival or fertility that also carry a recoded cDNA restoring endogenous gene activity would benefit from two forms of positive selection in populations[2, 105, 27, 54, 29]. The first advantage results from gradual Mendelian culling of recessive deleterious NHEJ alleles. The second, a more rapid process, depends on an actively dominant phenomenon referred to as lethal/sterile mosaicism, in which maternal deposition of Cas9/gRNA complexes in the embryo mutates the paternal allele in a mosaic fashion[2, 75]. If progeny inherit the recoded drive, they

[2]This work has been previously published[192]

are protected from this maternal effect since they carry one unassailable functional allele. Individuals inheriting a non-functional NHEJ allele, however, will either die or be infertile if the targeted gene is required broadly in many cells of the organism.

Two main types of CRISPR-based gene drives, or "active genetic" elements, have been designed and tested. The first, full gene drives (fGD)[2, 80, 61, 62], have linked sources of Cas9 and gRNA that are inserted together as one unit at a single genomic site. The second, split gene drives (sGD)[123, 106, 60, 30], carry a gRNA-only cassette capable of copying when combined with a "static" source of Cas9 inherited in a Mendelian fashion and located at a second locus. Split systems also have been deemed safer for laboratory research[1, 3] and more amenable for localized release purposes[123, 60, 129].

In this study, we design and test several sGDs in *Drosophila melanogaster* with various genetic parameters and strategies either to limit or extend sGD drive potential. We demonstrate that resistance can be largely overcome if the drive allele carries an efficient gRNA and recoded target gene sequences while generating few functional drive-resistant NHEJ alleles. sGDs inserted into target genes essential for survival or fertility were efficiently transmitted through females, but generally to a lesser degree through males. Genomic context (Cas9 promoter, chromosomal location) also contributed to gene-drive efficiency and rates of NHEJ mutagenesis. In multigenerational cage studies, sGDs benefited from the dual effect of a dominant maternal process (lethal/sterile-mosaicism) and classic gradual zygotic Mendelian negative selection. These advantages were amplified in the case of sGDs inserted into loci that demonstrate Cas9-dependent fitness costs. In such cases, sGDs behave as self-limiting and inherently confinable drive systems by virtue of imposing a fitness cost when Cas9 is co-inherited with the drive element.

2.2.2 Results

2.2.2.1 Design of Split Gene-Drive (sGD) Elements in Autosomal Recessive Lethal Loci

We tested whether super-Mendelian transmission efficiencies (i.e., $> 50\%$ expected by Mendelian inheritance) could be achieved in bipartite gene-drive systems inserted into essential gene targets on the II or III chromosomes (Fig. 2.7a, d) targeting conserved DNA coding sequences of recessive lethal (*rab5*, *rab11*, *prosalpha2*) or sterile (*mei-W68*[146], the *D. melanogaster* ortholog of *spo11* - referred from here on as *spo11*) genes. These elements, referred to hereafter as split gene-drives (sGDs), contain a gRNA selected to target sequences at, or close to, critical amino acids (e.g., catalytic centers or membrane-tethering sequences) of each gene of interest to minimize the potential generation of functional NHEJ events; a recoded cDNA portion of the gene that is seamlessly fused to endogenous coding sequences to restore its functionality upon successful insertion; a 3xP3-tdTomato dominant marker, and a partial inactive fragment of Opie2-eGFP (designed to permit future conversion of each sGD into

fGD[126]). All the aforementioned cargo is flanked by 1 kb homology arms to the gRNA cut site to support HDR-mediated integration of the cassette into the genome (Fig. 2.7a). The sGDs were targeted to loci selected based on various criteria including: function and evolutionary conservation (all loci); propitious localization features (*rab5*, *rab11*[188, 36]); broad cellular and/or dosage sensitive requirements (*prosalpha2*[190]). If possible, gRNAs targeted locations near the carboxy-terminus to minimize the extent of recoding necessary (e.g., few terminal amino acids of *rab5* and *rab11*). Also, by choosing essential target genes, more prevalent classes of loss-of-function (LOF) alleles will incur severe fitness costs (e.g., lethality or sterility) resulting from homozygosis or from the potent filtering phenomenon of dominantly-acting somatic lethal/sterile mosaicism[2, 75] (Supplementary Fig. 1).

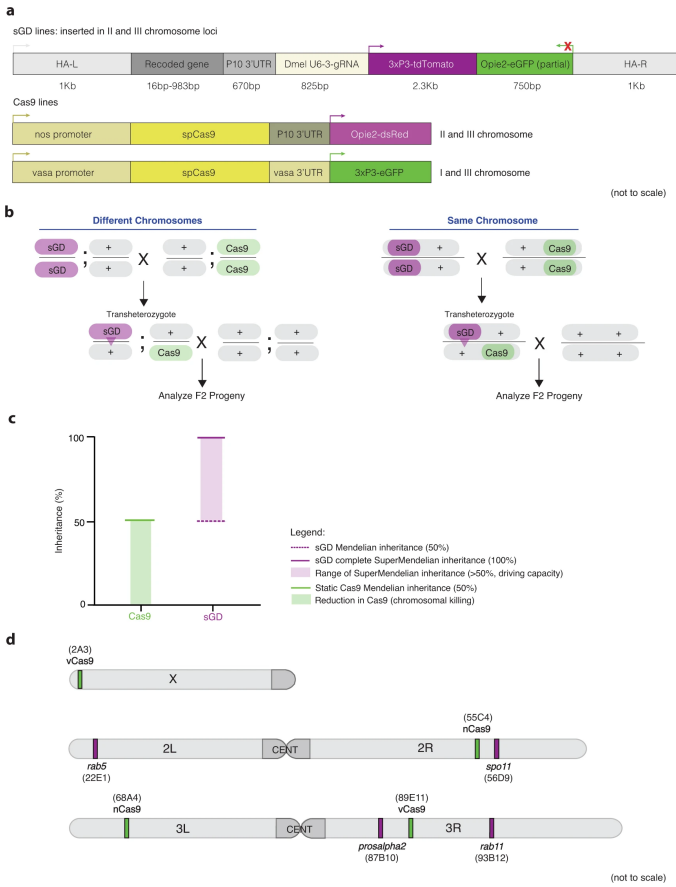

Figure 2.7: **Experimental Design of the Split Gene-Drive System in Essential Loci** (**A**) Schematic of the genetic constructs engineered and tested in the study. All constructs contain a recoded cDNA of the target gene that restores its functionality upon insertion of the transgene, a specific gRNA, and expression of 3xP3-tdTomato. Static Cas9 lines encode a *nanos* or *vasa*-driven Cas9 and a selectable marker, Opie2-dsRed or 3xP3-GFP, respectively. (**B**) Outline of the genetic cross schemes used to demonstrate the driving efficiency of each sGD, comparing systems where the sGD, Cas9 and wildtype (WT, +) are located in the same (right panel) or different chromosomes (left panel). F_1 trans-heterozygotes (carriers of both Cas9 and sGD *in trans*) were singly crossed to WT individuals to assess germline transmission rates by scoring % of the fluorescence markers in F_2 progeny. The conversion event at the sGD locus is shown with a triangle in F_1 individuals. (**C**) Overview of how data is plotted throughout the paper. F_1 germline inheritance is plotted in two independent columns, one that refers to the static Cas9, which should be inherited at Mendelian ratios since it does not have driving capacity, and a second column that displays the biased inheritance of the sGD transgene. Graph contains no empirical data. (**D**) Chromosomal location and insertion sites of all sGD and static Cas9 transgenes in the *Drosophila melanogaster* genome.

2.2.2.2 sGDs Display Super-Mendelian Inheritance

We first tested the frequency at which sGDs copied onto a wildtype (WT) chromosome in combination with a static source of Cas9 provided *in trans* from the X, II or III chromosomes (Fig. 2.7b). All sGD and Cas9 elements were inherited at Mendelian frequencies as separate elements. However, when combined with Cas9 sources inserted in different chromosomes and driven by different promoters (*vasa* - vCas9 or *nanos* - nCas9), sGD cassettes copied to the homologous chromosome leading to super-Mendelian transmission (Fig. 2.8). G_0 sGD homozygous males or virgin females were crossed to Cas9-bearing lines to obtain trans-heterozygous F_1 individuals carrying both transgenes (sGD; Cas9), which we refer to as 'master females' or 'master males'. Master males or females (virgin) were single-pair mated to WT individuals of the opposite sex, and resulting F_2 progeny were assessed for efficiency of sGD copying by scoring the prevalence of red-eye (3xP3-tdTomato, for the sGD element), and inheritance of Cas9-linked fluorescent markers (green eyes: 3XP3-eGFP, for vCas9 or red bodies: Opie2-DsRed, for nCas9).

F_2 progeny displayed super-Mendelian inheritance of sGDs ranging from 64.8% to 99.9%, depending on insertion locus, source of Cas9, and sex of the F_1 trans-heterozygote parent (Fig. 2.8). We detected no differential drive performance between Cas9 being contributed by G_0 males versus females. However, sex-dependent differences were observed for transmission of the *rab5*, *rab11*, or *spo11* sGDs from F_1 parents to F_2 progeny. When passed through F_1 master females, *rab* sGDs were inherited with high efficiency (95.2-97.1% for *rab5*; 97.6-98.5% for *rab11*, Fig. 2.8a and c) for all sGD-Cas9 combinations, while transmission of the *spo11* sGD was somewhat lower (75.6-82.5%, Fig. 2.8b). In contrast, when transmitted from F_1 master males, inheritance of these sGDs dropped by $\sim 15\%$ (76.1-83.4% for *rab5*; 70.9-84.6% for *rab11*; 64.8-69.4% for *spo11*, Fig. 2.8) relative to rates observed with F_1 master females. These trends are readily apparent when data are compiled for each locus across Cas9 sources (*rab5* sGD (t-test): $U = 1010$, $n_1 = 197$, $n_2 = 116$, $p < 0.0001$; *rab11* sGD (t-test): $U = 1052$, $n_1 = 154$, $n_2 = 122$, $p < 0.0001$; *spo11* sGD (t-test): $U = 2613$, $n_1 = 145$, $n_2 = 109$, $p < 0.0001$, Supplementary Fig. 2). An exception to this sex-biased drive was the *prosalpha2* sGD, which was transmitted with similarly high efficiency through F_1 master females (99.3-99.7%, Fig. 2.8d) and F_1 master males (97.4-99.9%, Fig. 2.8d; *prosalpha2* sGD (t-test): $U = 7640$, $n_1 = 143$, $n_2 = 113$, $p = 0.256$, Supplementary Fig. 2).

Several observations and trends can be extracted from the data summarized above. First, observed locus-specific variation in copying efficiencies most likely depends on several factors including the genomic context of the insertion locus, gRNA sequence, the strength (e.g., vCas9-III > nCas9-III $\sim$ vCas9-X > nCas9-II) and degree of germline specificity (nCas9 > vCas9) of the Cas9 source (Supplementary Figs. 1 and 2). Second, the essential nature of the target gene is highly relevant in this context: genes that are broadly required in the organisms for viability (*rab5*, *rab11*, *prosalpha2*) may display varying degrees of lethal mosaicism, depending on whether their activities are required in many or few cells or in a cell-autonomous

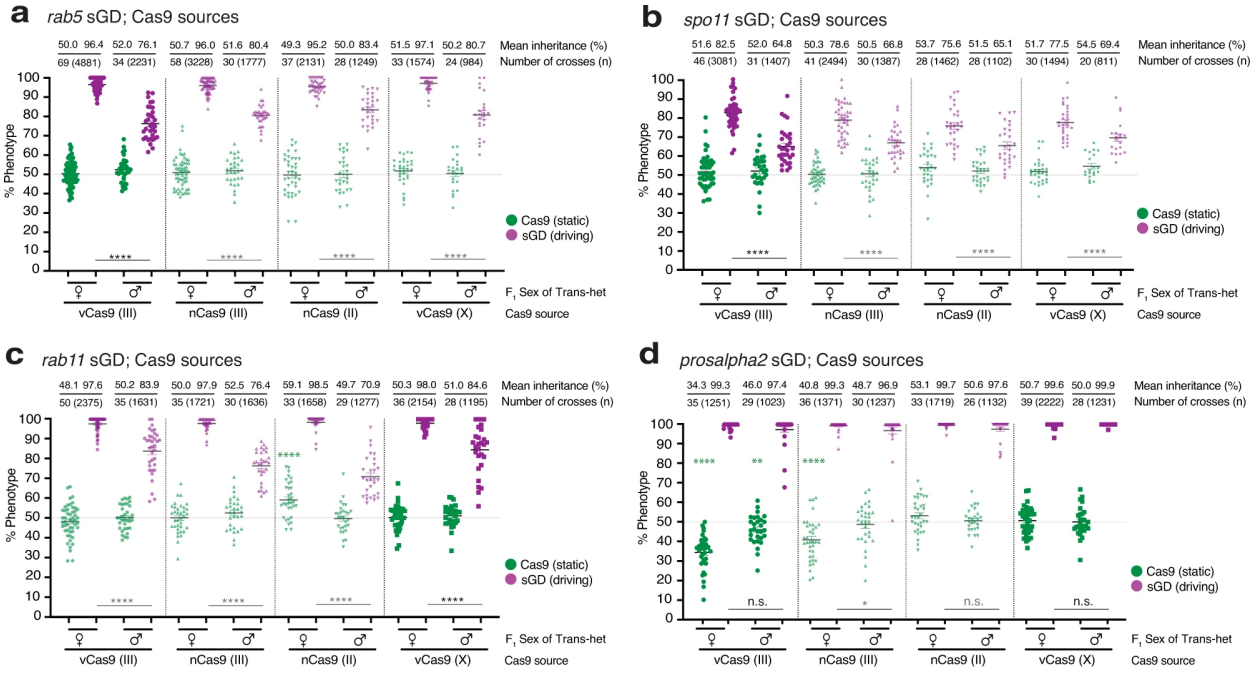

Figure 2.8: **sGD Elements Display Different Super-Mendelian Inheritance Patterns Depending on the Trans-Heterozygote Progenitor Sex**
Genetic crosses performed using the sGD transgenes in combination with *vasa* or *nanos*-driven Cas9 lines located in the first (X), second (II) or third (III) chromosomes. Graph contains data for (**A**) *rab5*; (**B**) *spo11*; (**C**) *rab11*; and (**D**) *prosalpha2* sGDs. Single F_1 germline conversion was assessed by scoring the markers for both transgenes in the F_2 progeny. Inheritance of Cas9 and sGD is depicted using green and purple dots, respectively. Each single cross is shown as a single data point. Values for inheritance mean, number of crosses (N) and individuals scored (n) are shown atop of the graph in line with each respective dataset. Sex of the parental (F_1) trans-heterozygote is indicated in the X-axis. sGD-Cas9 combinations depicted in bold represent pairings that were progressed to cage trials and further genotypic studies. Error bars represent mean values ±SEM. Stars represent statistical significance ($^{****}p < 0.0001$; $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$) for F_1 male-to-female sGD-copying differences (black, two-sided t-test) and to detect Cas9 being inherited below Mendelian frequencies (green, χ^2). Raw phenotypical data is provided as "Supplementary Data 1" .

fashion. These results suggest that a stringent requirement for locus function results in more complete and immediate elimination of LOF alleles (*prosalpha*2 > *rab*5), whereas for other genes (*rab11*) individuals heterozygous for LOF alleles may suffer a less significant fitness reduction. Targets essential for reproduction, such as *spo11*, will generate sterile mosaic progeny that survive but are infertile, and most NHEJ events will not be immediately

eliminated from the population, but will lead to dominant sterility in subsequent generations.

As summarized above, the highly-efficient *prosalpha2* sGD was an exception to the sex-biased pattern of transmission. Crosses with the *prosalpha2* sGD employing the 3^{rd} chromosome vCas9 source, which also served as a marker for the receiver chromosome due to its close linkage to the sGD cleavage site ($\sim$ 2 cM, Fig. 2.7d), revealed another behavior unique to this sGD in which sub-Mendelian inheritance of the GFP-marked Cas9 element was observed (Fig. 2.8d). In all other instances, recombination between the Cas9 source and the sGD (in females) or independent chromosome assortment precluded discrimination between donor versus receiver chromosomes. Allele-specific analysis in the case of the 3^{rd} chromosome vCas9 source revealed that the receiver chromosome was being transmitted at significantly lower rates than expected in both sexes (35-40% versus expected 50%, chi-square goodness-of-fit-in females: $\chi^2(1, N = 1251) = 104.2, p < 0.0001$; in males: $\chi^2(1, N = 1023) = 7.1, p < 0.05$, Fig. 2.8d). In principle, this observation could result from the copying process damaging the receiver chromosome and resulting in the homolog being inadequately repaired and lost prior to fertilization or shortly thereafter. Alternatively, the chromosome-III sources of Cas9 (both vCas9 and nCas9) may be stronger (e.g., higher levels or more timely delivery of Cas9) leading to individuals that carry both the sGD transgene and Cas9 source being less viable than those bearing the drive element alone. Selective sub-Mendelian transmission of Cas9 in master female crosses employing nCas9-III (chi-square goodness-of-fit: $\chi^2(1, N = 1371) = 44.5, p < 0.0001$, Fig. 2.8d) is consistent with a greater relative activity of this Cas9 source, since it is unlinked to the *prosalpha2* sGD locus and should freely recombine with any chromosomal damage incurred at the gRNA cut site. In crosses of the *prosalpha2* sGD to Cas9 sources located on the 2^{nd} or the X chromosomes we did not observe such deficits in GFP inheritance, suggesting that these Cas9 sources may express the Cas9 transgene at lower levels or in a more germline specific fashion than those delivered by the 3^{rd} chromosome sources of Cas9.

We also evaluated the sGD systems for the phenomenon of 'shadow drive', in which maternally-deposited Cas9 in the egg biases the inheritance of a transgene for one extra generation even if the Cas9 element is not transmitted to F_2 females[106, 75]. We crossed F_2 sGD$^+$/Cas9$^-$ virgin females (that only carried maternally-deposited Cas9 protein) to WT males and scored for presence of the eye-tdTomato marker in F_3 individuals. In the absence of shadow drive, F_3 progeny should display Mendelian inheritance of the marker. However, progeny of sGD$^+$/Cas9$^-$ F_2 females exhibited modest levels of super-Mendelian transmission of sGDs, which was most pronounced for the *rab5* sGD$^+$ vCas9-III ($64.8 \pm 11.9\% = \sim 30\%$ conversion, Supplementary Fig. 3). We also performed the reciprocal cross of F_2 sGD$^+$/Cas9$^-$ males to wildtype virgin females. Consistent with little appreciable paternal Cas9 protein being transmitted in the sperm of such F_2 males, sGDs were inherited at expected Mendelian rates in their F_3 progeny ($\sim 50\%, = 0\%$ conversion, Supplementary Fig. 3).

2.2.2.3 Resistant Alleles Do Not Impede sGD Performance

Because lethal mosaicism is hypothesized to dominantly eliminate LOF NHEJ mutations[2, 75], we analyzed production of these alleles in single-generation crosses as well as their predicted elimination in multigenerational cage studies. In single-generation experiments (Fig. 2.8), we recovered individual non-fluorescent male and female F_2 individuals (i.e., flies lacking the sGD dominant marker) that carried potential NHEJ alleles from one parent and a WT allele on the homologous chromosome provided by the other parent. Knowing the WT sequence, we were able to distinguish WT Sanger sequencing chromatogram reads from potential indel alleles. This analysis revealed the production of indels for all genes (including LOF frameshift and *in frame* mutants) as well as intact WT alleles (Fig. 2.9). For the *rab5* sGD, *in frame* NHEJ alleles accounted for 34.5% of all NHEJ events (Fig. 2.9a), equating to 1.2% of total alleles in F_2 progeny. Rab proteins contain carboxy-terminal prenylation sites, consisting of two cysteine residues that form disulfide bonds essential for their localization to vesicular membranes[68]. Some *in frame* NHEJ events mutated one or both of these two cysteines since, by design, the gRNA cleavage sites were chosen to be very near these codons. Consistent with the role of Rab protein tail residues in forming essential disulfide bonds, no cysteine-altering NHEJ alleles were recovered (even as balanced heterozygotes) among 12 tested balanced F_2 alleles despite 31% of the individuals carrying frameshift mutations that presumably disrupt disulfide bond formation and prevent protein localization to vesicle membranes.

Similar categories of mutant alleles were recovered in the *spo11* and *rab11* sGD crosses. The *spo11* sGD generated few frameshift mutations (Fig. 2.9b), and in frame mutations accounted for 7.7% of total F_2 alleles. Since the gRNA cut site for this drive targets a sequence adjacent to critical catalytic residues required for its topoisomerase activity, many of these *in frame* mutants are likely to be sterile when homozygous, as would be frameshift mutations leading to truncated non-functional proteins[146]. *Rab11* has a similar prenylated 3' end structure as *rab5*[120], with dual cysteines located at the tail end of the protein. Interestingly, fewer mutations altered these cysteine residues, even when in a heterozygous condition (13%, Fig. 2.9c). However, a higher percentage (63%, Fig. 2.9c) of the non-fluorescent individuals harbored *in frame* NHEJ alleles, accounting for 1.2% of the overall target alleles, similar to the those recovered for the *rab5* sGD. This marked skew in recovering predominantly *in frame* alleles is consistent with LOF frameshift alleles being created and then immediately eliminated by lethal mosaicism, and validates the strategy of selecting gRNA cut sites in functionally critical domains. In the case of the *prosalpha2* sGD, we did not detect a single WT allele (Fig. 2.9d) among non-fluorescent F_2 progeny, suggesting Cas9-mediated $\sim 100\%$ cleavage at this locus. The very low rate of total recovered NHEJ events, with *in frame* NHEJ events accounting for only 0.16% of the total alleles, correlates with the high transmission frequencies observed in single pair crosses. This high degree of sGD transmission may result from particularly efficient copying at the *prosalpha2* locus and/or individuals carrying NHEJ-induced LOF alleles or damaged chromosomes rarely surviving

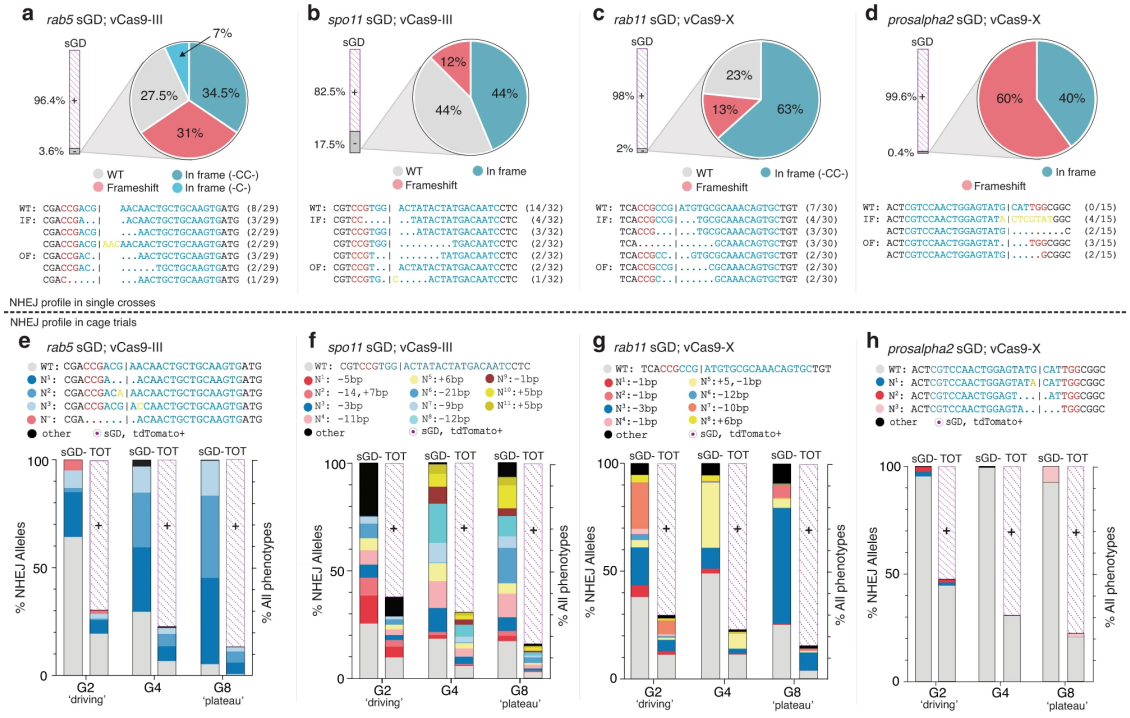

Figure 2.9: **Profiles of NHEJ Events in Single-Generation Crosses Versus Multi-generational Cages**
Production of NHEJ events in single crosses (**a-d**) and cage trials (**e-h**). For single-cross data, target regions were amplified from single non-fluorescent F_2 individuals generated in Fig. 2.8, sequenced through Sanger sequencing and analyzed. A bar depicts the % of sGD^+ (purple) and % of non-fluorescent (sGD^-, gray) flies for each tested locus. Genotypic data is depicted in pie charts representing the prevalence of specific indel mutations in sGD^- individuals for (**A**) *rab5*, (**B**) *spo11*, (**C**) *rab11*, (**D**) *prosalpha2* loci. Each section of the pie chart describes the kind of NHEJ that is formed and its percentage among the total tested sGD^- (NHEJ/WT) heterozygotes. The specific sequence of prominent NHEJ events, along with its corresponding prevalence, is reported under each pie chart. gRNA sequence of each sGD is depicted in blue with its PAM sequence shown in red. To generate the NHEJ cage trial data (**e-h**), non-fluorescent sGD^- individuals were pooled at every generation and used to amplify their target site region, which was deep sequenced to assess formation of NHEJ alleles. WT sequences (gray), *in frame* deletions (blue), frameshift deletions (red) and insertions (yellow) are shown in bars at each generation to represent the distribution of alleles in the sGD^- population (left) and among the total population (sGD^+ and sGD^-, right). Purple diagonally-dotted bars show the sGD^+ population percentage.

(e.g., due to strong lethal mosaicism).

2.2.2.4 sGD Exhibit Differing Degrees of Drive and Cas9 Depletion in Multigenerational Cages

We next assessed the potential of the various sGDs to mediate population modification by testing their performance in small laboratory population cages. G_0 drive-competent master males and master female virgins, heterozygous for both the sGD and vCas9 elements (sGD/+; vCas9/+), were combined with male and virgin female WT flies at a seeding ratio of 1:3 sGD/+; vCas9/+ to WT. Since the drive-bearing flies were heterozygous for the transgenes, this initial introduction corresponds to 12.5% sGD and Cas9 relative to total alleles. Except for the *prosalpha2* sGD trials in which vCas9-III is closely linked to the sGD insertion locus, the sGD and Cas9 elements were unlinked and could segregate independently. At each generation, we scored half of the progeny in a cage for prevalence of both the tdTomato+ (sGD) and eGFP+ (vCas9) elements based on their fluorescent phenotypes. The other half was used to seed the next generation.

The *rab5* sGD (Fig. 2.10a) and *spo11* sGD (Fig. 2.10b) cage trials were conducted for 20 generations using a vCas9-III source. Both drives achieved 80-90% introduction by G_{5-7}, and remained stable with little variation in overall population size or bias in male to female ratios over the remaining generations (Supplementary Fig. 4). The long plateau phases observed in all replicates, extending until G_{20}, most likely reflect an equilibrium being attained between the sGD and a stable percentage of drive-resistant functional NHEJ alleles (Fig. 2.9e, f). The NHEJ allele dynamics differ substantially between sGDs. A class of non-functional resistant allele was observed at G_2 for the *rab5* sGD (Fig. 2.9e) and then disappeared (G_4 and G_8). A few residual in frame cysteine-containing, and potentially functional, resistant alleles fixed in the population. In contrast, for the *spo11* sGD (Fig. 2.9f), similar patterns of NHEJ alleles persisted throughout the sequenced generations. Thus, the *spo11* locus may be significantly less constrained by selection than *rab5*, tolerating a greater array of alleles with reduced activity (*in* or *out-of-frame*). Alternatively, particular *rab5* alleles may become fixed in a biased fashion by outcompeting other NHEJ events that carry greater fitness costs.

Another salient trend in the *rab5* and *spo11* cages was the gradual progressive reduction in the prevalence of the Cas9 transgene in many cage replicates. In the *rab5* sGD experiments, Cas9 prevalence decreased with variable kinetics among cage replicates, whereas similar slopes of Cas9 decrease were observed in all of the *spo11* sGD cage trials. It is notable that following the disappearance of Cas9, whenever it occurred, the prevalence of the *rab5* and *spo11* sGDs remained stable in subsequent generations (Fig. 2.10a, b), indicating that the genetic recoded alleles carried by these elements bore no substantial fitness costs relative to other alleles present in the cages (either functional NHEJ or WT alleles of the target locus). Because we hypothesized that lethal (*rab5*) or sterile (*spo11*) mosaicism may contribute to the drive dynamics, we determined rates of egg-laying, hatching and egg-to-adult ratios for the different sGD±Cas9. All percentages were normalized to hatchability and adults observed on WT crosses. We observed reduced hatchability and development to adulthood

for trans-heterozygote *rab5* sGD; vCas9-III (84% and 73.7%, respectively) and *spo11* sGD; vCas9-III (91.4% and 81.4%) crosses relative to both sGD/+ (95.5% and 92.2%), vCas9-III (100.5% and 95.3%) crosses (Supplementary Fig. 5). The higher hatching rates of the *spo11* sGD drive compared to the *rab5* sGD may reflect the differential effects of lethal versus sterile mosaicism (see 2.2.3).

The drive dynamics for *rab11* sGD cages differed from those of the *rab5* and *spo11* sGDs in that, accompanying an increase of the sGD to a plateau level of 80% prevalence from G_8 through G_{20}, vCas9-X remained constant in frequency (Fig. 2.10c). The higher proportion of *in frame* NHEJ alleles observed in single crosses (Fig. 2.9c) for the *rab11* sGD compared to the *rab5* or *spo11* sGDs may account for its reduced level of maximal introduction, even with stable maintained inheritance of the vCas9-X transgene. Lending further support to this interpretation, deep-sequencing of non-fluorescent flies at different generations (Fig. 2.9g) revealed that a variety of mutations in *rab11* were generated during the drive process, only a small fraction of which fixed in the population by G_8. At that point, nearly all remaining non-fluorescent flies contained a single-codon deleted allele which presumably encoded a functional protein variant lacking a single amino acid (ΔAsp).

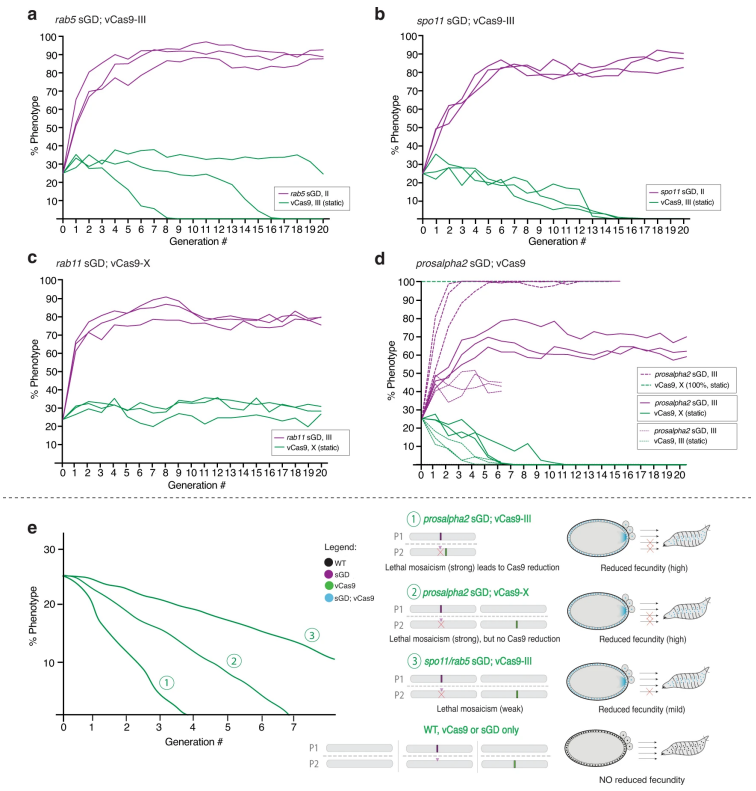

Figure 2.10: **sGD Driving Experiments in Cage Trials**
Virgin sGD/Cas9 trans-heterozygotes and WT flies were seeded at a 1:3 ratio. Each generation, flies in a cage were randomly split in half. One half was scored for eye fluorescence (G_n) while the other was used to seed fresh cages (G_{n+1}). Purple traces indicate sGD^+ progeny, green traces indicate $Cas9^+$ progeny. Experiments were done in triplicate and each line represents a separate cage. (**A**) *rab5* sGD;vCas9-III. The sGD prevalence increases exponentially in the cage ($84 \pm 6\%$) up to G_4 and then plateaus slowly. sGD highest percentage occurs at G_{10} ($92 \pm 4\%$). Cas9 decreases in two of the three cages from 25% to 0% by G_8 and G_{15}, respectively. (**B**) *spo11* sGD;vCas9-III. All three replicates reach their highest prevalence in the cage ($85 \pm 2\%$) by G_{6-7} and then plateau. Cas9 decreases linearly from 25% to 0% by G_{15} in all three replicates. (**C**) *rab11* sGD;vCas9-X. sGD proportions in the cage slowly increase linearly ($83 \pm 6\%$) up to G_8. Cas9 remains steady at seeding levels ($28 \pm 3\%$), suggesting continuous Mendelian transmission. (**D**) *prosalpha2* sGD drive dynamics depend on the location of the static Cas9, as well as seeding ratios. Bold lines reflect sGD;vCas9-X, thin dashes show cage trials using vCas9-III and thicker dashes depict drive in a vCas9-X-saturated population. *Prosalpha2* sGD;vCas9 combinations produce different driving fates and outcomes, providing a flexible tool for deployment. (**E**) Hypothesis on cage and drive behavior of the different sGDs and vCas9 reduction over time. Raw phenotypic scoring is provided as Supplementary Data 2.

2.2.2.5 The *prosalpha2* sGD Displays Differential Drive Behavior Depending on the Cas9 Source

As described above, the *rab5* and *spo11* sGD elements exhibited similar drive trajectories in cage experiments, consistent with their performance in single-generation crosses (considering also that the *spo11* sGD is expected to benefit from sterile mosaicism a generation later than *rab5* sGD, which profits from lethal mosaicism). In contrast, *prosalpha2* sGD; vCas9-III cages displayed a curtailed trajectory in which the drive leveled off much sooner than for the other drives (Fig. 2.10d, thin dashed lines). This population-level behavior was surprising as it contrasted markedly with the efficient transmission of this sGD through both males and females in single-generation crosses, where it displayed the greatest drive potential, using the vCas9-III source (>99%). Yet, with the same Cas9 strain in cages, this sGD achieved only a modest level of introduction (Fig. 2.10d, thin dashed lines). Another notable feature of all three cage replicates was a very rapid decline of the Cas9 transgene (lost in all cages by G_6), in line with the sub-Mendelian (<50%) transmission of the genetically-linked vCas9-III transgene observed in single crosses (Fig. 2.8d). We further analyzed the drive performance of the *prosalpha2* sGD using an X-linked source of Cas9 (vCas9-X), unlinked to the sGD transgene. vCas9-X also declined over time when combined with the *prosalpha2* sGD, although did so more gradually than with its vCas9-III counterpart (Fig. 2.10d, bold lines). Accordingly, the *prosalpha2* sGD reached higher levels of introduction with the unlinked vCas9-X (60-70%, Fig. 2.10d, bold lines) than it did with vCas9-III, yet still fell short of that achieved by the *rab5* sGD or *spo11* sGD driven by the action of vCas9-III. However, unlike the *rab5* and *spo11* sGD drives, in which functional NHEJ alleles were generated and selected, very few such NHEJ alleles accompanied the partial spread of the *prosalpha2* sGD (Fig. 2.9h). We observed the transient generation of a small number of varied mutations at G_2 and G_4. By G_8, nearly all remaining non-drive alleles had the WT sequence that should remain susceptible to conversion. The decline in vCas9-X in these experiments contrasts with its stable maintenance in *rab11* sGD in cages, and with its Mendelian inheritance in single-generation *prosalpha2* sGD crosses. Hatching rates (79.8%) from *prosalpha2* sGD; vCas9-X mothers revealed the greatest reduction among all sGDs relative to heterozygous *prosalpha2* sGD/+ (99.2%) and vCas9-X/+ (99.4%) controls (Supplementary Fig. 5). Similarly, decrements in embryonic survival (65.3%) and adult viability (80% of the hatched larvae) were observed. This substantial reduction in fecundity most likely contributes to the observed cage dynamics, particularly when acting over several generations in populations mating randomly.

Since the *prosalpha2* sGD displayed the greatest transmission in single-generation crosses (>99%, Fig. 2.8d), we wondered whether we could increase its level of introduction in population cage experiments given that nearly all non-drive alleles that remained following elimination of the Cas9 source were WT. We tested this possibility by seeding *prosalpha2* sGD; vCas9-X drive competent individuals into a population of pure vCas9-X homozygous flies. This scenario forces maintenance of Cas9 in the population, approximating a full drive

configuration in this respect. We seeded cages at the same 25% trans-heterozygous sGD; vCas9 rate as in the other cage experiments, but replaced the WT population with vCas9-X homozygous flies, thereby increasing initial Cas9 prevalence to 100%. In contrast to the self-attenuated drive manifested in the previous examples, the sGD now rapidly achieved complete fixation in 3 generations in 2 of the 3 cages, and in 5 generations in the third cage (Fig. 2.10d, thick dashed lines). No NHEJ events were recovered from the *prosalpha2* sGD; vCas9-X cages (100% vCas9) analyzed for these events. This absence of mutant alleles is consistent with the highly-efficient drive trajectory of the *prosalpha2* sGD in a Cas9 background and with the large fraction of unaltered target alleles remaining in cages when this sGD was seeded into WT populations. The divergent drive outcomes under these differing scenarios suggests a hypothesis for behavior of the *prosalpha2* sGD (Fig. 2.10e) that may also pertain in a less potent mode to other sGDs, as discussed below.

2.2.2.6 Modeling Captures Varied Cage Drive Dynamics and Reveals a Gradient of Allelic Fitness Costs

We generated four mathematical models to account for the results observed in single generation crosses and cage trials of sGDs inserted into: 1-an autosomal locus required for fecundity (*spo11*, vCas9-III), 2-an autosomal sGD locus essential for zygote viability (*rab5*, vCas9-III), 3-linked autosomal sGD and Cas9 elements with the sGD inserted in a gene essential for zygote viability (*prosalpha2*, vCas9-III), and 4-an X-linked Cas9 source combined with sGDs required for zygote viability (*rab11* and *prosalpha2*, vCas9-X). Drive efficacy was refined and fitness cost parameters were estimated from respective cage trial data using predictive mathematical models and a likelihood-based Markov chain Monte Carlo (MCMC) algorithm, with initial parameter estimates taken from single-pair mating data (Supplementary Figs. 6–11). Model-estimated rates for cleavage and accurate HDR events were consistent with frequencies observed in single-pair crosses shown in Fig. 2.8, except for *rab11* (Supplementary Tables 8–13). Modeled HDR frequencies for *rab11* females were significantly lower ($\sim$ 8%) than those observed in single-pair crosses, and slightly lower in males ($\sim$ 2%). Also, we note that the initial drive ascent in the modeling is somewhat less than the observed experimentally for the several of the sGDs, which may reflect an initial mating bias with females transmitting the gene drive more efficiently than males as seen in single-generation crosses. Consistent with such a density-dependent phenotype mechanism, the modeling more closely resembles the early dynamics of the *prosalpha2* sGD experimental cages, which display little sex-biased transmission in single crosses. The parameters describing the proportion of generated NHEJ alleles that are functional vs. non-functional generally had a broad credible interval (CrI), e.g., for the X-linked Cas9-*rab11* split-drive design, the proportion of functional NHEJ alleles that were retained under selection was estimated to be 95.7% (95% CrI: 69.8-99.9%) in females and 99% (95% CrI: 75.1-99.9%) in males. Lack of resolution for this parameter stems from the high rate of accurate HDR and hence relatively low number of NHEJ resistant alleles. Estimated fitness costs demonstrate the potential for deliberate tuning through drive design and target gene selection. Different target genes showed various

fitness costs, with *rab11* sGD estimated to have the lowest associated cost combined with Cas9 (0.0%, 95% CrI: 0.0-0.1%), followed by the *rab5* sGD (4.8%, 95% CrI: 1.0–5.1%) and *spo11* (11.3%, 95% CrI: 10.6-12.3%), and the *prosalpha2* sGD with the highest cost (15.8%, 95% CrI: 15.1-16.2%) (Supplementary Tables 8–13). Using two designs for prosalpha2 (an autosomal split-drive with vCas9 located on the same chromosome (III) and a second using an X-linked vCas9), we found that targeting a chromosome *in cis* was costlier than targeting a chromosome *in trans* for this locus (cost of 27.5%, 95% CrI 26.0–28.3% *in cis* c.f. 15.8%, 95% CrI 15.0–16.2% *in trans*). Collectively, these findings suggest it may be possible to vary fitness costs by design through thoughtful pairing of a target gene and Cas9 source, with the caveat that fitness costs measured under laboratory conditions may differ from those experienced by organisms in the field.

Finally, employing model-estimated drive efficacy and fitness cost parameters (Supplementary Figs. 6–11), predictions from a stochastic implementation of the model were compared to experimental cage trial data (Fig. 2.11). The stochastic model captured the potential role of chance events such as mate choice (multinomial-distributed), egg production (Poisson), progeny genotype (multinomial), and the finite sampling of the next generation (multivariate hypergeometric). Stochastic model trajectories were consistent with the observed experimental data for each construct, both in terms of mean behavior and stochastic variation. This provides reassurance that: (1) the model is capturing dynamic features pertaining to each of the drive systems, (2) the variation observed in the laboratory cage trials can be explained by stochasticity given cage population sizes and fitness costs involved, and (3) plausible relative fitness costs associated with sGD+ Cas9 combinations can be inferred.

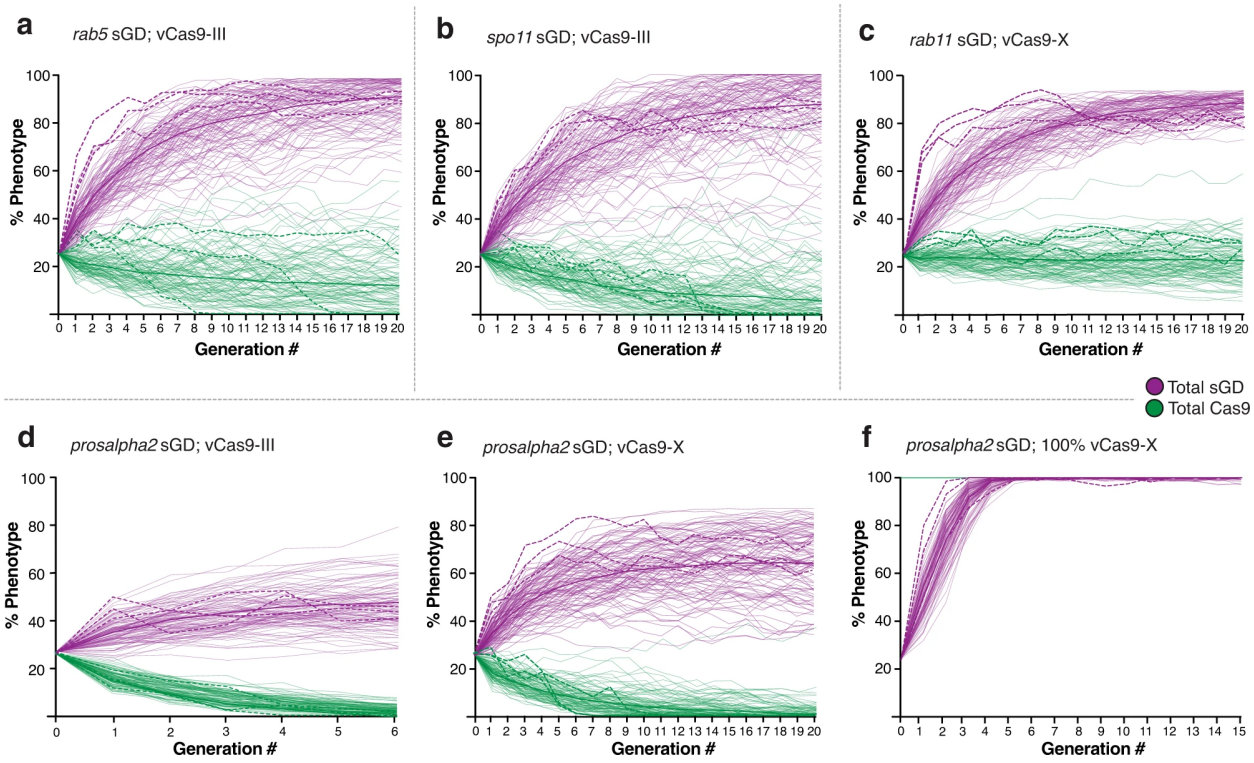

Figure 2.11: **Mathematical Model Simulations Recapitulate Cage Trial Experimental Data**

Four mathematical models were designed based on target gene biology and behavior: split drive in autosomes targeting viability (**A**), split drive in autosomes affecting fecundity (**B**), linked split drive (**D**) and X-linked Cas9 split drive targeting viability (**C**, **E**, **F**). Simulations were run using estimated and fitted parameter values and 100 stochastic model realizations depicted in thin purple (sGD) and green (Cas9) lines. Thicker lines show the mean of those 100 simulations. For comparison, dashed curves represent the collected experimental data.

2.2.3 Discussion

In this study, we design and test a range of inherently confinable split drives targeting essential recessive genes. These sGD elements carry recoded target gene sequences that both mitigate the formation of cleavage-resistant alleles and lead to the loss of separately-encoded Cas9 transgenes in population experiments. Specifically, the elimination of cleavage-resistant alleles relies on LOF NHEJ events being rapidly eliminated after NHEJ alleles are created by dominantly acting maternal lethal-sterile mosaicism[2, 75], followed by gradual culling by standard Mendelian homozygosis and/or moderate haplo-insufficient phenotypes. The rapid loss of Cas9 transgenes appears to depend on two putative mechanisms: (1) reduced viability or fecundity of individuals carrying both a Cas9 source and certain sGDs due to partial haplo-insufficiency of the locus and/or off-target Cas9/gRNA affects, and (2) damage to the Cas9/gRNA-targeted chromosome (if linked genetically to the Cas9 source). We hypothesize that these distinct mechanisms may reduce the overall fitness of sGD+; Cas9+ individuals.

All sGDs tested in single-generation crosses displayed significant super-Mendelian inheritance ranging up to $\sim$ 100%. One clear trend from these experiments was that all sGDs (except *prosalpha2*) exhibited significantly higher inheritance rates ($\sim$ 15%) when transmitted by F_1 master females than by F_1 master males. Since many components of the HDR DNA repair pathway are shared between repair of damage-induced DSB and meiotic recombination[77], these shared features may underlie both the peculiar lack of male recombination in *Drosophila* and the reduced rates of HDR-mediated gene conversion resulting in the divergent sex-specific sGD drive frequencies we observed. This phenomenon was also observed in other studies using a full gene drive[33], as well as in a trans-complementing framework[129]. Why this sex difference was not also observed for the *prosalpha2* sGD will require further analysis, but may be related to a combination of highly-efficient copying and particularly strong forms of lethal mosaicism and/or haplo-insufficiency that eliminate nearly all LOF alleles (NHEJs or damaged chromosomes) when generated. It is noteworthy that when the *prosalpha2* sGD drive was placed *in trans* to a vCas9-III source located on the homologous 3^{rd} chromosome at a site tightly linked to the *prosalpha2* locus, the vCas9 source was inherited at a sub-Mendelian rate in both males and females, suggesting that the target chromosome may have been damaged by an imperfect repair process leading to its loss during or shortly after transmission. Additionally, the two III-chromosome sources of Cas9 may be expressed either at higher levels or in temporal patterns that promote greater overall lethal/mosaic activity.

The different sGD; Cas9 combinations displayed varying levels of drive in population cages. The *rab5* and *spo11* sGDs demonstrated intermediate levels of drive resulting in introduction of the sGD into 80-95% of the population accompanied by a gradual loss of the vCas9-III transgene. This Cas9 elimination mechanism (summarized above regarding the *prosalpha2* sGD) is inferred to operate at varying levels of intensity for the *prosalpha2*, *rab5*, and *spo11* sGDs since we only observed multigenerational loss of Cas9 or reductions in egg-laying and

hatching when the sGD element and Cas9 were carried together (when sGD conversion occurs), but not for each of the separate elements. These results demonstrate that sGDs targeting genes essential for viability or fertility can effectively spread in a super-Mendelian fashion, provided that a recoded version of the gene is carried by the propagating sGD element and that the gRNA site is chosen to target functionally-critical elements of the gene so as to generate as few functional NHEJ alleles as possible. Such loci can be used as docking sites for a split form of drive as shown here, as well as a full gene-drive system or in other strategies such as integral gene drives[153], Medea-like toxin-antitoxin systems[35, 20, 5, 200] or daisy-chain drives[155]. We note that our system was designed to be readily converted to a full-drive configuration using "homology assisted CRISPR knock-in"[126] to create a self-spreading full-drive element (gRNA and Cas9 would spread together), which is similar in outcome to the RMCE strategy in which a Cas9 transgene was mobilized into pre-existing docking sites[80]. Indeed, we have obtained preliminary proof-of-principle data that such conversion to autonomous drives can be achieved (for the *spo11* sGD), and a comparison between such full and split-drive versions of these recoded drives will be assessed in future studies.

The behavior of the *prosalpha2* sGD in population cages did not conform to simple expectations based on its performance in single-generation experiments. We observed different behaviors of this sGD depending on the source of Cas9. When seeded with an autosomal vCas9-III at a 25% initial frequency, little sGD drive was observed and the Cas9 transgene was rapidly eliminated from the population (within 3 generations). This precipitous decrease in Cas9 levels may reflect the damage incurred by the receiver chromosome, evidence for which was observed in single-generation crosses (Fig. 2.8d), as well as by fitness costs associated with carrying both elements. As noted above, the vCas9-III source also declined to varying degrees in different cage replicates when coupled with the other sGDs. Combining the *prosalpha2* sGD with an unlinked Cas9 source (vCas9-X) resulted in increased drive, but fell short of the levels attained by the *rab5* and *spo11* sGDs, most likely due to the fitness costs of carrying vCas9-X and the sGD, which again resulted in speedy elimination of the vCas9-X element from the population. It is noteworthy that the vCas9-X source remained stable in the presence of the *rab11* sGD, indicating that this Cas9 source carries little if any fitness burden on its own or coupled with this particular sGD. Importantly, in *prosalpha2* sGD cages employing the less aggressive vCas9-X source, most of the remaining non-fluorescent fly population carried WT alleles. This retention of WT alleles suggests that further spread of the sGD transgene should occur upon a second introduction of trans-heterozygote or Cas9-only flies in the population, providing a serially-dosable dynamic drive system. Indeed, this prediction was born out when the *prosalpha2* sGD was seeded into a homozygous vCas9-X population, leading to a rapid and complete introduction of the drive ($\sim$ 3 generations with a 12.5% allelic seeding frequency).

We offer the following hypothesis to account for *prosalpha2* sGD cage results (Fig. 2.10e) and their relevance to performance of the other sGDs. According to this model, the *prosalpha2*

sGD incurs two types of drive-dependent fitness costs in different scenarios. When crossed to vCas9-III, it both damages the target chromosome (as evident in the sub-Mendelian inheritance of the Cas9 target chromosome in single-generation crosses) and also induces a significant heterozygous fitness cost. The latter fitness cost, estimated by modeling, may be associated with efficient and penetrant lethal mosaicism leading to moderate haplo-insufficient and/or off-target phenotypes and consequent potent selection against individuals carrying both Cas9 and the *prosalpha2* sGD. These two processes act in concert to rapidly reduce Cas9 prevalence. When the *prosalpha2* sGD is combined with the unlinked and presumed less active vCas9-X source, it does not inflict appreciable damage to the Cas9 chromosome (which displays Mendelian transmission), but still generates a sGD/Cas9-dependent fitness lethal mosaic cost resulting in a more gradual loss of the Cas9 transgene. This, combined sGD/Cas9 effect is also displayed, albeit in milder forms (estimated again by modeling), by the *rab5* and *spo11* sGDs.

In aggregate, this study offers insights into important design features of the sGD system that can be exploited to achieve specific levels of spread and confinement of these systems when applied into native populations. Thus, modest introduction of the sGD followed by rapid removal of the endonuclease can be achieved by placing the *prosalpha2* sGD across from a strongly-linked Cas9 source. Intermediate levels of drive result from moderating the intensity of lethal mosaicism or by employing sterile mosaicism to delay fitness penalties by one generation. Full population introduction can be attained by increasing the proportion of a moderately active Cas9 source to the *prosalpha2* sGD. Such tunable systems could serve as updating platforms to sustain allelic drives[75] that bias inheritance of beneficial traits such as susceptibility to insecticides, or to disseminate cargo with desired features such as new rounds of anti-pathogenic molecules to supplement those carried by an original modification drive[99, 19].

2.2.4 Methods

2.2.4.1 Plasmid Construction

All plasmids were cloned using standard recombinant DNA techniques. Recoded cDNA fragments were designed by using non sub-optimal alternative codons from cut site to stop codon and thus maintaining the exact amino acid sequence of each target gene upon insertion. In all instances, codon usage was kept as similar as possible to that of the endogenous sequence. All fragments were synthesized as gBlocks™ (Integrated DNA Technologies) and cloned into the desired vector. Plasmid and genomic DNA sequences were amplified using Q5 Hotstart Master Mix (New England Biolabs, Cat. #M0494S) and Gibson assembled with NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs, Cat. #E2621). Resulting plasmids were transformed into NEB 5-alpha chemically-competent *E. coli* (New England Biolabs, Cat. #C2987), isolated and sequenced. Primer sequences used for the creation of the different plasmids can be found in Supplementary Table 1.

2.2.4.2 Microinjection of Constructs

Plasmids were purified using the PureLink Fast Low-endotoxin Maxi Plasmid Purification kit (ThermoFisher Scientific, Cat. #A35895). All constructs were fully re-sequenced prior to injection. Embryo injections were carried out at Rainbow Transgenic Flies, Inc. (http://www.rainbowgene.com). Each gRNA construct was injected into a vasa-Cas9 expressing line in the 3^{rd} chromosome (Bloomington #51324).

Injected embryos were received as G_0 larvae, were allowed to emerge and 3-4 females were intercrossed to 3-4 males in different tubes. G_1 progeny were screened for the eye-tdTomato positive marker that indicates transgene insertion. All transgenic flies that displayed the red marker were then balanced using Sco/CyO (for genes located on the II chromosome) or TM3/TM6 (III chromosome) and kept on a w^{1118} background. Homozygous stocks were kept in absence of any balancer alleles. Correct insertions in homozygous transgenic stocks were confirmed through Sanger sequencing using primers found in Supplementary Table 1.

2.2.4.3 Fly Genetics and Crosses

Fly stocks were kept and reared on regular cornmeal medium under standard conditions at 20-22°C with a 12-hour day-night cycle. sGD housekeeping and crosses were performed in glass vials in an ACL1 fly room, freezing the flies for 48 h prior to their discard. To assess each locus' copying efficiency, we genetically crossed each sGD construct to different Cas9 lines. Since our genes of interest are autosomal, individual trans-heterozygote F_1 males or virgin females were collected for each G_0 cross and crossed to a wildtype fly of the opposite gender. Single one-on-one crosses were grown at 25°C. Inheritance of both gRNA ($> 50\%$) and Cas9 ($\sim 50\%$) were calculated using the resulting F_2 progeny by scoring the phenotypic markers associated to each transgenic cassette.

2.2.4.4 Multigenerational Cage Trials

All population cage experiments were conducted at 25°C with a 12-hour day-night cycle using 250 ml bottles containing standard cornmeal medium. Crosses between flies carrying the gRNA and flies carrying vCas9 (X or III) were carried out to obtain F_1 trans-heterozygotes used to seed the initial generation. Wildtype or trans-heterozygote males and virgin females were collected and separately matured for 3-5 days. Cages for all loci were seeded at a phenotypic frequency of 25% gRNA/Cas9 trans-heterozygotes (15 males, 15 females) to 75% w^{1118} (45 males, 45 females). In each generation, flies were allowed to mate and lay eggs for ~ 72 h, when parents were removed from the cage (G_n), and kept for 10 days. Subsequent progeny (G_{n+1}) were randomly separated into two pools and scored; one was collected for sequencing analyses while the other was used to seed the following generation. If the two pools differed much phenotypically, frequencies were averaged in order to reduce variability and stochastic extremes. This process of sampling and passage was continued for 10-20 generations.

2.2.4.5 Molecular Analysis of Resistant Alleles

To extract fly genomic DNA for single fly resistant allele sequence analysis, single flies were squashed in lysis solution (10 mM Tris-Cl pH 8.2, 1 mM EDTA, 25 mM NaCl and 0.2 mg/ml proteinase K), incubated at 37°C for 30 min and deactivated at 95°C for 2 min, as described previously[67]. After extraction, each sample was diluted 1:3 (sample:H2O) and stored at -20°C if needed. 1-2 μl of each diluted DNA extraction was used as template for a 25μl PCR reaction that covered the flanking regions of the gRNA cut site, which were used to sequence the alleles. Sanger sequencing in individual non-fluorescent flies was performed at Genewiz, Inc. in San Diego, CA to obtain the single-cross NHEJ data. NHEJ allele sequences were obtained from Sanger chromatographs by isolating the WT sequence first and then annotating the remaining allelic sequence. For cage trials, 20 non-fluorescent flies were pooled together and DNA was extracted at each sampling generation (2, 4 and 8). Target sequences were amplified using specific gene primers that also contained adapter sequences. Non-fragmented amplicons were sequenced using Illumina-based technology (2x250bp reads, Amplicon-EZ, Genewiz), with final data being delivered as FASTQ reads and aligned to a reference sequence for each gene of interest to detect indel formation. For all datasets, an average of 72,594±8712 (mean±SD) paired-end amplicon reads were obtained per sample.

Primer sequences used for either single fly or cage trial deep sequencing analyses can be found in Supplementary Table 1.

2.2.4.6 Viability Assays

For embryo viability counts, 2 to 3-day old virgin female flies were mated to wild-type males for 24 h and 15 females (in triplicate) were placed in egg-collection chambers to lay eggs during a 18 h period, then the laying plate was removed and eggs counted. All embryos were counted and kept on an agar surface at 20°C for 48 h and hatchability of those eggs was calculated at that point by counting the unhatched embryos. In all plates, females laid between 211 and 332 eggs. Each experiment was carried out in triplicate, and the results presented are averages from these three experiments. For adult fly counts, the larvae obtained in each embryo count assay replicate were transferred from egg collection plates to 250 ml bottles containing modified cornmeal medium. All adult flies that emerged from these bottles were counted and the results of the three replicates for each experiment averaged together. Egg-to-larvae and egg-to-adult ratios were calculated by normalizing the experimental values to the survival observed in parallel experiments carried out with WT flies.

2.2.4.7 Mathematical Modeling

Model fitting was carried out for all four constructs and six distinct cage trials, using a discrete-generation adaptation of the Mosquito Gene Drive Explorer (MGDrivE)[177]. A likelihood-based Markov chain Monte Carlo (MCMC) procedure was used to estimate gene drive efficacy and genotype-specific fitness costs, providing Maximum a Posteriori (M.A.P.)

estimates and 95% credible intervals for each parameter of each distinct cage trial[91]. Mendelian inheritance was assumed except under co-occurrence of the Cas9 and gRNA constructs, when the split-drive design allowed active cleavage of the target chromosome and the possibility of super-Mendelian inheritance. When cleavage occurred, a fraction of the cut alleles could be properly repaired via HDR, and the remaining cut alleles underwent NHEJ repair, generating *in* or *out-of-frame* resistant alleles. The effects of shadow-drive, in which Cas9 protein is deposited in the embryo of a female individual who does not carry the Cas9 allele, but whose own mother does, were also accommodated. Fitness costs were implemented as fractional reductions in male and female fecundity (*spo11*) or male-mating competitiveness and egg viability reductions (*rab5*, *rab11*, and *prosalpha2*), again only under active cleavage conditions with both the Cas9 and gRNA alleles present. In the model, it was assumed that any individual with two *out-of-frame* alleles was completely infertile (*spo11*) or did not mature past the egg stage (*rab5*, *rab11*, *prosalpha2*). More details on the model implementation and likelihood function used in the model fitting can be found in the Supplementary Information file (as a Supplementary Method). The results of each fit, using the estimated parameters, are plotted in Supplementary Figs. 6-11, along with corresponding experimental cage trial data. Parameter descriptions and estimates for each cage trial are provided in Supplementary Tables 2-13.

Simulated model trajectories for Fig. 2.11 were generated using a stochastic implementation of the discrete-generation model. At each generation, adult females mate with males, thereby obtaining a composite mated genotype (their own, and that of their mate) with mate choice following a multinomial distribution determined by adult male genotype frequencies, modified by mating efficacy. Egg production by mated adult females then follows a Poisson distribution, proportional to the genotype-specific lifetime fecundity of the adult female. Offspring genotype follows a multinomial distribution informed by the composite mated female genotype and the inheritance pattern of the gene drive system. Sex distribution of offspring follows a binomial distribution, assuming equal probability for each sex. Female and male adults from each generation are then sampled equally to seed the next generation, with sample size proportional to the average size of the cage trials at that generation, following a multivariate hypergeometric distribution. All simulations were performed and analyzed in R[169].

2.2.5 Safety Measures

All sGD crosses were performed in glass vials in an ACL1 facility, in accordance with the Institutional Biosafety Committee-approved protocol from the University of California San Diego. All vials were frozen for 48 h prior to autoclaving and discarding the flies.

2.2.5.1 Figure Generation and Statistical Analysis

Graphs were generated using Prism 8 (GraphPad Software Inc., San Diego, CA) and potentially modified using Adobe Illustrator CS6 (Adobe Inc., San Jose, CA) to visually fit the rest of the non-data figures featured in the paper.

Conclusion

This chapter demonstrates two of several[165, 104, 208, 165] collaborations at the forefront between gene drive design and mathematical modeling. These projects demonstrate the utility of a flexible model that can be tightly integrated with experimental designs to maximize information gain. However, they also demonstrate deficiencies and suggest directions for improvement. Gene drives with large genotype space[208] illustrate the need for more efficient or more scalable inference procedures. Towards this end, future projects may test evolutionary algorithms[151, 9, 197, 143], a stochastic, gradient-free optimization method that draws inspiration from the biological problem we are studying, genetic inheritance, and has highly-parallel implementations for efficient computation. Additionally, several projects have issues with parameter identifiability. Future work could incorporate multi-objective Markov chains[125] into the fitting procedure to provide orthogonal views of the underlying process. Alternatively, one could apply methods for filtering parameters that are unidentifiable due to model or data limitations. Power analysis[127] are appropriate for this, as are global sensitivity analysis, with the addition that sensitivity analysis are capable of assigning importance to the variables estimated[176, 134, 168, 147, 143, 197].

Chapter 3

Predictive Simulation Applying Gene Drives to Real-World Scenarios

Introduction

Chapter 1 introduced novel simulation frameworks for exploring gene drives. Chapter 2 surveyed cutting-edge analysis of lab-based gene drive experiments. This final chapter integrates and applies that work to forecast the behavior of gene drives in realistic landscapes. We begin this chapter by combining our computational framework with previous work releasing mosquitoes carrying an intracellular bacteria[95] to predict the population dynamics of two recent gene drive systems on a large landscape in Australia. Our goal is to predict the situational effort required for population replacement, in terms of the number of releases performed and the amount of community support for the effort. However, as there exists a possibility of unforeseen behavior that may require the removal of a gene drive, we also interrogate the difficulty in undoing our work and restoring the population to a pre-gene drive state.

We end this chapter by exploring one of the largest gene drive designs we have tested, capable of rapid introgression of transgenes and subsequent removal of driving components. The trans-complementing gene drive[129] builds on the classic split gene drive system[123] by driving both pieces of the split drive. Additionally, we add a fitness cost against the Cas9 portion, testing the impact on removal of Cas9 from the population once the drive is complete. This creates a drive capable of rapid spread into a population and autonomous removal of the driving pieces, without additional effort by the management program or the community, potentially making this design easier to implement and more effective than previous constructs.

3.1 Modeling Confinement and Reversibility of Threshold-Dependent Gene Drive Systems in Spatially-Explicit *Aedes aegypti* Populations[1]

3.1.1 Background

The discovery of CRISPR and its application as a gene editing tool has enabled gene drive systems to be engineered with much greater ease[50, 28]. Recent attention has focused on homing-based drive systems and their potential to control mosquito-borne diseases on a wide scale, either by spreading disease-refractory genes[62] or by spreading genes that confer a fitness load or sex bias and thereby suppress mosquito populations[80, 114]. The increased ease of gene editing has also advanced the entire field of gene drive, including systems appropriate during the trial phase of the technology[139]. Such systems would ideally be capable of enacting local population control by (a) effectively spreading into populations to achieve the desired epidemiological effect, and (b) being recallable from the environment in the event of unwanted consequences, public disfavor, or the end of a trial period.

As gene drive technology has progressed, a number of systems have been proposed with the potential to enact localized population control without spreading on a wide scale[139, 117]. Sterile male releases provide one option[6], a recent version of which is based on the same molecular components as CRISPR gene drive systems[6, 107]. At the interface between homing-based and localized suppression systems, an autosomal X chromosome-shredding system has been proposed that induces a transient male sex bias and hence population suppression before being selected out of the population[59]. Population modification drive systems that display transient drive activity before being eliminated by virtue of a fitness cost could also spread disease-refractory genes into populations in a localized way. Examples of this variety of drive system include split-gene drive[3], daisy drive[155], and killer-rescue systems[70]. Each system has its own strengths and weaknesses and could be suited to a different situation. In this paper, we theoretically explore the potential for two recently engineered threshold-dependent gene drive systems to achieve localized and reversible population modification in structured populations-reciprocal chromosomal translocations[21] and a toxin-antidote-based system known as UD^{MEL} [4].

Threshold-dependent gene drive systems must exceed a critical threshold frequency in a population in order to spread. Based on this dynamic, simple population models, in which two randomly mating populations exchange small numbers of migrants with each other, predict that these systems can be released at high frequencies in one population and spread to near-fixation there, but never take off in the neighboring population because they do not exceed the required threshold there[140, 7]. These systems can also be eliminated through

[1] This work has been previously published[178]

dilution to sub-threshold levels with wild-type organisms at the release site, making them excellent candidates for the trial phase of a population modification gene drive strategy, or when localized population modification is desired[140]. Elimination of a non-driving transgene can in fact be more difficult, as the dynamics of threshold-dependent systems actively drive them out of populations at sub-threshold levels. However, whether these dynamics hold in real ecosystems depends crucially on the dispersal patterns and population structure of the species being considered. First steps towards modeling the spatial dynamics of these systems have been taken by Champer et al.[32], who model spatially structured releases of various threshold-dependent systems without life history, and Huang et al.[97], who model engineered underdominance[43] on a grid-based landscape incorporating life history for *Aedes aegypti*, the mosquito vector of dengue, Zika, and other arboviruses.

Here, we present a detailed ecological analysis of the expected population dynamics of two recently engineered threshold-dependent drive systems, translocations and UD^{MEL} , in *Ae. aegypt* in a well-characterized landscape-Yorkeys Knob, a suburb approximately 17 km northwest of Cairns, Australia (Fig. 3.1c)-suitable for confineable and reversible releases. Yorkeys Knob and the nearby town of Gordonvale were field sites for releases of *Wolbachia*-infected mosquitoes in 2011[95], and the prevalence of *Wolbachia* infection over time provided information on the number of adult *Ae. aegypti* mosquitoes per household and other mosquito demographic parameters for that location[170], as well as an opportunity to validate our modeling framework. *Wolbachia* is an intracellular bacterium and biocontrol agent that biases inheritance in its favor if infected females are present, and blocks transmission of dengue and other arboviruses[199]. Yorkeys Knob is a partially isolated suburb, separated by a 1-2-km-wide uninhabited, vegetated area from the nearest suburb, Trinity Park, a control site for the *Wolbachia* trial. This allowed us to simulate trials of transgenic mosquitoes in a well-characterized population, while also theoretically exploring their potential of spread to a neighboring community.

A number of other ecological details are relevant to the spread of threshold-dependent gene drive systems that have not been considered in previous modeling studies. Perhaps of greatest importance, the frequency of the introduced transgene in the mating pool is markedly different from the frequency of introduced adults. It is typical to release only male mosquitoes as part of an intervention, as only females are involved in human disease transmission. Life cycle and mating structure therefore become relevant, as immature life stages are not available for mating, and female adults are thought to mostly mate only once soon after emergence[90]. This means that many of the released adult males will not find a mating partner, and hence, larger releases will be required to exceed threshold frequencies than predicted in simple population frequency models.

The nature of mosquito dispersal behavior is also relevant to the spatial dispersal of transgenes. Our species of interest, *Ae. aegypti*, is understood to display leptokurtic dispersal behavior in a suburban setting, in which mosquitoes tend to remain in the same household

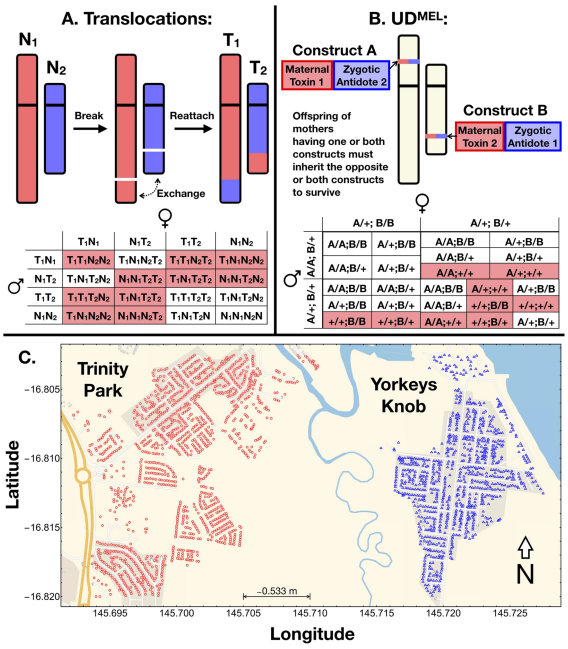

Figure 3.1: **Inheritance and Landscape Features of the Modeling Framework** (**A**) Reciprocal translocations (T_1 and T_2) result from the mutual exchange between terminal segments of two non-homologous chromosomes (N_1 and N_2). The cross here depicts possible parental gametes, with respect to the translocation, and offspring that result from matings between them. Matings between wild-type organisms or translocation homozygotes result in viable offspring, but translocation heterozygotes produce unbalanced gametes, and many of the resulting offspring are unviable (shaded). This results in a heterozygote disadvantage and threshold-dependent population dynamics. (**B**) UD^{MEL} is composed of two unlinked constructs (here referred to as A and B), each consisting of a maternally expressed toxin and a zygotically expressed antidote for the toxin on the opposite construct. The cross here represents matings between two of the nine possible parental genotypes ("+" represents the wild-type allele, and "A" and "B" represent alleles corresponding to the two UD^{MEL} constructs). The complete inheritance pattern is depicted in Additional file 2: Fig. S1. Offspring lacking the antidotes to the maternal toxins produced by their mother are unviable (shaded). At high population frequencies, the selective advantage on the constructs, by virtue of the antidotes, outweighs the fitness load due to the toxins and hence results in frequency-dependent spread. (**C**) Distribution of households in Yorkeys Knob (blue) and Trinity Park (red) in Cairns, Queensland, Australia. Households serve as individual *Aedes aegypti* populations in our metapopulation framework, with movement of adult *Ae. aegypti* between them. Yorkeys Knob serves as a simulated release site, and Trinity Park as a simulated control site.

for the majority of their lifespan, while a few mosquitoes disperse over larger distances[180]. With these landscape, dispersal, and life cycle considerations accounted for, we theoretically explore the ability to drive two threshold-dependent systems, translocations and UD^{MEL} , into populations of *Ae. aegypti* in one community, Yorkeys Knob, without them spreading in significant numbers to a neighboring community, Trinity Park, and to be remediated from Yorkeys Knob at the end of the simulated trial period.

3.1.2 Results

3.1.2.1 Model Framework

We use the Mosquito Gene Drive Explorer (MGDrivE) modeling framework[177] to model the spread of translocations and UD^{MEL} through spatially structured mosquito populations (Fig. 3.1). This is a genetic and spatial extension of the lumped age-class model of mosquito ecology[88] modified and applied by Deredec et al.[46] and Marshall et al.[142] to the spread of homing gene drive systems. The framework incorporates the egg, larval, pupal, and adult life stages, with egg genotypes being determined by maternal and paternal genotypes and the allelic inheritance pattern of the gene drive system. Spatial dynamics are accommodated through a metapopulation structure in which lumped age-class models run in parallel and migrants are exchanged between populations according to a zero-inflated exponential dispersal kernel with parameters defined in Additional file 1: Table S1[170, 162, 53, 184, 55, 96, 193, 179, 89, 74]. Further details of the framework are described in the Methods (section 3.1.5).

Applying the MGDrivE modeling framework to our research questions, we incorporate the inheritance patterns of reciprocal chromosomal translocations and UD^{MEL} into the inheritance module of the model (Fig. 3.1a, b, Additional file 2: Fig. S1), the life cycle parameters of *Aedes aegypti* (Additional file 1: Table S1) into the life history module, and the distribution of households in Yorkeys Knob (923 households) and Trinity Park (1301 households) along with their expected mosquito population sizes and movement rates between them into the landscape module (Fig. 3.1c). The suburb of Trinity Park served as a control site for field releases of *Wolbachia*-infected mosquitoes, to quantify the extent to which the *Wolbachia* infection could spread from one community to another, and plays a similar role for our simulated releases of threshold-dependent gene drive systems.

The inheritance patterns that result from chromosomal translocations are depicted in Fig. 3.1a. Chromosomal translocations result from the mutual exchange between terminal segments of two non-homologous chromosomes. When translocation heterozygotes mate, several crosses result in unbalanced genotypes and hence unviable offspring, resulting in a heterozygote reproductive disadvantage. This results in bistable, threshold-dependent population dynamics, confirmed in laboratory drive experiments[20]. The inheritance patterns produced by the UD^{MEL} system are depicted in Fig. 3.1b. This system consists of two unlinked

constructs, each possessing a maternally expressed toxin active during oogenesis and a zygotically active antidote expressed by the opposite construct. Offspring are more likely to have the antidote(s) to the maternal toxin(s) when transgenes are present at moderate-to-high population frequency. This produces threshold-dependent dynamics since, above a critical frequency, the selective advantage of the antidotes outweighs the selective disadvantage of the toxins, and below the critical frequency, the selective disadvantage of the toxins dominates. Mathematical models predict a threshold frequency of about 24% in the absence of an additional fitness cost-a result consistent with laboratory drive experiments for a construct having a modest additional fitness cost[4].

3.1.2.2 Model Validation

Using data from field trials of *Wolbachia*-infected *Ae. aegypti* mosquitoes in Yorkeys Knob and Gordonvale, Australia[95], we validated our modeling framework prior to application to other threshold-dependent drive systems. *Wolbachia* biases the offspring ratio in favor of those carrying *Wolbachia* through a mechanism known as cytoplasmic incompatibility, in which offspring of matings between infected males and uninfected females result in the death of some or all progeny, while matings involving infected females tend to produce infected offspring[195]. For the *wMel* strain of *Wolbachia* that was used in the Australian field trials, incompatible crosses produce no viable offspring, and *Wolbachia* is inherited by all offspring of infected females. There is also a fitness cost associated with *Wolbachia* infection, the value of which has been estimated between 0 and 20%[95, 172, 87].

We used *Wolbachia* surveillance data from Fig. 1 of Hoffmann et al.[95] to validate our model framework and parameter values. Based on the description of the field trials in Hoffmann et al.[95], we simulated weekly releases of 20 *Wolbachia*-infected mosquitoes (10 female and 10 male) per household at a coverage of 30% over 10 weeks in the spatially explicit landscapes of Yorkeys Knob and Gordonvale with the exception that in Gordonvale, the fifth release was postponed by a week due to a tropical cyclone. Model predictions were calculated for a variety of literature-based values of adult mortality rate and *Wolbachia*-associated fitness cost[95, 55, 172, 87, 149] and were compared to observed *Wolbachia*-prevalence over time. We found that model predictions most closely matched field data for a baseline adult mortality rate of 0.090 per day[55] and that predictions matched field data quite well for both 5% and 10% fitness costs, with a 10% fitness cost being closer to that estimated elsewhere[95, 172, 87] (Additional file 3: Fig. S2). Model predictions in Fig. 3.2 use these parameter values, along with others listed in Additional file 1: Table S1, and their agreement with the observed field data provides good validation of our modeling framework.

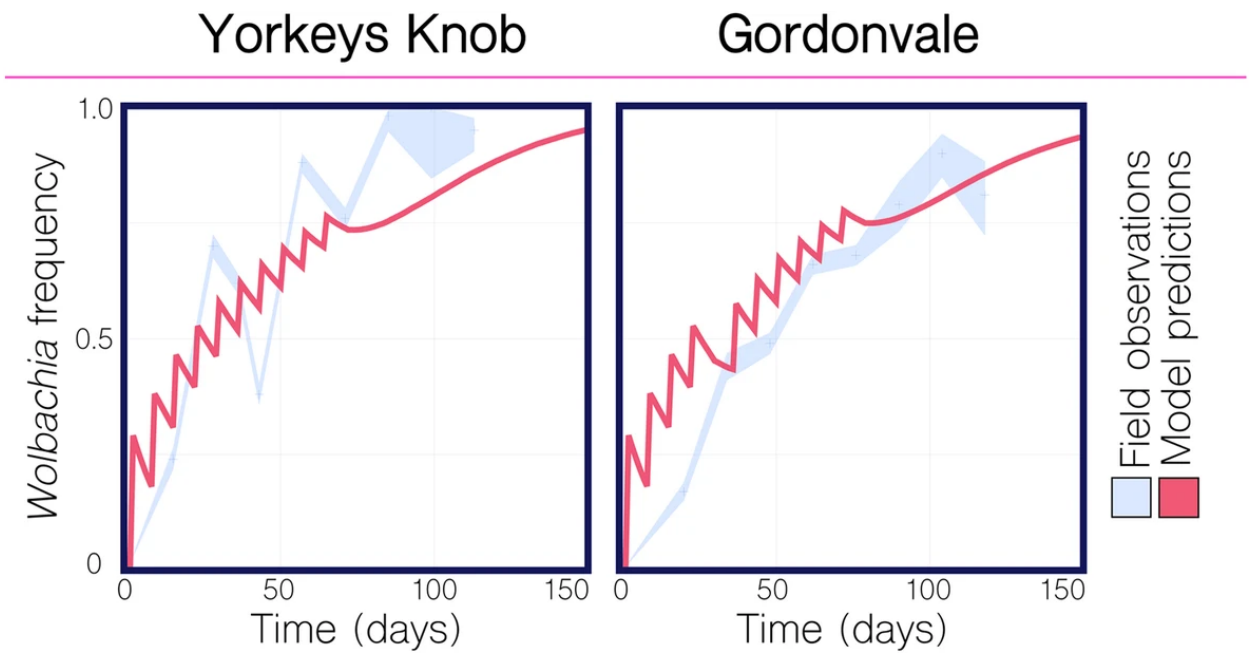

Figure 3.2: **Comparison of Model Predictions to *Wolbachia* Field Trial Data from Yorkeys Knob and Gordonvale, Australia**
Field observations of *Wolbachia* population frequency are depicted in light blue, with 95% binomial confidence intervals based on the frequency and sample size reported for the 2011 field trials in Yorkeys Knob and Gordonvale[95]. Model predictions are depicted for an analogous release scheme consisting of 20 *Wolbachia*-infected mosquitoes (10 female and 10 male) per household at a coverage of 30% over 10 weeks with the exception that in Gordonvale, the fifth release was postponed by a week due to a tropical cyclone. Parameter values listed in Additional file 1: Table S1. *Wolbachia* infection is associated with a 10% fitness cost. Agreement between observations and predictions is strong, providing good validation of the modeling framework.

3.1.2.3 Population Replacement and Remediation for Translocations

The use of translocations for transforming pest populations was initially suggested by Serebrovskii[181] and later Curtis[40] for the introduction of disease-refractory genes into mosquito populations. A number of models have been proposed to describe their spread through randomly mating populations[21, 140, 41, 206]; however, with one recent exception addressing spatial structure[32], these have largely ignored insect life history and mating structure. Such models suggest that the translocation need only exceed a population frequency of 50%, in the absence of a fitness cost associated with the translocation, to spread to fixation in a population, which could conceivably be achieved through a single seeding release round. Here, we find that incorporating life history and population structure into mosquito population dynamic models significantly increases release requirements.

In Figs. 3.3 and 3.4, based on the precedent set by the 2011 *Wolbcahia* field trial[95], we consider weekly releases of 20 adult *Ae. aegypti* males homozygous for the translocation per household for given durations and coverage levels, where coverage level is the proportion of households that receive the releases. Releases are simulated in the community of Yorkeys Knob, in which prior releases of *Wolbachia*-infected mosquitoes suggested a local population of approximately 15 adult *Ae. aegypti* per household[170], and for mosquito movement rates inferred from previous studies[180, 89, 74] (Additional file 1: Table S1). For a coverage level of 100%, and in the absence of a fitness cost, four weekly releases of 20 *Ae. aegypti* males ($\sim 3:1$ released to local males) are required for the translocation to spread to fixation in the community (Fig. 3.3), as opposed to the single release expected when ignoring life history and population structure[41]. As coverage is reduced to 50%, the required number of releases increases to 7, and for a coverage level of 25%, as seen for the World Mosquito Program in Yorkeys Knob, the required number of releases increases to 16 (Figs. 3.3 and 3.4). Although large, these releases are achievable, considering the much larger releases conducted for sterile insect programs[26].

To simulate remediation of a translocation, we consider weekly releases of 20 adult *Ae. aegypti* wild-type males in the community of Yorkeys Knob, whereby the translocation has already reached fixation in that community. In the absence of a fitness cost associated with the translocation, translocations are symmetrical in their threshold dynamics, and so, for a coverage level of 100%, four weekly releases are required for the translocation to be completely remediated from the community, and for a coverage of 25%, 16 weekly releases are required for the translocation to be completely remediated (Figs. 3.3 and 3.4). Encouraging features of these results are that (i) remediation can be achieved through releases of non-biting, non-disease-transmitting males; (ii) release sizes are achievable; and (iii) despite the spatial household structure, both replacement and remediation are complete within the community. The time to replacement is highly dependent on the coverage level and number of releases, but is reasonably quick given sufficient releases. At a coverage of 50%, 20 weekly releases led to the translocation spreading to a frequency $> 99\%$ within half a year of the final release

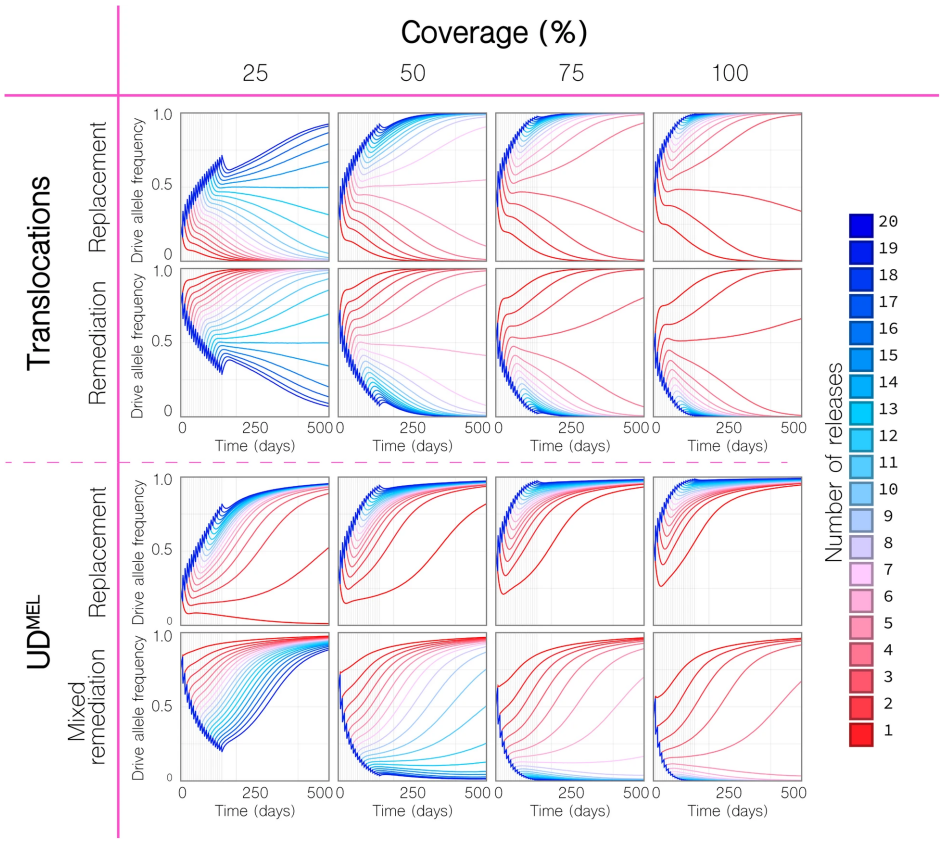

Figure 3.3: **Replacement and Remediation Results for Translocations and UD$^{\mathbf{MEL}}$** Time-series results are shown for a given number of weekly releases of 20 adult *Ae. aegypti* per household with the intent of population replacement or remediation in the community of Yorkeys Knob (Fig. 3.1c), and at given coverage levels, where coverage is the proportion of households that receive the releases. For population replacement, releases are of males homozygous for the translocation or UD$^{\mathrm{MEL}}$ into a wild-type population. For remediation of translocations, releases are of wild-type males into a population homozygous for the translocation. For UD$^{\mathrm{MEL}}$, remediation is not possible through releases of males only, and so "mixed remediation" is considered, in which releases consist of 10 females and 10 males. (**Top**) Replacement and remediation are symmetric for translocations. At a coverage of 50%, seven or more releases result in the translocation being driven to fixation or remediated from the population. (**Bottom**) Release requirements for UD$^{\mathrm{MEL}}$ are smaller for population replacement, but larger for mixed remediation. At a coverage of 50%, a single release results in UD$^{\mathrm{MEL}}$ being driven to fixation.

(or within 300 days of the first release). For equivalent wild-type releases, this is the same as the time to $> 99\%$ elimination.

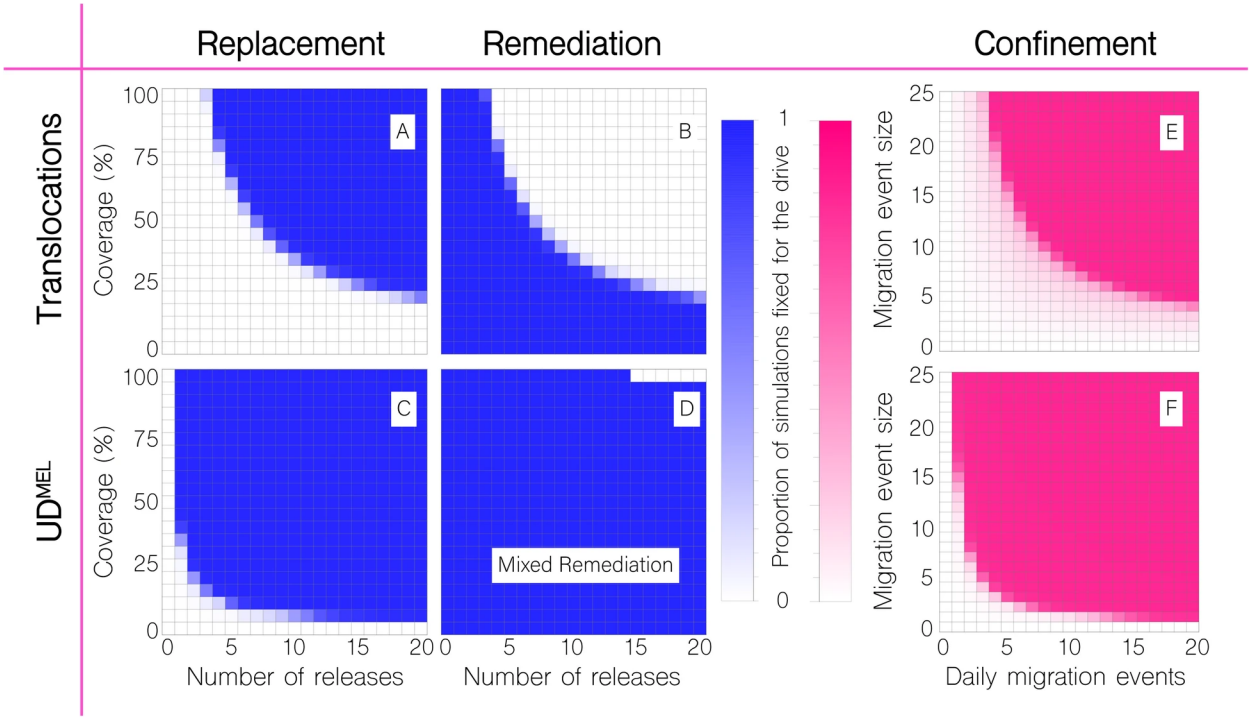

Figure 3.4: **Replacement, Remediation, and Confinement Outcomes for Translocations and UD^{MEL}**
Outcomes are depicted for the proportion of 50 stochastic simulations of population replacement, remediation, and confinement of translocations and UD^{MEL} that result in fixation of each system. (**A**-**D**) For replacement and remediation, each cell corresponds to a given number of releases (horizontal axis) and coverage level (vertical axis), given 20 adult *Ae. aegypti* per household per release. For replacement, releases are of males homozygous for the system into a wild-type population. For remediation of translocations, releases are of wild-type males into a population homozygous for the translocation, and for mixed remediation of UD^{MEL} , releases are of wild-type females and males into a population homozygous for UD^{MEL} . Blue cells represent cases where all simulations result in fixation of the system, and white cells represent cases where the wild-type is fixed in all simulations. (**E**, **F**) For confinement, each cell corresponds to a daily number of batch migration events (horizontal axis) of a given size (vertical axis) from Yorkeys Knob, where the system is fixed, to Trinity Park, where the system is initially absent. Dark pink cells represent cases where all simulations result in fixation of the system in Trinity Park, and white cells represent cases where the wild-type is fixed in all simulations. These results are encouraging for translocations as systems for introducing transgenes in a local and reversible way as (i) they can be remediated through an achievable number of male-only releases and (ii) they require more batch migration events to spread to neighboring communities.

3.1.2.4 Population Replacement and Remediation for UD^{MEL}

UD^{MEL} was the first synthetic gene drive system to be engineered that displays threshold-dependent dynamics[4]. The system consists of two unlinked constructs, each possessing a maternally expressed toxin active during oogenesis and a zygotically active antidote expressed by the opposite construct (Fig. 3.1b). At low population frequencies, the maternal toxin confers a significant selective disadvantage, leading to elimination, while at high population frequencies, the zygotic antidote confers a selective benefit in the context of a prevalent toxin, leading to fixation. The dynamics of this system in randomly mating populations have been characterized by Akbari et al.[4], suggesting that the system need only exceed a population frequency of $\sim 24\%$, in the absence of a fitness cost, to spread to fixation, while the wild-type must exceed a population frequency of $\sim 76\%$ to eliminate the construct. Both replacement and remediation should therefore be achievable with 1-2 releases of transgenic and wild-type organisms, respectively; however, as for translocations, we find that incorporating life history and population structure into our models increases release requirements in both cases.

In Figs. 3.3 and 3.4, we consider weekly releases of 20 adult *Ae. aegypti* males homozygous at both loci for the UD^{MEL} system in the community of Yorkeys Knob. The lower threshold for UD^{MEL} as compared to translocations means that replacement is much easier to achieve for UD^{MEL} . For a coverage level of 50% or higher, and in the absence of a fitness cost, a single release of 20 *Ae. aegypti* males leads to the UD^{MEL} system spreading to fixation throughout the community (Fig. 3.3). As coverage is reduced to 25%, the required number of releases to achieve fixation increases to two (Figs. 3.3 and 3.4). As for translocations, the time to replacement is highly dependent on the coverage level and number of releases. From Fig. 3.3, it is apparent that UD^{MEL} reaches total allele fixation slowly, although the number of individuals having at least one copy of the transgene increases quickly. At a coverage of 50%, 10 weekly releases lead to wild-type individuals falling to a frequency $< 2\%$ within 2.6 years of the final release (or within 3 years of the first release).

Remediation, however, is more difficult to achieve for UD^{MEL} compared to translocations due to the higher threshold that wild-type organisms must exceed to eliminate UD^{MEL} . Additionally, wild-type females must be included in the releases to propagate the wild-type allele because, assuming continued functioning of UD^{MEL} components, the maternal toxins of females having UD^{MEL} at both loci kill all offspring that do not inherit UD^{MEL} at both loci. To simulate remediation, we first consider weekly releases of 10 adult *Ae. aegypti* wild-type females and 10 adult males in the community of Yorkeys Knob. In the absence of a fitness cost associated with the UD^{MEL} construct, and for a coverage level of 75%, nine weekly releases are required for a reduction in UD^{MEL} allele frequency over the first year (Fig. 3.3); however, a closer inspection of the simulation results reveals that complete remediation of UD^{MEL} from the community is not possible even with 20 releases, as the UD^{MEL} allele frequency bounces back. Comparison of these results to those for a panmictic population with a population

size equal to that of Yorkeys Knob reveals that complete remediation of UD^{MEL} can occur at coverage levels as low as 25% (for 16 or more weekly releases) (Additional file 4: Fig. S3). Inspection of the spatially explicit simulation results suggests that the rebound in UD^{MEL} allele frequency in the structured population is due to UD^{MEL} remaining at super-threshold levels after the wild-type releases in a small number of households, and slowly recolonizing the landscape following that. Complete remediation of UD^{MEL} is possible, however, for 15 or more releases at a coverage level of 100% (Fig. 3.4). These results make a strong case for translocations as preferred systems to introduce transgenes in a local and reversible way as (i) remediation of UD^{MEL} requires releases of biting, vector-competent females and (ii) release requirements for these biting, vector-competent females are burdensomely high due to the high threshold that must be surpassed consistently throughout a spatially structured population.

3.1.2.5 Confinement of Translocations and UD^{MEL} to Release Site

Confinement of translocations and UD^{MEL} to partially isolated populations has previously been modeled by Marshall and Hay[140] and Akbari et al.[4]. In both cases, two randomly mating populations were modeled that exchange migrants at given rates. Population structure was otherwise ignored, as was mosquito life history. Results from these analyses suggest that translocations would spread and remain confined to populations for which the migration rate is less than approximately 5.8% per mosquito per generation[140], and that UD^{MEL} would remain confined to populations for which the migration rate is less than about 1.6% per mosquito per generation[4]. These migration rates are relatively low; however, this may be beneficial for the types of landscapes we are considering here, whereby the system may spread between neighboring households, but not from one suburb to another. Recently, Champer et al.[32] showed that translocations would remain confined to and persist in a population connected to another by a "migration corridor" under a range of parameter values.

For our landscape of interest-the suburbs of Yorkeys Knob and Trinity Park-it is very unlikely that *Ae. aegypti* mosquitoes will travel from one suburb to another by their own flight. Extrapolating the exponential dispersal kernel used in our simulations, fitted to data from mark-release-recapture experiments collated by Guerra et al.[74], suggests these events to be negligible, before accounting for the fact that the intervening vegetated area may serve as a barrier to *Ae. aegypti* flight[179]. Furthermore, rare migrant mosquitoes are unlikely to cause the threshold frequency for either drive system to be exceeded, thus making spatial spread due to such movements unlikely. In considering confineability to the release suburb, we therefore model "batch migration", in which several mosquitoes are carried, perhaps by a vehicle, from one community to another at once. Batch migration events could be thought of as several adult mosquitoes being carried at once, or perhaps more likely, as a larval breeding site, such as a tyre, being carried from one household to another, with several adults emerging from the tyre following transport. We model batch migration events as occurring

between randomly chosen households, and vary the number of daily migration events and the effective number of adults carried per event. For computational simplicity, we focus on migration events from Yorkeys Knob, in which either system has already reached fixation, to households in Trinity Park, which is initially fixed for wild-type mosquitoes.

In Fig. 3.4e, f, we see that both the number and size of daily batch migration events affect the chance of either system establishing itself in the neighboring suburb, Trinity Park. For translocations, about 16 daily migration events of batches of 5 adults are required for spread in Trinity Park. For batches of 10 adults, approximately 9 daily migration events are required, and for batches of 20 adults, only 5 daily migration events are required. For UD^{MEL}, only 3 daily migration events of batches of 5 adults are required for spread in Trinity Park, and for batches of 10 adults, around 2 daily migration events are required.

These results continue to make a strong case for translocations as preferred systems to introduce transgenes in a local and reversible way as (i) many more batch migration events are required to lead to spread for translocations as opposed to UD^{MEL} and (ii) the rate of migration events required for translocations to spread is higher than what would be expected between these communities. Specifically, *Wolbachia* releases in Yorkeys Knob in 2011 provide evidence for occasional batch migrations to the nearby suburb of Holloways Beach; however, the spatio-temporal pattern of *Wolbachia* spread, as inferred from monitored trap data, suggests only $\sim 1-2$ batch migration events over the course of a month, consisting of less than 5 adult females per event[95].

3.1.2.6 Sensitivity Analysis

A theoretical study by Khamis et al.[111] on toxin-antidote-based underdominance gene drive systems, similar to UD^{MEL} but for which the toxins are zygotic rather than maternal[43], found that the gene drive threshold frequency is highly sensitive to (i) the increase in adult mortality rate in organisms having the transgene, (ii) the duration of the larval life stage, and (iii) the parameters determining the character or strength of larval density dependence. In Fig. 3.5 and Additional file 5: Fig. S4, we explore the sensitivity of our model outcomes of replacement, remediation, and confinement for translocations and UD^{MEL} as we vary (i) the duration of the larval life stage, (ii) the baseline adult mortality rate, (iii) the fitness cost associated with the gene drive system, and (iv) the mean adult dispersal distance. For translocations, we model a 10% fitness cost as a 10% reduction in mean adult lifespan for organisms homozygous for the translocation and a 5% reduction for organisms heterozygous for the translocation. For UD^{MEL}, since its inheritance bias is induced through the action of maternal toxins, we model a 10% fitness cost as a 10% reduction in female fecundity for organisms homozygous for UD^{MEL} at both loci, with 2.5% additive fitness costs contributed by each transgenic allele.

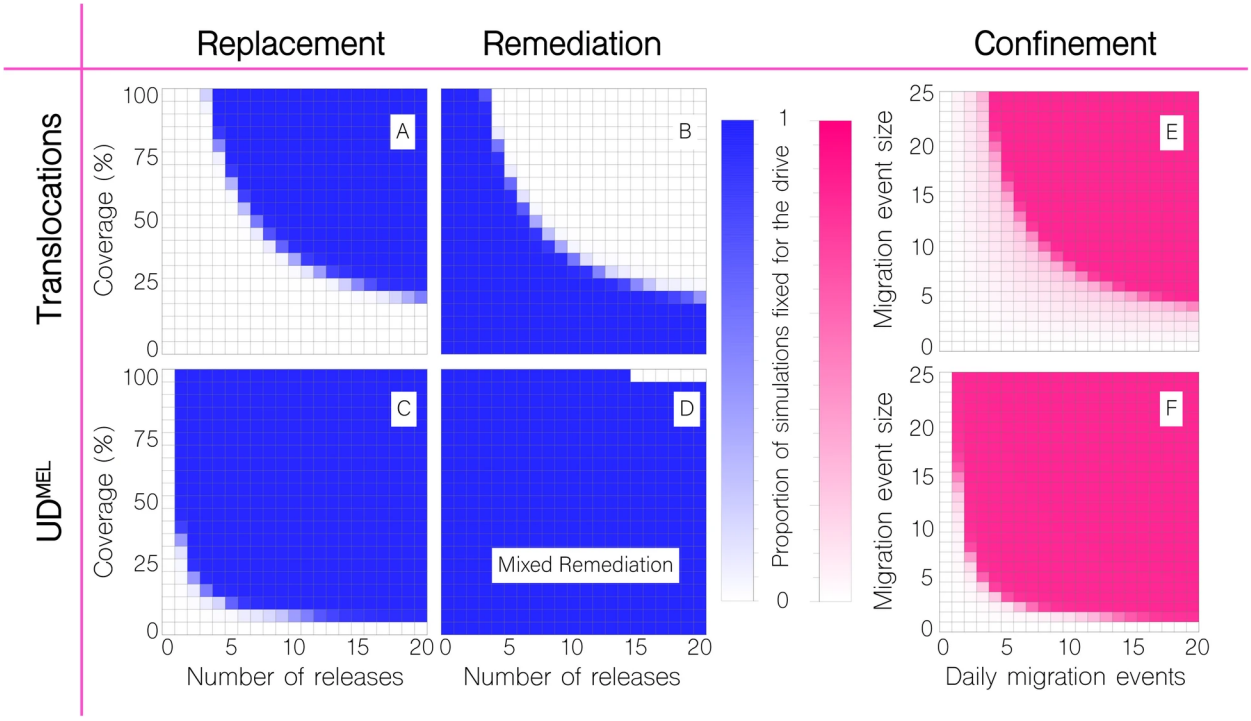

Figure 3.5: **Sensitivity of Model Outcomes for Replacement, Remediation, and Confinement of translocations**
Changes are depicted in the proportion of 50 stochastic simulations that result in fixation for replacement, remediation, and confinement of translocations. Proportions are compared to those in the first row of Fig. 3.4 as we vary (i) the mean dispersal distance of adult mosquitoes (±50%), (ii) the duration of the larval life stage (±2 days), (iii) the baseline adult mortality rate (±2%), and (iv) the fitness cost associated with being homozygous for the translocation (+10% or +20%). Fitness costs have the greatest impact on the release scheme required for the system to be fixed or remediated from the population, given the life parameters considered. Fitness costs also lead to more batch migration events being required for invasion of Trinity Park. A small increase in the baseline adult mortality rate leads to slightly fewer batch migration events being required for invasion of Trinity Park; however, comparison to migration rates inferred from field data suggests that confinement is still expected.

For translocations, the associated fitness cost had the greatest impact on the release scheme required for the system to be fixed or remediated from the population, given the life parameters considered (Fig. 3.5). A 10% fitness cost led to around 10 weekly releases at a coverage of 50% being required for the translocation to reach fixation (an increase of 3 releases), while a 20% fitness cost led to about 13 weekly releases being required (an increase of 6 releases).

Small changes in the duration of the larval life stage had minor impacts on the release requirements, with an increase in larval lifespan of 2 days leading to one more weekly release being required for the translocation to reach fixation, and vice versa. A 2% change in the baseline adult mortality rate and 50% change in the mean migration distance had negligible impact on release requirements. Remediation, on the other hand, requires fewer wild-type releases when there is a fitness cost associated with the translocation. A 10% fitness cost led to 5 weekly releases at a coverage of 50% being sufficient to eliminate the translocation (a decrease of 2 releases), and a 20% fitness cost led to 4 weekly releases at a coverage of 50% being sufficient for elimination (a decrease of 3 releases). Small changes in the duration of the larval life stage had minor impacts on the wild-type release requirements for elimination, with an increase in larval lifespan of 2 days or a 2% decrease in the adult mortality rate leading to one fewer release being required.

The sensitivity of our predictions regarding confinement to the release site is of particular interest, as invasion of a neighboring community may be more likely under some parameter values than others. Fortunately, a fitness cost associated with the translocation leads to a higher threshold and hence more batch migration events required for invasion of a neighboring community. A 10% fitness cost led to $\sim 2-3$ additional daily migration events of 10 adults required for spread to Trinity Park, and a 20% fitness cost led to $\sim 6-7$ additional daily migration events required (Fig. 3.5). Also noteworthy, a 2% increase in the adult mortality rate led to around 2 fewer daily migration events required for spread to Trinity Park-i.e., ~ 7 migration events for batches of 10 adults and ~ 14 migration events for batches of 5 adults. While still above inferred batch migration rates, this highlights that there could exist parameter sets beyond those explored for which invasion is feasible.

UD^{MEL} displays similar parameter sensitivities regarding fixation and batch migration outcomes as for translocations, with the exception that these outcomes are less sensitive to fitness costs (Additional file 5: Fig. S4), likely due to the fact that fitness is accommodated through a reduction in female fecundity rather than an increase in adult mortality. A 20% fitness cost led to almost 1 additional weekly release being required for the system to spread to fixation, whether at a coverage of 25% or 50%. Similarly, for an invasion of Trinity Park, a 10% fitness cost required nearly 1 additional daily migration event of 5 adults, and a 20% fitness cost required ~ 2 additional daily migration events. Of note, a 2% increase in the adult mortality rate or a 2-day increase in the duration of the larval stage led to almost 1 fewer daily migration event required for spread to Trinity Park, making this now very achievable-i.e., $\sim 2-3$ migration events for batches of 5 adults and $\sim 1-2$ migration events for batches of 10 adults.

Finally, we conducted an analysis of the sensitivity of our results to population structure, exploring the impact of (i) removing all population structure by treating Yorkeys Knob and Trinity Park as randomly mixing populations and (ii) incorporating heterogeneity in mosquito household population size. Results of the comparison to panmictic populations are

depicted in Additional file 6: Fig. S5. Previously, we had seen that introducing population structure greatly increases the release requirements to eliminate UD^{MEL} from a community (Additional file 4: Fig. S3). The trend of higher release requirements in structured populations is also seen for translocations, although to a lesser extent, with one additional release required for either replacement or remediation at a coverage of 75%, two additional releases required at a coverage of 50%, and 7-8 additional releases required at a coverage of 25%. Invasion of a neighboring population, on the other hand, requires around 1-2 fewer daily migration events in structured populations for batches of 10 adults having either the translocation or UD^{MEL} .

Incorporating heterogeneity in household mosquito population size, we retain a mean of 15 adults per household, as inferred from *Wolbachia* field trial data in Yorkeys Knob[170], and distribute population sizes across households according to a zero-inflated, truncated exponential distribution with 55% of households having no mosquitoes and none having more than 45 adults. This distributional form, including zero inflation, is based on results of a large field survey conducted across a set of households in Kamphaeng Phet province, Thailand[112]. Including this source of heterogeneity substantially increases release requirements for replacement and remediation with translocations, with 3-4 additional releases required at a coverage of 75%, 5-6 additional releases required at a coverage of 50%, and 20 releases being insufficient at a coverage of 25% (Additional file 7: Fig. S6). Release requirements are marginally increased for UD^{MEL} , with 1-2 additional releases required at coverages of 25–100%. These results are likely due to the threshold frequency being more difficult to exceed in households with large numbers of mosquitoes, and this being less of an issue for UD^{MEL} due to its lower threshold frequency. Fortunately, population size heterogeneity makes confinement more promising for both systems, increasing the required number of daily migration events for batches of 10 adults by 3-4 for translocations and by 1-2 for UD^{MEL}.

3.1.3 Discussion

The idea of using threshold-dependent gene drive systems to replace local populations of disease vectors with varieties that are unable to transmit diseases has been discussed for over half a century now, since Curtis[40] famously proposed the use of translocations to control diseases transmitted by mosquitoes. While Curtis had been primarily concerned with introducing and spreading genes into a population, as the technology nears implementation, confining them is also becoming a significant concern. As CRISPR-based homing gene drive technology edges closer to field application, concerns are increasingly being raised regarding the invasiveness of these systems[138, 154], and systems such as split drive[3], daisy drive[155], and threshold-dependent underdominant systems[117] are gaining interest, at least during the trial phase of population replacement technology[139]. In this paper, we model the introduction of two drive systems, chromosomal translocations and UD^{MEL}, that have been engineered in the laboratory and shown to display threshold-dependent spread[21, 4]. While previous papers have described the population dynamics of these two systems in

randomly mating populations ignoring life history[21, 4, 140, 40], with one recent paper including spatial structure[32], we present the first analysis of these systems in a spatially structured population including mosquito life history and reflecting a well-characterized landscape where field trials could conceivably be conducted[95].

Our results provide strong support for the use of translocations to implement confineable and reversible population replacement in structured *Ae. aegypti* populations. Regarding reversibility, translocations are preferable to UD^{MEL} as (i) they can be remediated through releases of non-disease-transmitting male *Ae. aegypti* and (ii) required releases sizes are achievable (∼10 weekly releases at a coverage level of 50%). UD^{MEL} requires less effort to introduce into a population, but is much more difficult to remove once it has been introduced, requiring a very large number of both males and disease-transmitting females to be released. This highlights the benefit of a 50% threshold for reversible population replacement: the symmetry allows both replacement and remediation to be achieved with similar effort. Extreme underdominance is another example of system with a 50% threshold[43, 144]. Regarding confineability, translocations again outperform UD^{MEL} as around 16 daily migration events of batches of 5 *Ae. aegypti* adults are required for translocations to spread to the neighboring suburb of Trinity Park, while UD^{MEL} can spread to Trinity Park given only 3 daily migration events (or 2-3 daily migration events for alternative model parameterizations). The true batch migration rate between suburbs is expected to be smaller than either of these[95]; however, the rate required for translocations to spread is highly unlikely to be reached, while the rate for UD^{MEL} is conceivable.

As with any modeling study, there are limitations inherent in our analysis. Several of the parameters we assumed to be constant here would indeed be dynamic in a real intervention scenario. At the genetic level, lab experiments suggest non-outbred individuals homozygous for the translocation had a fitness cost that largely disappeared once offspring were produced that were the product of at least one wild-type individual[21]. Models fitted to data from UD^{MEL} drive experiments also suggested dynamic fitness costs that depended on the frequency of transgenic organisms in the population[4]. Another recent modeling study highlights the possibility of toxin and antidote mutational breakdown for underdominance constructs; however, this is expected to be impactful over a larger timescale than considered here (hundreds of generations)[52]. At the ecological level, our model of *Ae. aegypti* life history[177], based on the lumped age-class model of Hancock and Godfray[88], assumes the existence of a constant equilibrium population size and other constant ecological parameters, such as adult death rate and larval development times. These parameters have indeed been shown to vary in space and time, and in response to local mosquito density[86, 173], which our sensitivity analyses suggest could have significant impacts on release thresholds and gene drive outcomes[111] (Fig. 3.5 and Additional file 5: Fig. S4). At the landscape level, we have assumed a relatively homogenous distribution of mosquitoes per household, and movement rates between households that depend only on distance and household distribution. Extensive landscape heterogeneities have been shown to slow and alter the spread of *Wolbachia*[87,

84] and would likely impact the spread of translocations and UD^{MEL} as well. Future work that helps to characterize the environmental drivers of mosquito population dynamics will inform iterative model development to address this.

3.1.4 Conclusions

In conclusion, our analysis supports the use of translocations as a threshold-dependent drive system capable of spreading disease-refractory genes into structured *Ae. aegypti* populations in a confineable and reversible manner. If such a system were engineered in *Ae. aegypti*, it would be an excellent candidate for the introduction of disease-refractory genes during the trial phase of population replacement technology, or whenever a localized release was otherwise desired. As the technology nears implementation, further ecological work characterizing the density dependencies, seasonality, and spatial heterogeneities of *Ae. aegypti* populations will be essential to enhance model predictions in preparation for field trials.

3.1.5 Methods

We used the MGDrivE framework[177] (https://marshalllab.github.io/MGDrivE/) to simulate releases of adult *Ae. aegypti* males homozygous for one of two threshold-dependent gene drive systems-reciprocal chromosomal translocations or UD^{MEL} -in the community of Yorkeys Knob in Queensland, Australia. To simulate remediation, we modeled releases of wild-type adult *Ae. aegypti* into populations in Yorkeys Knob already fixed for the gene drive system. Houses receiving releases were randomly chosen from a uniform distribution. These were conserved within simulation runs, but varied between runs. To determine confineability, we simulated batch migration events from Yorkeys Knob (fixed for the gene drive system) to the neighboring community of Trinity Park (initially wild-type). The MGDrivE framework models the egg, larval, pupal, and adult (male and female) mosquito life stages implementing a daily time step, overlapping generations and a mating structure in which adult males mate throughout their lifetime, while adult females mate once upon emergence, retaining the genetic material of the adult male with whom they mate for the duration of their adult lifespan. Density-independent mortality rates for the juvenile life stages are assumed to be identical and are chosen for consistency with the population growth rate in the absence of density-dependent mortality. Additional density-dependent mortality occurs at the larval stage, the form of which is taken from Deredec et al.[46]. Full details of the modeling framework are available in the S1 Text of Sánchez et al.[177] and in the software documentation available at https://marshalllab.github.io/MGDrivE/docs/reference/. Parameters describing *Ae. aegypti* life history and the gene drive systems and landscape of interest are listed in Additional file 1: Table S1.

The inheritance patterns for reciprocal chromosomal translocations (depicted in Fig. 3.1a) and UD^{MEL} (depicted in Additional file 2: Fig. S1) are modeled within the inheritance module of the MGDrivE framework[177], and their impacts on female fecundity and adult lifespan

are implemented in the life history module. The distribution of households in Yorkeys Knob, Trinity Park, and Gordonvale was taken from OpenStreetMap (Fig. 3.1c). We implement the stochastic version of the MGDrivE framework to capture the randomness associated with events that occur in small populations, such as households, which serve as nodes in the landscape modeled here. In the stochastic implementation of the model, the number of eggs produced per day by females follows a Poisson distribution, the number of eggs having each genotype follows a multinomial distribution, all survival/death events follow a Bernoulli distribution, and female mate choice follows a multinomial distribution with probabilities given by the relative frequency of each adult male genotype in the population.

3.1.6 Availability of Data and Materials

The distribution of households in Yorkeys Knob, Trinity Park, and Gordonvale was obtained from OpenStreetMap. The software package, MGDrivE, is available on GitHub[177]. All information required to reproduce the simulations is available in the GitHub deposition, the manuscript, and its additional files.

3.2 Trans-Complementing Gene Drive Removes Cas9 after Population Modification: A Modeling Analysis

3.2.1 Introduction

Mosquito-borne diseases cause significant health and social burdens throughout much of the world[16, 161]. Long-standing approaches to mosquito-borne disease control, including insecticide-treated nets and indoor residual spraying with insecticides, are becoming less effective as resistance to insecticides spreads[92, 38]. Novel and existing methods involving population suppression through releases of sterile insects are promising for local disease control[26, 39, 124]; however, these approaches are less feasible at larger spatial scales, as they require large, sustained releases and disease vector populations are likely to rebound once efforts are relaxed.

Gene drive technology provides scalable options for relatively fast and enduring control of mosquito populations[62, 114], provided the necessary community and regulatory approvals can be gained. In general, two approaches for malaria vector control have been proposed: i) population modification, in which a disease-refractory gene is driven into the mosquito population, preventing pathogen transmission between mosquitoes and humans, and ii) population suppression, in which a gene conferring a fitness load or sex bias is driven into the population, potentially eliminating the mosquito population.

A drawback of standard CRISPR-based population modification strategies is that a Cas9 nuclease gene is driven into the population along with the effector gene[2, 24], and that enduring Cas9 nuclease activity could result in ongoing off-target mutagenic effects[102, 163]. Here, we describe how a recently-proposed trans-complementing gene drive (tGD) system[129] could provide a solution to this by retaining the effector gene in the population while removing the engineered Cas9 allele. This strategy results in a Cas9-free modified population where only the gRNAs and effector genes remain in the insect population.

The tGD system was originally proposed as a CRISPR-based gene drive design that addresses laboratory biosafety concerns by separating the drive components - Cas9 and guide RNAs (gRNAs) - in two separate transgenic individuals, and studying their properties in these different strains (Fig. 3.6b). In the tGD design, the gRNAs target wild-type alleles of genes at two different loci, inducing gene drive-based allelic conversion at both loci only when the separate engineered individuals are crossed together, limiting gene drive activity to experimentation windows[129]. Mathematical modeling has shown the tGD design could also be advantageous for population modification programs as, if effector genes are linked to both the Cas9 and gRNA components, the proportion of individuals having at least one effector gene may increase more quickly and to a higher population frequency than for an

equivalent single-locus population modification system[129].

Here, we explore another advantage of the tGD design - that the separation of the Cas9 and gRNA components allows the Cas9 allele to be reduced or eliminated from a modified population while the gRNA allele and linked effector gene remain at high frequency. A result of the separation of tGD components is that, as the inheritance bias induced by the homing process wanes due to depletion of cleavable wild-type alleles, fitness costs dominate and one or both of the tGD components may be selected out of a population. Fitness costs due to Cas9 have been documented in the presence of gRNAs, and may be leveraged for these purposes. Fitness costs are also commonly associated with drive systems and effector genes relative to wild-type alleles[62]; however these can be minimized by including a recoded copy of the target wild-type allele within the drive system[2, 104], preventing the gRNA and effector gene from being selected out of the population.

Using computational modeling, we explore the performance of a tGD system with an effector gene, carrying a recoded DNA sequence of the targeted allele at the gRNA locus, at introducing the effector gene for an extended period and Cas9 allele transiently for a range of conditions. We compare this tGD design to two other gene drive designs that could also be used to achieve this - a split gene drive (sGD) in which the Cas9 allele is not driven[123] (Fig. 3.6c), and a two-locus cleave-and-rescue (ClvR) system that cleaves an essential gene and restores its function[159] (Fig. 3.6d). A single-locus gene drive (GD)[62] (Fig. 3.6a) is also modeled for comparison. Of these gene drive architectures, we show that tGD outperforms the other designs in terms of removing the Cas9 allele while spreading and retaining the effector gene at high frequency in the population.

3.2.2 Results

We explored the effectiveness of tGD under several fitness cost and release scenarios. The behavior of a GD depends on its design: the choice of promotor, Cas9, gRNAs, effector gene, target loci, and more. These choices impact the cutting/HDR/NHEJ rates and the fitness costs associated with each piece of the construct, thereby influencing the effectiveness of the drive. To responsibly account for this, we applied fitness costs across several different aspects (Fig. 3.6f, g) of our design and the corresponding non-drive split-gene design. Cas9 was provided a copy-number dependent cost, due to protein expression and off-target cutting, and non-functional NHEJ alleles at that locus were applied a recessive cost, due to gene disruption. It is assumed that the effector gene is designed to be as cost-independent as possible, to increase the potential for success, and targets an essential gene, so non-functional alleles at the target site are recessive non-viable. For the traditional GD, with linked Cas9/gRNAs, the costs are combined and applied and apply at the same locus, analogous to the split-gene drive designs (Fig. 3.6e). ClvR was implemented similar to[160], where fitness costs were applied additively to the Cas9 locus (Fig. 3.6h), the effector/gRNA locus has a negligible

fitness cost, and the target locus is recessive lethal.

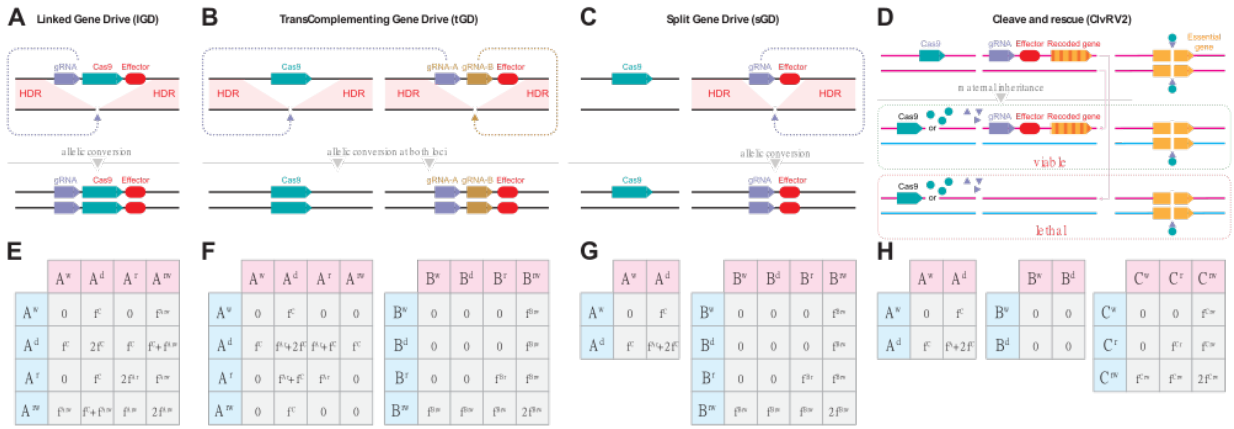

Figure 3.6: **Genetic Design and Fitness Cost Implementation**
Molecular designs (**A** - **D**) and fitness cost implementations (**E** - **H**) for linked drive (**A** and **E**), transComplementing drive (**B** and **F**), split drive (**C** and **G**), and a two-locus Cleave and Rescue (**D** and **H**).
(**A**-**D**) Molecular designs show where the pieces of each drive are located, on one or several autosomes, and which pieces undergo allelic conversion. Arrows from the gRNAs indicate target sites. Grey arrows indicate allelic conversion. The parental chromosomes are indicated on top and chromosomes after proper allelic conversion are indicated beneath them.
(**E**-**H**) Independent chromosomes are labeled "A", "B", or "C". The impact of allelic conversion is indicated as super scripts: X^w is wild-type, X^d is the drive construct, X^r represents non-functional resistance alleles, and X^{rw} represents functional resistance alleles. The fitness cost ("f") of each allele is indicated with the same superscripts, with the additional label for the cost associated with Cas9 ("f^c"). For example, cells with "$2f^c$" (in **E**) indicates 2 times the fitness cost of Cas9.

Figure 3.7a demonstrates two archetypal results for release of the tGD construct into a panmictic population, with low (10%, Fig. 3.7a top) or high (90%, Fig. 3.7a bottom) fitness costs applied to the Cas9 allele. In both instances, the effector gene spreads through the population, with near-complete coverage in carriers, in under 2.5 years. For a dominant-acting effector gene, where a single copy provides protection, this is exactly the outcome desired. Concurrently, the Cas9 allele drives into the population and subsequently falls out of the population over the 5 year simulation period. In the upper panel of figure 3.7a, where the Cas9 allele has a 10% fitness reduction, the Cas9 allele is driven out of the population through the accumulation of functional (ergo, low-fitness cost) NHEJ alleles. In the lower panel of figure 3.7a, where the Cas9 allele possesses a 90% fitness cost, the Cas9 allele falls out of the population prior to accumulation of NHEJ alleles, leaving the population only

partially changed. This demonstrates an ability to remove Cas9 alleles from the population, thereby removing a constitutive source of mutagenesis and preventing desensitization of populations to future gene drives.

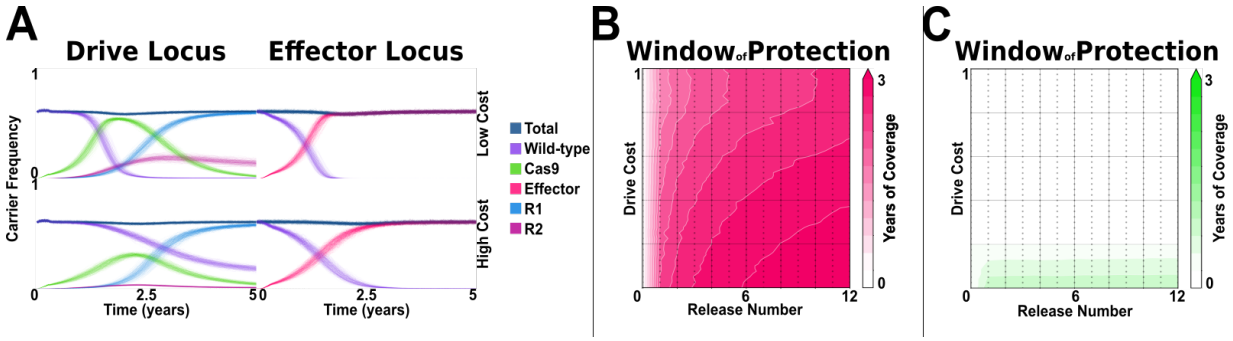

Figure 3.7: **Expected Behavior of transComplementing Gene Drive in a Panmictic Population**
(**A**) Two examples of tGD, one with a 10% fitness cost on Cas9 ("Low Cost", top) and one with a 90% fitness cost ("High Cost", bottom), released into a panmictic population of 10,000 individuals. 12 weekly releases of male mosquitoes, at 1% of the total population size, were performed. The dynamics of the Cas9 locus are shown on the left ("Drive Locus"), as measured in the total population, and the gRNA/effector gene are shown on the right ("Effector Locus"), as measured in the female population only, since female mosquitoes transmit disease.
(**B**) Sweep over the Cas9 cost ("f^c" in Fig. 3.6f) and the number of weekly releases, from 0 to 12. All other parameters are as in panel A. The sweep measures the amount of time the effector gene remains above 75% carrier frequency, individuals homozygous or heterozygous for the allele, in the female portion of the population over the course of the simulation.
(**C**) Same as panel **B**, but measuring the frequency of Cas9 carriers above10% in the whole population.
Supplemental Table A.1 contains life-cycle parameters for all simulations and Supplemental Table A.2 contains drive parameters.

Given a potential design, implementation of our construct is subject to economic, social, and operational constraints. Therefore, we explore the efficacy of our design as a function of release sizes (as a fraction of the total endemic population) or release number (Fig. 3.7b and c, Sup. Fig. A.1). We explore fitness costs of the Cas9 allele, ranging from no cost to complete cost, against up to 3 months, approximately one field season, of weekly releases. The tGD system is most effective at low cost and high release number, as expected, but drives the effector carrier frequency above 50% in the population for all parameter combinations tested (Fig. 3.7b). Additionally, the speed of introgression remains $> 80\%$ of the optimal time for

all parameters, indicating that the rate of tGD is relatively insensitive to fitness costs of the Cas9 allele (Sup. Fig. A.1a).

The behavior of Cas9 in the population is critical to reducing population mutagenesis and preventing horizontal transfer and spread into naive populations. Figure 3.7a demonstrates rapid rise and subsequent elimination of the Cas9 allele under low or high fitness costs. We measured how long Cas9 remained above 10% in the population, over the entire parameter range (Fig. 3.7c), and found that it spends less than 20 months above 10% under the most lenient parameterization, and under some conditions never rose above about 70% (Sup. Fig. A.1b). These results indicate potential to reduce fitness reduction in the target population and unintended spread through hybridization with non-target species, while indicating directing designs towards a sweet-spot that maximizes introgression of our effector gene and egression of the Cas9.

Confineability, controlling or preventing the spread of a GD from one population to another, is an issue that has arisen in other contexts due to concerns of drives spreading over political or territorial borders. To compare confinement and efficacy of our design, we performed multi-deme simulations against comparable designs, the linked GD and the split GD, as well as a recent threshold system, ClvR[160] (Fig 3.8). Simulations involved 2 demes connected by a 1% per generation migration rate, where releases were performed, and efficacy measured, in the first deme, and confinement was measured in the second.

As before, we query drives for introgression of an effector gene and spread or removal of Cas9. Linked and trans-complimenting GDs rapidly drive the effector gene through not only the release population but also the non-release, linked population (Fig. 3.8b). In contrast, split GD and ClvR introgress an effector gene above 90% only in a subset of parameter space, but have the bonus of being confineable to the release population over the majority of parameter space. To obtain this introgression, split GD and ClvR required more than 6 releases 40% the size of the expected population size, while linked and trans-complimenting GDs achieved this success using only 1 release at 1% of the expected population size, a feature of the threshold-less designs.

Behavior of the Cas9 locus indicates that trans-complimenting GD resembles split GD and ClvR instead of linked GD (Fig. 3.8c). Given the size and amount of releases we perform, and because split GD and ClvR do not drive the Cas9 locus, these drives never cross a 10% introgression frequency. tGD does cross that threshold under some parameter combinations, but the Cas9 allele falls out of the population in a fraction of our 5 year period. Conversely, it appears the linked drive never removes the Cas9 allele from the population. This is partially true at lower fitness costs, it replaces the entire population and the fitness of that population, as a whole, is reduced, while at higher fitness costs it causes population suppression, thereby removing the population as it removes the Cas9 allele. It also has the unfortunate side-effect of removing the effector gene with the Cas9, a consequence of the linked design. This

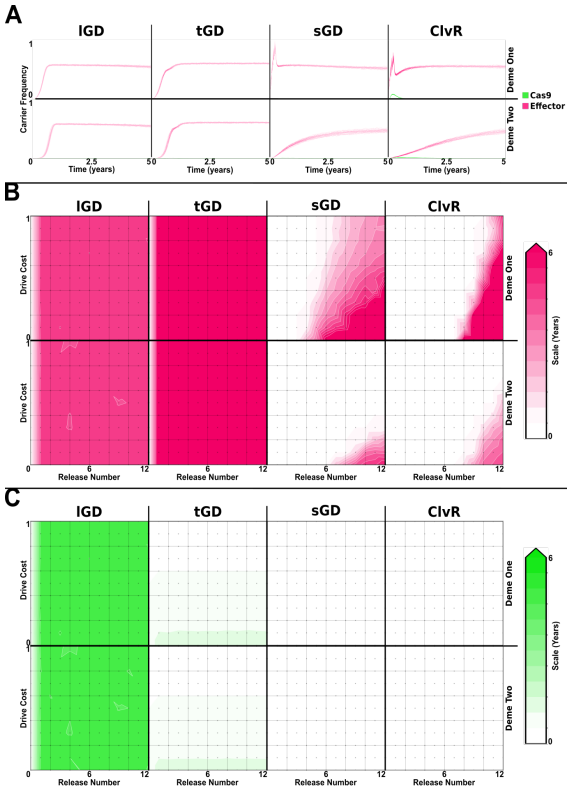

Figure 3.8: **Gene Drive Technologies in a Two-Population Landscape**
A comparison between the four competing technologies described in figure 3.6.
(**A**) 12 weekly releases of male mosquitoes were performed in "Deme One" (top), which is connected to "Deme Two" (bottom) by a 1% migration rate per adult generation. Linked and transComplementing drives were released at 1% of the total population size, while split drive and ClvR were released at 40% of the total population size. All drives had a 90% cost on the Cas9 allele. The green Cas9 line is covered in the linked drive panels ("lGD", left column) due to the linkage between Cas9 and effector genes.
(**B**) Heatmaps exploring the length of time the effector gene spent at over 75% carrier frequency, individuals homozygous or heterozygous for the allele, in the female portion of the populations.
(**C**) Heatmaps exploring the length of time Cas9 remained above 10% carrier frequency in the whole population, male and female.
Supplemental Table A.1 contains life-cycle parameters for all simulations and Supplemental Table A.2 contains drive parameters.

situates tGD at an interesting crossroad - being as fast and effective, but also as invasive, as typical linked designs (Sup. Fig. A.2a), while providing the benefit of removing Cas9, and preventing further mutagenesis, as typical split and ClvR designs.

3.2.3 Discussion

We have explored the application of a trans-complimenting gene drive over varying parameters, representing uncertainties possible in field applications, and compared it to three contemporary drive systems - basic linked homing, split homing, and two-locus Cleave-and-Rescue - for several metrics of interest. The tGD system demonstrates strong performance in terms of introgression fraction and speed, leading to a rapid impact from effector genes, as well as natural loss of the Cas9 element for the prevention of future off-target mutagenic effects.

However, there are caveats with the performance of a tGD design. As the Cas9 element links with the gRNAs, it spreads rapidly, significantly out-performing split and ClvR style gene drives. The ability to separate and only lose the Cas9 portion reduces fitness effects at high levels of introgression, demonstrating higher levels of introgression than the linked drive design. These traits cause tGD designs to be highly invasive, equal to or even greater than linked-drive designs. This makes it ideal for use in locations that are separated from other populations, or where such global spread would not be a negative result. In instances where control or containment is desired, split and ClvR style drives demonstrate superior performance.

Continued action by Cas9 in a population poses the potential for unwanted spread, via hybridization between closely-related species[194], mutagenic effects, from off-target action or mutagenesis of the Cas9 itself, and issues for future application of gene drives, because the population will either maintain the current Cas9 or evolve ways of silencing Cas9 action[57]. Therefore, removal of the Cas9 element is of critical importance, and three designs explored here accomplish that goal. However, there are differences in the surviving population. Split and ClvR designs do not propagate their Cas9 elements in a Super-Mendelian fashion, so there is no damage to the sister chromosomes at that location. This is why these designs are relatively slow to spread, compared to tGD or linked-drive designs, and why the population remains wild-type at that locus after the Cas9 element has died out. In contrast, tGD actively propagates its Cas9 element, demonstrating superior invasion dynamics, but also generating NHEJ/MMEJ alleles at that locus. After the Cas9 element has fallen out of the population, the remaining alleles are primarily, if not completely, functional NHEJ alleles. Thus, while the Cas9 is removed from the population, the population is not left unchanged as illustrated by split and ClvR designs (Sup. Fig. A.1b and Sup. Fig. A.2b).

As with any modeling study, there are limitations inherent to our analysis. While we vary several parameters to account for environmental unknowns, we did not vary all of them. In particular, we did not explore the impact of temperature and rainfall on population size, maturation rate, or adult mortality. It has been shown that life-history parameters vary in space and time, in response to local mosquito density[173, 86] and environmental conditions. Additionally, CRISPR-based gene drives demonstrate environmental variability, specifically

in relation to temperature-sensitive cleavage rates[103, 133]. Next steps would include the impact of environmental variability in explorations to generate ecologically applicable expectations.

As insecticides and bed nets become less effective, it is only a matter of time before CRISPR-based gene drives are used to combat mosquito-vectored diseases. It is our responsibility to design, and thoroughly evaluate, the safety and efficacy of novel drives[128]. Towards these goals, we propose tGD as a lab-safe, split-gene design combining the efficacy of a linked drive with the safety and stability of a split drive[30]. It demonstrates rapid efficacy and, through the removal of Cas9, low long-term risk of background mutation (Fig. 3.7). It combines the best aspects of linked and split drive designs - rapid and high introgression levels combined with removal of Cas9 (Fig. 3.8, Sup. Fig. A.2). The tGD system is successful with few, small releases, significantly fewer and smaller than split GDs and ClvR designs, but with the complication of being more invasive. This makes tGD an excellent design for small-scale releases to rapidly convert isolated populations.

3.2.4 Materials and Methods

We used the MGDrivE framework[177] (https://marshalllab.github.io/MGDrivE/) to simulate releases of adult *Ae. aegypti* males homozygous for each of the drive systems - tGD, sGD, two-locus CleavR and GD - in a two-node network, and in the community of Yorkeys Knob in Queensland, Australia. Briefly, the MGDrivE framework models the egg, larval, pupal, and adult (male and female) mosquito life stages implementing a daily time step, overlapping generations and a mating structure in which adult males mate throughout their lifetime, while adult females mate once upon emergence, retaining the genetic material of the adult male with whom they mate for the duration of their adult lifespan. Density-independent mortality rates for the juvenile life stages are assumed to be identical and are chosen for consistency with the population growth rate in the absence of density-dependent mortality. Additional density-dependent mortality occurs at the larval stage, the form of which is taken from Deredec *et al.*[46]. Full details of the modeling framework are available in the S1 Text of Sánchez *et al.*[177] and in the software documentation available at https://marshalllab.github.io/MGDrivE/docs/reference/. Parameters describing *Ae. aegypti* life history and the gene drive systems and landscape of interest are listed in Table A.1.

The inheritance patterns of tGD, sGD, two-locus CleavR and GD are modeled within the inheritance module of the MGDrivE framework[177], and genotype-specific effects on female fecundity, male mating competitiveness and adult lifespan influence parameters in the life history module. For the linked GD construct, the Cas9, gRNAs and effector gene are all present at the same locus[62]. We consider a linked GD construct with 100% DNA cleavage efficiency, a variable rate of accurate homology-directed repair (HDR), and 33% of inaccurate DNA repair events leading to functional homing-resistant alleles. For consistency across

all modeled gene drive systems, parameter values are shared, including between female and male gene drive organisms. For sGD, the Cas9 and gRNAs are unlinked, segregating on independent, autosomal chromosomes[123], with the effector gene linked to the gRNAs. Drive parameters at the gRNA locus are the same as for the linked GD construct, and drive does not occur for the Cas9 element, which is inherited in a Mendelian fashion. For tGD, the Cas9 and gRNAs/effector gene are again present on separate autosomal chromosomes; however, gRNAs are included that target wild-type alleles for both elements, leading to independent drive of both the Cas9 and gRNAs/effector gene[129]. Drive parameters at both loci are the same as for the linked GD construct, and a constant HDR rate of 99% is modeled at the gRNA locus. As these three gene drive systems - GD, sGD and tGD - are different versions of the same construct (linked Cas9/gRNAs, separated Cas9/gRNAs with Mendelian inheritance, or separated Cas9/gRNAs driven independently), they behave in a similar fashion. In the presence of both Cas9 and gRNAs, wild-type alleles are cleaved at a rate of cM(cF for females). 1-cM(cF) of those alleles undergo HDR, at a rate chM(chF). Alleles that are not repaired via HDR, 1-chM(chF), undergo NHEJ at the rates specified above.

This double-stranded break undergoes either homology-directed repair (HDR) or non-homologous end-joining (NHEJ). HDR events, using the Cas9 or gRNA/effector elements as repair templates, experience gene conversion, propagating the Cas9 or gRNA/effector element to the sister chromosome and biasing inheritance above expected Mendelian ratios[61]. NHEJ events lead to in-frame, thus functional, alleles or frame-shifts, resulting in nonsense mutations and non-functional alleles.

CleavR is implemented according to Oberhofer *et al.*[160]. It is assumed that all pieces are unlinked and the target gene is haplosufficent.

Fitness costs are associated with the Cas9 element, the gRNA element, and non-functional NHEJ alleles at both loci (if separated). Cas9 and gRNA costs are implemented additively, with a maximum of 100%. Non-functional NHEJ alleles are implemented as a recessive cost. ClvR assumes inviability of individuals homozygous for NHEJ alleles at the target locus.

All simulations were performed as 2, well-mixed populations of 10,000 wild-type *Ae. aegypti*. Releases of modified mosquitoes were performed weekly, using homozygous-drive mosquitoes (homozygous at both loci if split-gene drive or tGD). For confineability sweeps, migration between populations is defined as a rate per generation. This is implemented as a multinomial distribution over nodes. Simulations were performed in R[169], with analysis in a custom python package. All code is available upon request.

3.2.5 Availability of Data and Materials

The distribution of households in Yorkeys Knob and Trinity Park was obtained from OpenStreetMap. The software package, MGDrivE, is available on GitHub[177]. All information

required to reproduce the simulations is available in the GitHub deposition, the manuscript, and its additional files.

Conclusion

Our final chapter probed two very different predictive simulations - classic drive designs applied to a more complex landscape than any previously, and an extremely complex drive applied to a relatively simple landscape. These two explorations not only demonstrate the flexibility of our simulation software, but also the wide range of questions representative of where a gene drive design is in the development process. Translocations and underdominant designs are well known, simply generated using a novel CRISPR-based approach here, and while the basic dynamics are understood, it is time to explore their behavior on more realistic landscapes. The trans-complementing design is extremely young and very little is understood about the dynamics it may follow, indicating that preliminary explorations on simple landscapes are necessary to understand the basics.

The dichotomy of early experiments delving into basic behaviors of novel designs, contrasted with detailed explorations of well-known constructs applied to biologically-relevant ecosystems, is exemplified throughout this thesis. Chapter 1 begins with a basic mathematical model capable of rapidly integrating and testing novel designs in marginally realistic landscapes. It ends with a new model, with all of the capabilities of the first model, integrating time-varying ecology for seasonal dynamics. Chapter 2 begins with first-generation, linked CRISPR drive targeting a non-essential gene. It then explores five split gene drive designs targeting various essential and non-essential genes, required at different stages of the life cycle, with multiple Cas9 promotors for differential expression. This is a powerful demonstration of how far the field has come in just a few years, and provides some indication of how far we may expect it to go in the next few years.

The rate of progress in this field sets the context for my thesis. For all of the progress made, all of the novel ideas and effort, this work will be outdated before it is finished. The lasting impact stems from the solid foundation of statistically robust, and biologically representative, approaches to mathematical modeling and predictive simulation. The models in chapter 1 are powerful, but fall short in terms of epidemiological impact. The methods outlined in chapter 2 are rapidly advancing, in terms of computational performance and response to sensitivity metrics, but there are several variance-based techniques that may benefit the sensitivity metrics. And while chapter 3 is a tour de force in predictive simulation, it represents a snapshot of possible gene drive designs and locations that need explored. Every novel design, at any potential location, requires its own predictive work to inform the community about the ecological risks, and possible epidemiological benefits, for that place and time. This risk-benefit assessment may never end, but it will benefit from the tools and approaches provided herein.

Bibliography

[1] Z. N. Adelman, D. Pledger, and K. M. Myles. "Developing standard operating procedures for gene drive research in disease vector mosquitoes". In: *Pathog Glob Health* 111.8 (Dec. 2017), pp. 436–447. DOI: 10.1038/nbt.3926.

[2] Adriana Adolfi et al. "Efficient population modification gene-drive rescue system in the malaria mosquito Anopheles stephensi". In: *Nature Communications* 11.1 (Nov. 2020), p. 5553. ISSN: 2041-1723. DOI: 10.1038/s41467-020-19426-0. URL: https://doi.org/10.1038/s41467-020-19426-0.

[3] O. S. Akbari et al. "BIOSAFETY. Safeguarding gene drive experiments in the laboratory". In: *Science* 349.6251 (Aug. 2015), pp. 927–929. DOI: 10.1126/science.aac7932. URL: https://doi.org/10.1126/science.aac7932.

[4] Omar S. Akbari et al. "A Synthetic Gene Drive System for Local, Reversible Modification and Suppression of Insect Populations". In: *Current Biology* 23.8 (Apr. 2013), pp. 671–677. ISSN: 0960-9822. DOI: 10.1016/j.cub.2013.02.059. URL: https://doi.org/10.1016/j.cub.2013.02.059.

[5] Omar S. Akbari et al. "Novel Synthetic Medea Selfish Genetic Elements Drive Population Replacement in Drosophila; a Theoretical Exploration of Medea-Dependent Population Suppression". In: *ACS Synthetic Biology* 3.12 (Dec. 2014), pp. 915–928. DOI: 10.1021/sb300079h. URL: https://doi.org/10.1021/sb300079h.

[6] Luke Alphey. "Re-engineering the sterile insect technique". In: *Insect Biochemistry and Molecular Biology* 32.10 (Oct. 2002), pp. 1243–1247. ISSN: 0965-1748. URL: https://www.sciencedirect.com/science/article/pii/S0965174802000875.

[7] Philipp M. Altrock et al. "Using underdominance to bi-stably transform local populations". In: *Journal of Theoretical Biology* 267.1 (Nov. 2010), pp. 62–75. ISSN: 0022-5193. URL: https://www.sciencedirect.com/science/article/pii/S0022519310004030.

[8] M. Amaral et al. "Structural basis of kynurenine 3-monooxygenase inhibition". In: *Nature* 496.7445 (Apr. 2013), pp. 382–385.

[9] David Ardia et al. "Differential Evolution with DEoptim: An Application to Non-Convex Portfolio Optimization". In: *R Journal* 3.1 (2011), pp. 27–34. DOI: 10.32614/RJ-2011-005.

[10] Olle AU - Terenius, Jennifer AU - Juhn, and Anthony A. AU - James. "Injection of An. stephensi Embryos to Generate Malaria-resistant Mosquitoes". In: *JoVE* 5 (July 2007), e216. ISSN: 1940-087X. DOI: 10.3791/216. URL: https://doi.org/10.3791/216.

[11] M N Bayoh and S W Lindsay. "Effect of temperature on the development of the aquatic stages of Anopheles gambiae sensu stricto (Diptera: Culicidae)". en. In: *Bull. Entomol. Res.* 93.5 (Oct. 2003), pp. 375–381.

[12] Andrea K. Beaghton et al. "Gene drive for population genetic control: non-functional resistance and parental effects". In: *Proceedings of the Royal Society B: Biological Sciences* 286.1914 (Nov. 2019), p. 20191586. DOI: 10.1098/rspb.2019.1586. URL: https://doi.org/10.1098/rspb.2019.1586.

[13] Lindsay M Beck-Johnson et al. "The importance of temperature fluctuations in understanding mosquito population dynamics and malaria risk". en. In: *R Soc Open Sci* 4.3 (Mar. 2017), p. 160969.

[14] M. Q. Benedict et al. "Recommendations for Laboratory Containment and Management of Gene Drive Systems in Arthropods". In: *Vector Borne Zoonotic Dis* 18.1 (Jan. 2018), pp. 2–13.

[15] Mark Benedict et al. *Guidance framework for testing of genetically modified mosquitoes*. 2014.

[16] Samir Bhatt et al. "The global distribution and burden of dengue". In: *Nature* 496.7446 (Apr. 2013), pp. 504–507. ISSN: 1476-4687. DOI: 10.1038/nature12060. URL: https://doi.org/10.1038/nature12060.

[17] A. M. Bolger, M. Lohse, and B. Usadel. "Trimmomatic: a flexible trimmer for Illumina sequence data". In: *Bioinformatics* 30.15 (Aug. 2014), pp. 2114–2120.

[18] Luca Bortolussi et al. "Continuous approximation of collective system behaviour: A tutorial". In: *Performance Evaluation* 70.5 (2013), pp. 317–349. ISSN: 0166-5316. DOI: https://doi.org/10.1016/j.peva.2013.01.001. URL: https://www.sciencedirect.com/science/article/pii/S0166531613000023.

[19] Anna Buchman et al. "Engineered resistance to Zika virus in transgenic Aedes aegypti expressing a polycistronic cluster of synthetic small RNAs". In: *Proceedings of the National Academy of Sciences* 116.9 (Feb. 2019), p. 3656. DOI: 10.1073/pnas.1810771116. URL: https://doi.org/10.1073/pnas.1810771116.

[20] Anna Buchman et al. "Synthetically engineered Medea gene drive system in the worldwide crop pest Drosophila suzukii". In: *Proceedings of the National Academy of Sciences* 115.18 (2018), pp. 4725–4730. ISSN: 0027-8424. DOI: 10.1073/pnas.1713139115. eprint: https://www.pnas.org/content/115/18/4725.full.pdf. URL: https://www.pnas.org/content/115/18/4725.

[21] Anna B. Buchman et al. "Engineered Reciprocal Chromosome Translocations Drive High Threshold, Reversible Population Replacement in Drosophila". In: *ACS Synthetic Biology* 7.5 (May 2018), pp. 1359–1370. DOI: 10.1021/acssynbio.7b00451. URL: https://doi.org/10.1021/acssynbio.7b00451.

[22] Austin Burt. "Site-specific selfish genes as tools for the control and genetic engineering of natural populations". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270.1518 (2003), pp. 921–928. DOI: 10.1098/rspb.2002.2319. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rspb.2002.2319. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2002.2319.

[23] R. Carballar-Lejarazú and A. A. James. "Population modification of Anopheline species to control malaria transmission". In: *Pathog Glob Health* 111.8 (Dec. 2017), pp. 424–435.

[24] Rebeca Carballar-Lejarazú et al. "Next-generation gene drive for population modification of the malaria vector mosquito," en. In: *Proc. Natl. Acad. Sci. U. S. A.* 117.37 (Sept. 2020), pp. 22805–22814.

[25] James R. Carey, Pablo Liedo, and James W. Vaupel. "Mortality dynamics of density in the mediterranean fruit fly". In: *Experimental Gerontology* 30.6 (1995), pp. 605–629. ISSN: 0531-5565. DOI: https://doi.org/10.1016/0531-5565(95)00013-5. URL: https://www.sciencedirect.com/science/article/pii/0531556595000135.

[26] Danilo O. Carvalho et al. "Suppression of a Field Population of Aedes aegypti in Brazil by Sustained Release of Transgenic Male Mosquitoes". In: *PLOS Neglected Tropical Diseases* 9.7 (July 2015), e0003864. DOI: 10.1371/journal.pntd.0003864. URL: https://doi.org/10.1371/journal.pntd.0003864.

[27] J. Champer et al. "A CRISPR homing gene drive targeting a haplolethal gene removes resistance alleles and successfully spreads through a cage population". In: *Proc Natl Acad Sci U S A* 117.39 (Sept. 2020), pp. 24377–24383.

[28] Jackson Champer, Anna Buchman, and Omar S. Akbari. "Cheating evolution: engineering gene drives to manipulate the fate of wild populations". In: *Nature Reviews Genetics* 17.3 (Mar. 2016), pp. 146–159. ISSN: 1471-0064. DOI: 10.1038/nrg.2015.34. URL: https://doi.org/10.1038/nrg.2015.34.

[29] Jackson Champer et al. "A toxin-antidote CRISPR gene drive system for regional population modification". en. In: *Nat. Commun.* 11.1 (Feb. 2020), p. 1082. DOI: 10.1038/s41467-020-14960-3.

[30] Jackson Champer et al. "Molecular safeguarding of CRISPR gene drive experiments". In: *eLife* 8 (Jan. 2019). eLife 2019;8:e41439, e41439. ISSN: 2050-084X. DOI: 10.7554/eLife.41439. URL: https://doi.org/10.7554/eLife.41439.

[31] Jackson Champer et al. "Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations". In: *PLOS Genetics* 13.7 (July 2017), pp. 1–18. DOI: 10.1371/journal.pgen.1006796. URL: https://doi.org/10.1371/journal.pgen.1006796.

[32] Jackson Champer et al. "Population Dynamics of Underdominance Gene Drive Systems in Continuous Space". In: *ACS Synthetic Biology* 9.4 (Apr. 2020), pp. 779–792. DOI: 10.1021/acssynbio.9b00452. URL: https://doi.org/10.1021/acssynbio.9b00452.

[33] Jackson Champer et al. "Reducing resistance allele formation in CRISPR gene drive". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 115.21 (May 2018), pp. 5522–5527. DOI: 10.1073/pnas.1720354115. URL: https://doi.org/10.1073/pnas.1720354115.

[34] H. H. Y. Chang et al. "Non-homologous DNA end joining and alternative pathways to double-strand break repair". In: *Nat. Rev. Mol. Cell Biol.* 18 (2017). DOI: 10.1038/nrm.2017.48. URL: https://doi.org/10.1038/nrm.2017.48.

[35] C. H. Chen et al. "A synthetic maternal-effect selfish genetic element drives population replacement in Drosophila". In: *Science* 316.5824 (Apr. 2007), pp. 597–600. DOI: 10.1126/science.1138595. URL: https://doi.org/10.1126/science.%201138595.

[36] Savvas Christoforidis et al. "The Rab5 effector EEA1 is a core component of endosome docking". In: *Nature* 397.6720 (Feb. 1999), pp. 621–625. ISSN: 1476-4687. DOI: 10.1038/17618. URL: https://doi.org/10.1038/17618.

[37] S. R. Christophers. *Aëdes aegypti (L.) the yellow fever mosquito; its life history, bionomics, and structure.* English. Cambridge [England: University Press, 1960.

[38] Jackie Cook et al. "Implications of insecticide resistance for malaria vector control with long-lasting insecticidal nets: trends in pyrethroid resistance during a WHO-coordinated multi-country prospective study". In: *Parasites & Vectors* 11.1 (Oct. 2018), p. 550. ISSN: 1756-3305. DOI: 10.1186/s13071-018-3101-4. URL: https://doi.org/10.1186/s13071-018-3101-4.

[39] Jacob E. Crawford et al. "Efficient production of male Wolbachia-infected Aedes aegypti mosquitoes enables large-scale suppression of wild populations". In: *Nature Biotechnology* 38.4 (Apr. 2020), pp. 482–492. ISSN: 1546-1696. DOI: 10.1038/s41587-020-0471-x. URL: https://doi.org/10.1038/s41587-020-0471-x.

[40] C. F. CURTIS. "Possible Use of Translocations to fix Desirable Genes in Insect Pest Populations". In: *Nature* 218.5139 (Apr. 1968), pp. 368–369. ISSN: 1476-4687. DOI: 10.1038/218368a0. URL: https://doi.org/10.1038/218368a0.

[41] C. F. Curtis and A. S. Robinson. "Computer simulation of the use of double translocations for pest control". eng. In: *Genetics* 69.1 (Sept. 1971). PMC1212692[pmcid], pp. 97–113. ISSN: 0016-6731. URL: https://pubmed.ncbi.nlm.nih.gov/5167305.

[42] M. Daszykowski and B. Walczak. "2.29 - Density-Based Clustering Methods". In: *Comprehensive Chemometrics*. Ed. by Steven D. Brown, Romá Tauler, and Beata Walczak. Oxford: Elsevier, 2009, pp. 635–654. ISBN: 978-0-444-52701-1. DOI: https://doi.org/10.1016/B978-044452701-1.00067-3. URL: https://www.sciencedirect.com/science/article/pii/B9780444527011000673.

[43] STEPHEN DAVIS, NICHOLAS BAX, and PETER GREWE. "Engineered Underdominance Allows Efficient and Economical Introgression of Traits into Pest Populations". In: *Journal of Theoretical Biology* 212.1 (Sept. 2001), pp. 83–98. ISSN: 0022-5193. URL: https://www.sciencedirect.com/science/article/pii/S0022519301923574.

[44] J. Dehairs et al. "CRISP-ID: decoding CRISPR mediated indels by Sanger sequencing". In: *Sci Rep* 6 (July 2016), p. 28973.

[45] Jean-Marc O. Depinay et al. "A simulation model of African Anopheles ecology and population dynamics for the analysis of malaria transmission". In: *Malaria Journal* 3.1 (July 2004), p. 29. ISSN: 1475-2875. DOI: 10.1186/1475-2875-3-29. URL: https://doi.org/10.1186/1475-2875-3-29.

[46] Anne Deredec, H. Charles J. Godfray, and Austin Burt. "Requirements for effective malaria control with homing endonuclease genes". In: *Proceedings of the National Academy of Sciences* 108.43 (2011), E874–E880. ISSN: 0027-8424. DOI: 10.1073/pnas.1110717108. eprint: https://www.pnas.org/content/108/43/E874.full.pdf. URL: https://www.pnas.org/content/108/43/E874.

[47] Alexandros D. Diamantidis et al. "Population-specific demography and invasion potential in medfly". In: *Ecology and Evolution* 1.4 (2011), pp. 479–488. DOI: https://doi.org/10.1002/ece3.33. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ece3.33. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.33.

[48] Nico Dissmeyer. "Conditional Modulation of Biological Processes by Low-Temperature Degrons". en. In: *Methods Mol. Biol.* 1669 (2017), pp. 407–416.

[49] Y. Dong, M. L. Simões, and G. Dimopoulos. "Versatile transgenic multistage effector-gene combinations for Plasmodium falciparum suppression in Anopheles". In: *Sci Adv* 6.20 (May 2020), eaay5898.

[50] Jennifer A. Doudna and Emmanuelle Charpentier. "The new frontier of genome engineering with CRISPR-Cas9". In: *Science* 346.6213 (Nov. 2014), p. 1258096. DOI: 10.1126/science.1258096. URL: https://doi.org/10.1126/science.1258096.

[51] Philip A. Eckhoff. "A malaria transmission-directed model of mosquito life cycle and ecology". In: *Malaria Journal* 10.1 (Oct. 2011), p. 303. ISSN: 1475-2875. DOI: 10.1186/1475-2875-10-303. URL: https://doi.org/10.1186/1475-2875-10-303.

[52] Matthew P. Edgington and Luke S. Alphey. "Modeling the mutation and reversal of engineered underdominance gene drives". In: *Journal of Theoretical Biology* 479 (Oct. 2019), pp. 14–21. ISSN: 0022-5193. URL: https://www.sciencedirect.com/science/article/pii/S0022519319302590.

[53] Lars Eisen et al. "The Impact of Temperature on the Bionomics of Aedes (Stegomyia) aegypti, With Special Reference to the Cool Geographic Range Margins". In: *Journal of Medical Entomology* 51.3 (May 2014), pp. 496–516. ISSN: 0022-2585. DOI: 10.1603/ME13214. URL: https://doi.org/10.1603/ME13214.

[54] Kevin M Esvelt et al. "Emerging Technology: Concerning RNA-guided gene drives for the alteration of wild populations". In: *eLife* 3 (July 2014). Ed. by Diethard Tautz, e03401. ISSN: 2050-084X. DOI: 10.7554/eLife.03401. URL: https://doi.org/10.7554/eLife.03401.

[55] D. A. Focks et al. "Dynamic Life Table Model for Aedes aegypti (Diptera: Culicidae): Analysis of the Literature and Model Development". In: *Journal of Medical Entomology* 30.6 (Nov. 1993), pp. 1003–1017. ISSN: 0022-2585. DOI: 10.1093/jmedent/30.6.1003. eprint: https://academic.oup.com/jme/article-pdf/30/6/1003/18180193/jmedent30-1003.pdf. URL: https://doi.org/10.1093/jmedent/30.6.1003.

[56] Dana A. Focks et al. "A Simulation Model of the Epidemiology of Urban Dengue Fever: Literature Analysis, Model Development, Preliminary Validation, and Samples of Simulation Results". In: *The American Journal of Tropical Medicine and Hygiene* 53.5 (Nov. 1995), pp. 489–506. DOI: 10.4269/ajtmh.1995.53.489. URL: https://www.ajtmh.org/view/journals/tpmd/53/5/article-p489.xml.

[57] Kevin J. Forsberg et al. "Functional metagenomics-guided discovery of potent Cas9 inhibitors in the human microbiome". In: *eLife* 8 (Sept. 2019). eLife 2019;8:e46540, e46540. ISSN: 2050-084X. DOI: 10.7554/eLife.46540. URL: https://doi.org/10.7554/eLife.46540.

[58] Linton C. Freeman. "Centrality in social networks conceptual clarification". In: *Social Networks* 1.3 (1978), pp. 215–239. ISSN: 0378-8733. DOI: https://doi.org/10.1016/0378-8733(78)90021-7. URL: https://www.sciencedirect.com/science/article/pii/0378873378900217.

[59] Roberto Galizi et al. "A synthetic sex ratio distortion system for the control of the human malaria mosquito". In: *Nature Communications* 5.1 (June 2014), p. 3977. ISSN: 2041-1723. DOI: 10.1038/ncomms4977. URL: https://doi.org/10.1038/ncomms4977.

[60] Valentino M Gantz and Ethan Bier. "The dawn of active genetics". en. In: *Bioessays* 38.1 (Jan. 2016), pp. 50–63. DOI: 10.1002/bies.201500102.

[61] Valentino M. Gantz and Ethan Bier. "Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations". eng. In: *Science (New York, N.Y.)* 348.6233 (Apr. 2015). science.aaa5945[PII], pp. 442–444. ISSN: 1095-9203. DOI: 10.1126/science.aaa5945. URL: https://doi.org/10.1126/science.aaa5945.

[62] Valentino M. Gantz et al. "Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi". In: *Proceedings of the National Academy of Sciences* 112.49 (2015), E6736–E6743. ISSN: 0027-8424. DOI: 10.1073/pnas.1521077112. eprint: https://www.pnas.org/content/112/49/E6736.full.pdf. URL: https://www.pnas.org/content/112/49/E6736.

[63] Daniel T. Gillespie. "Approximate accelerated stochastic simulation of chemically reacting systems". In: *The Journal of Chemical Physics* 115.4 (July 2001), pp. 1716–1733. ISSN: 0021-9606. DOI: 10.1063/1.1378322. URL: https://doi.org/10.1063/1.1378322.

[64] Daniel T. Gillespie. "Exact stochastic simulation of coupled chemical reactions". In: *The Journal of Physical Chemistry* 81.25 (Dec. 1977), pp. 2340–2361. ISSN: 0022-3654. DOI: 10.1021/j100540a008. URL: https://doi.org/10.1021/j100540a008.

[65] Daniel T. Gillespie. "Stochastic Simulation of Chemical Kinetics". In: *Annual Review of Physical Chemistry* 58.1 (May 2007), pp. 35–55. ISSN: 0066-426X. DOI: 10.1146/annurev.physchem.58.032806.104637. URL: https://doi.org/10.1146/annurev.physchem.58.032806.104637.

[66] Daniel T. Gillespie. "The multivariate Langevin and Fokker–Planck equations". In: *American Journal of Physics* 64.10 (Oct. 1996), pp. 1246–1257. ISSN: 0002-9505. DOI: 10.1119/1.18387. URL: https://doi.org/10.1119/1.18387.

[67] G. B. Gloor et al. "Type I repressors of P element mobility." In: *Genetics* 135.1 (Sept. 1993), pp. 81–95. ISSN: 1943-2631. DOI: 10.1093/genetics/135.1.81. URL: https://doi.org/10.1093/genetics/135.1.81.

[68] Anita Q. Gomes et al. "Membrane Targeting of Rab GTPases Is Influenced by the Prenylation Motif". In: *Molecular Biology of the Cell* 14.5 (May 2003), pp. 1882–1899. ISSN: 1059-1524. DOI: 10.1091/mbc.e02-10-0639. URL: https://doi.org/10.1091/mbc.e02-10-0639.

[69] P J Goss and J Peccoud. "Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 95.12 (June 1998), pp. 6750–6755.

[70] Fred Gould et al. "A Killer–Rescue system for self-limiting gene drive of anti-pathogen constructs". In: *Proceedings of the Royal Society B: Biological Sciences* 275.1653 (Dec. 2008), pp. 2823–2829. DOI: 10.1098/rspb.2008.0846. URL: https://doi.org/10.1098/rspb.2008.0846.

[71] S. J. Gratz et al. "Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila". In: *Genetics* 196.4 (Apr. 2014), pp. 961–971.

[72] Jamie T. Griffin et al. "Reducing Plasmodium falciparum Malaria Transmission in Africa: A Model-Based Evaluation of Intervention Strategies". In: *PLOS Medicine* 7.8 (Aug. 2010), e1000324. DOI: 10.1371/journal.pmed.1000324. URL: https://doi.org/10.1371/journal.pmed.1000324.

[73] Anja Gronewold and Michael Sonnenschein. "Event-based modelling of ecological systems with asynchronous cellular automata". In: *Ecological Modelling* 108.1 (May 1998), pp. 37–52. ISSN: 0304-3800. URL: https://www.sciencedirect.com/science/article/pii/S0304380098000179.

[74] Carlos A. Guerra et al. "A global assembly of adult female mosquito mark-release-recapture data to inform the control of mosquito-borne pathogens". In: *Parasites & Vectors* 7.1 (June 2014), p. 276. ISSN: 1756-3305. DOI: 10.1186/1756-3305-7-276. URL: https://doi.org/10.1186/1756-3305-7-276.

[75] A. Guichard et al. "Efficient allelic-drive in Drosophila". In: *Nat Commun* 10.1 (Apr. 2019), p. 1640. DOI: 10.1038/s41467-019-09694-w.

[76] Peter J Haas. *Stochastic Petri Nets: Modelling, Stability, Simulation*. en. Springer Science & Business Media, Apr. 2006.

[77] J. E. Haber. "TOPping off meiosis". In: *Mol. Cell* 57 (2015). DOI: 10.1016/j.molcel.2015.03.008. URL: https://doi.org/10.1016/j.molcel.2015.03.008.

[78] Benjamin C Haller and Philipp W Messer. "SLiM 2: Flexible, interactive forward genetic simulations". In: *Molecular biology and evolution* 34.1 (2017), pp. 230–240.

[79] Andrew Hammond et al. In: *Nature Portfolio* (June 2021). ISSN: 2693-5015. DOI: 10.21203/rs.3.rs-411410/v1. URL: https://doi.org/10.21203/rs.3.rs-411410/v1.

[80] Andrew Hammond et al. "A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae". In: *Nature Biotechnology* 34.1 (Jan. 2016), pp. 78–83. ISSN: 1546-1696. DOI: 10.1038/nbt.3439. URL: https://doi.org/10.1038/nbt.3439.

[81] Andrew Hammond et al. "Regulation of gene drive expression increases invasive potential and mitigates resistance". In: *bioRxiv* (Jan. 2020), p. 360339. DOI: 10.1101/360339. URL: https://doi.org/10.1101/360339.

[82] Andrew M. Hammond et al. "The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito". In: *PLOS Genetics* 13.10 (Oct. 2017), e1007039. DOI: 10.1371/journal.pgen.1007039. URL: https://doi.org/10.1371/journal.pgen.1007039.

[83] Q. Han et al. "Analysis of the wild-type and mutant genes encoding the enzyme kynurenine monooxygenase of the yellow fever mosquito, Aedes aegypti". In: *Insect Mol Biol* 12.5 (Oct. 2003), pp. 483–490.

[84] Penelope A. Hancock and H. Charles J. Godfray. "Modelling the spread of Wolbachia in spatially heterogeneous environments". In: *Journal of The Royal Society Interface* 9.76 (Nov. 2012), pp. 3045–3054. DOI: 10.1098/rsif.2012.0253. URL: https://doi.org/10.1098/rsif.2012.0253.

[85] Penelope A. Hancock, Steven P. Sinkins, and H. Charles J. Godfray. "Population Dynamic Models of the Spread of Wolbachia". In: *The American Naturalist* 177.3 (2011). PMID: 21460541, pp. 323–333. DOI: 10.1086/658121. eprint: https://doi.org/10.1086/658121. URL: https://doi.org/10.1086/658121.

[86] Penelope A. Hancock et al. "Density-dependent population dynamics in Aedes aegypti slow the spread of wMel Wolbachia". In: *Journal of Applied Ecology* 53.3 (June 2016), pp. 785–793. ISSN: 0021-8901. DOI: 10.1111/1365-2664.12620. URL: https://doi.org/10.1111/1365-2664.12620.

[87] Penelope A. Hancock et al. "Predicting the spatial dynamics of Wolbachia infections in Aedes aegypti arbovirus vector populations in heterogeneous landscapes". In: *Journal of Applied Ecology* 56.7 (July 2019), pp. 1674–1686. ISSN: 0021-8901. DOI: 10.1111/1365-2664.13423. URL: https://doi.org/10.1111/1365-2664.13423.

[88] Penny A. Hancock and H. Charles J. Godfray. "Application of the lumped age-class technique to studying the dynamics of malaria-mosquito-human interactions". In: *Malaria Journal* 6.1 (July 2007), p. 98. ISSN: 1475-2875. DOI: 10.1186/1475-2875-6-98. URL: https://doi.org/10.1186/1475-2875-6-98.

[89] L. C. Harrington et al. "Dispersal of the dengue vector Aedes aegypti within and between rural communities". In: *Am J Trop Med Hyg* 72.2 (Feb. 2005), pp. 209–220.

[90] W. K. Hartberg. "Observations on the mating behaviour of Aedes aegypti in nature". eng. In: *Bulletin of the World Health Organization* 45.6 (1971). PMC2428001[pmcid], pp. 847–850. ISSN: 0042-9686. URL: https://pubmed.ncbi.nlm.nih.gov/5317018.

[91] Florian Hartig, Francesco Minunno, and Stefan Paul. *BayesianTools: General-Purpose MCMC and SMC Samplers and Tools for Bayesian Statistics*. R package version 0.1.7. 2017. URL: https://github.com/florianhartig/BayesianTools.

[92] Janet Hemingway. "Resistance: A problem without an easy solution". In: *Pesticide Biochemistry and Physiology* 151 (Oct. 2018), pp. 73–75. ISSN: 0048-3575. URL: https://www.sciencedirect.com/science/article/pii/S0048357518302852.

[93] M. Hirai et al. "cDNA cloning, functional expression and characterization of kynurenine 3-hydroxylase of Anopheles stephensi (Diptera: Culicidae)". In: *Insect Mol Biol* 11.5 (Oct. 2002), pp. 497–504.

[94] Astrid Hoermann et al. "Converting endogenous genes of the malaria mosquito into simple non-autonomous gene drives for population replacement". In: *eLife* 10 (Apr. 2021). eLife 2021;10:e58791, e58791. ISSN: 2050-084X. DOI: 10.7554/eLife.58791. URL: https://doi.org/10.7554/eLife.58791.

[95] A. A. Hoffmann et al. "Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission". In: *Nature* 476.7361 (Aug. 2011), pp. 454–457. ISSN: 1476-4687. DOI: 10.1038/nature10356. URL: https://doi.org/10.1038/nature10356.

[96] W. E. Horsfall. *Mosquitoes. Their Bionomics and Relation to Disease. ID - 19552900863*. English. 15 East 26th Street, New York 10: The Eonald Press Co., 1955, viii + 723 pp.

[97] Yunxin Huang et al. "Gene-drive into insect populations with age and spatial structure: a theoretical assessment". In: *Evolutionary Applications* 4.3 (May 2011), pp. 415–428. ISSN: 1752-4571. DOI: 10.1111/j.1752-4571.2010.00153.x. URL: https://doi.org/10.1111/j.1752-4571.2010.00153.x.

[98] INSEED. *Annuaire Statistique des Comores. Moroni, Union of the Comoros*. National Institute of Statistics, Economic, and Demographic Studies (INSEED), 2015.

[99] A. T. Isaacs et al. "Engineered resistance to Plasmodium falciparum development in transgenic Anopheles stephensi". In: *PLoS Pathog* 7.4 (Apr. 2011), e1002017. DOI: 10.1371/journal.ppat.1002017. URL: https://doi.org/10.1371/journal.ppat.1002017.

[100] A. T. Isaacs et al. "Transgenic Anopheles stephensi coexpressing single-chain antibodies resist Plasmodium falciparum development". In: *Proc Natl Acad Sci U S A* 109.28 (July 2012), E1922–1930.

[101] S. James et al. "Pathway to Deployment of Gene Drive Mosquitoes as a Potential Biocontrol Tool for Elimination of Malaria in Sub-Saharan Africa: Recommendations of a Scientific Working Group". In: *Am J Trop Med Hyg* 98.6 (June 2018), pp. 1–49.

[102] Stephanie L James et al. "Toward the Definition of Efficacy and Safety Criteria for Advancing Gene Drive-Modified Mosquitoes to Field Testing". en. In: *Vector Borne Zoonotic Dis.* 20.4 (Apr. 2020), pp. 237–251.

[103] Fuguo Jiang et al. "Temperature-Responsive Competitive Inhibition of CRISPR-Cas9". In: *Molecular Cell* 73.3 (Feb. 2019), 601–610.e5. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2018.11.016. URL: https://doi.org/10.1016/j.molcel.2018.11.016.

[104] Nikolay P. Kandul et al. "A confinable home-and-rescue gene drive for population modification". In: *eLife* 10 (Mar. 2021). eLife 2021;10:e65939, e65939. ISSN: 2050-084X. DOI: 10.7554/eLife.65939. URL: https://doi.org/10.7554/eLife.65939.

[105] Nikolay P. Kandul et al. "A home and rescue gene drive efficiently spreads and persists in populations". In: *bioRxiv* (Jan. 2020), p. 2020.08.21.261610. DOI: 10.1101/2020.08.21.261610. URL: https://doi.org/10.1101/2020.08.21.261610.

[106] Nikolay P. Kandul et al. "Assessment of a Split Homing Based Gene Drive for Efficient Knockout of Multiple Genes". In: *G3: Genes—Genomes—Genetics* 10.2 (Feb. 2020), p. 827. DOI: 10.1534/g3.119.400985. URL: https://doi.org/10.1534/g3.119.400985.

[107] Nikolay P. Kandul et al. "Transforming insect population control with precision guided sterile males with demonstration in flies". In: *Nature Communications* 10.1 (Jan. 2019), p. 84. ISSN: 2041-1723. DOI: 10.1038/s41467-018-07964-7. URL: https://doi.org/10.1038/s41467-018-07964-7.

[108] K. Katoh, J. Rozewicki, and K. D. Yamada. "MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization". In: *Brief Bioinform* 20.4 (July 2019), pp. 1160–1166.

[109] William Ogilvy Kermack, A. G. McKendrick, and Gilbert Thomas Walker. "A contribution to the mathematical theory of epidemics". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115.772 (Aug. 1927), pp. 700–721. DOI: 10.1098/rspa.1927.0118. URL: https://doi.org/10.1098/rspa.1927.0118.

[110] William Ogilvy Kermack, A. G. McKendrick, and Gilbert Thomas Walker. "Contributions to the mathematical theory of epidemics. II. —The problem of endemicity". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 138.834 (Oct. 1932), pp. 55–83. DOI: 10.1098/rspa.1932.0171. URL: https://doi.org/10.1098/rspa.1932.0171.

[111] Doran Khamis et al. "Ecological effects on underdominance threshold drives for vector control". In: *Journal of Theoretical Biology* 456 (Nov. 2018), pp. 1–15. ISSN: 0022-5193. URL: https://www.sciencedirect.com/science/article/pii/S0022519318303461.

[112] C. J. Koenraadt et al. "Spatial and temporal patterns in pupal and adult production of the dengue vector Aedes aegypti in Kamphaeng Phet, Thailand". In: *Am J Trop Med Hyg* 79.2 (Aug. 2008), pp. 230–238.

[113] Adam J. Kucharski et al. "Transmission Dynamics of Zika Virus in Island Populations: A Modelling Analysis of the 2013–14 French Polynesia Outbreak". In: *PLOS Neglected Tropical Diseases* 10.5 (May 2016), pp. 1–15. DOI: 10.1371/journal.pntd.0004726. URL: https://doi.org/10.1371/journal.pntd.0004726.

[114] Kyros Kyrou et al. "A CRISPR-Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes". en. In: *Nat. Biotechnol.* 36.11 (Dec. 2018), pp. 1062–1066.

[115] Wayne G Landis. "The Origin, Development, Application, Lessons Learned, and Future Regarding the Bayesian Network Relative Risk Model for Ecological Risk Assessment". In: *Integrated Environmental Assessment and Management* 17.1 (2021), pp. 79–94. DOI: https://doi.org/10.1002/ieam.4351. eprint: https://setac.onlinelibrary.wiley.com/doi/pdf/10.1002/ieam.4351. URL: https://setac.onlinelibrary.wiley.com/doi/abs/10.1002/ieam.4351.

[116] Gilbert Le Goff et al. "Field evaluation of seasonal trends in relative population sizes and dispersal pattern of Aedes albopictus males in support of the design of a sterile male release strategy". en. In: *Parasit. Vectors* 12.1 (Feb. 2019), p. 81.

[117] Philip T. Leftwich et al. "Recent advances in threshold-dependent gene drives for mosquitoes". In: *Biochemical Society Transactions* 46.5 (Sept. 2018), pp. 1203–1212. ISSN: 0300-5127. DOI: 10.1042/BST20180076. URL: https://doi.org/10.1042/BST20180076.

[118] Mathieu Legros et al. "Assessing the Feasibility of Controlling Aedes aegypti with Transgenic Methods: A Model-Based Evaluation". In: *PLOS ONE* 7.12 (Dec. 2012), pp. 1–12. DOI: 10.1371/journal.pone.0052235. URL: https://doi.org/10.1371/journal.pone.0052235.

[119] Caroline M. Leitschuh et al. "Developing gene drive technologies to eradicate invasive rodents from islands". In: *Journal of Responsible Innovation* 5.sup1 (2018), S121–S138. DOI: 10.1080/23299460.2017.1365232. eprint: https://doi.org/10.1080/23299460.2017.1365232. URL: https://doi.org/10.1080/23299460.2017.1365232.

[120] K. F. Leung et al. "Rab GTPases containing a CAAX motif are processed post-geranylgeranylation by proteolysis and methylation". In: *J. Biol. Chem.* 282 (2007). DOI: 10.1074/jbc.M605557200. URL: https://doi.org/10.1074/jbc.M605557200.

[121] H. Li. "A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data". In: *Bioinformatics* 27.21 (Nov. 2011), pp. 2987–2993.

[122] Heng Li and Richard Durbin. "Fast and accurate short read alignment with Burrows–Wheeler transform". In: *Bioinformatics* 25.14 (July 2009), pp. 1754–1760. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btp324. URL: https://doi.org/10.1093/bioinformatics/btp324.

[123] Ming Li et al. "Development of a confinable gene drive system in the human disease vector". en. In: *Elife* 9 (Jan. 2020).

[124] Ming Li et al. "Eliminating Mosquitoes with Precision Guided Sterile Males". In: *bioRxiv* (Jan. 2021), p. 2021.03.05.434167. DOI: 10.1101/2021.03.05.434167. URL: https://doi.org/10.1101/2021.03.05.434167.

[125] Xiaolin Li, M. Parizeau, and R. Plamondon. "Training hidden Markov models with multiple observations-a combinatorial method". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.4 (2000), pp. 371–377. DOI: 10.1109/34.845379.

[126] C. C. Lin and C. J. Potter. "Editing Transgenic DNA Components by Inducible Gene Replacement in Drosophila melanogaster". In: *Genetics* 203.4 (Aug. 2016), pp. 1613–1628. DOI: 10.1534/genetics.116.191783. URL: https://doi.org/10.1534/genetics.116.191783.

[127] Jingxian Liu et al. "Maximum Likelihood Estimation of Fitness Components in Experimental Evolution". In: *Genetics* 211.3 (Mar. 2019), pp. 1005–1017. ISSN: 1943-2631. DOI: 10.1534/genetics.118.301893. URL: https://doi.org/10.1534/genetics.118.301893.

[128] Kanya C. Long et al. "Core commitments for field trials of gene drive organisms". In: *Science* 370.6523 (Dec. 2020), p. 1417. DOI: 10.1126/science.abd1908. URL: https://doi.org/10.1126/science.abd1908.

[129] Víctor López Del Amo et al. "A transcomplementing gene drive provides a flexible platform for laboratory investigation and potential field deployment". In: *Nature Communications* 11.1 (Jan. 2020), p. 352. ISSN: 2041-1723. DOI: 10.1038/s41467-019-13977-7. URL: https://doi.org/10.1038/s41467-019-13977-7.

[130] George Macdonald. *The epidemiology and control of malaria.* English. London: Oxford Univ. Press, 1957.

[131] Vanessa M. Macias, Johanna R. Ohm, and Jason L. Rasgon. "Gene Drive for Mosquito Control: Where Did It Come from and Where Are We Headed?" In: 14.9 (2017). ISSN: 1660-4601. DOI: 10.3390/ijerph14091006. URL: https://doi.org/10.3390/ijerph14091006.

[132] Krisztian Magori et al. "Skeeter Buster: A Stochastic, Spatially Explicit Modeling Tool for Studying Aedes aegypti Population Replacement and Population Suppression Strategies". In: *PLOS Neglected Tropical Diseases* 3.9 (Sept. 2009), pp. 1–18. DOI: 10.1371/journal.pntd.0000508. URL: https://doi.org/10.1371/journal.pntd.0000508.

[133] Aimee A. Malzahn et al. "Application of CRISPR-Cas12a temperature sensitivity for improved genome editing in rice, maize, and Arabidopsis". In: *BMC Biology* 17.1 (Jan. 2019), p. 9. ISSN: 1741-7007. DOI: 10.1186/s12915-019-0629-5. URL: https://doi.org/10.1186/s12915-019-0629-5.

[134] Simeone Marino et al. "A methodology for performing global uncertainty and sensitivity analysis in systems biology". In: *Journal of Theoretical Biology* 254.1 (Sept. 2008), pp. 178–196. ISSN: 0022-5193. URL: https://www.sciencedirect.com/science/article/pii/S0022519308001896.

[135] Clare D. Marsden et al. "An analysis of two island groups as potential sites for trials of transgenic mosquitoes for malaria control". In: *Evolutionary Applications* 6.4 (June 2013), pp. 706–720. ISSN: 1752-4571. DOI: 10.1111/eva.12056. URL: https://doi.org/10.1111/eva.12056.

[136] J. M. Marshall et al. "Winning the Tug-of-War Between Effector Gene Design and Pathogen Evolution in Vector Population Replacement Strategies". In: *Front Genet* 10 (2019), p. 1072.

[137] John M Marshall et al. "Semele: A Killer-Male, Rescue-Female System for Suppression and Replacement of Insect Disease Vector Populations". In: *Genetics* 187.2 (Feb. 2011), pp. 535–551. ISSN: 1943-2631. DOI: 10.1534/genetics.110.124479. eprint: https://academic.oup.com/genetics/article-pdf/187/2/535/37384388/genetics0535.pdf. URL: https://doi.org/10.1534/genetics.110.124479.

[138] John M. Marshall. "The effect of gene drive on containment of transgenic mosquitoes". In: *Journal of Theoretical Biology* 258.2 (May 2009), pp. 250–265. ISSN: 0022-5193. URL: https://www.sciencedirect.com/science/article/pii/S0022519309000496.

[139] John M. Marshall and Omar S. Akbari. "Can CRISPR-Based Gene Drive Be Confined in the Wild? A Question for Molecular and Population Biology". In: *ACS Chemical Biology* 13.2 (Feb. 2018), pp. 424–430. ISSN: 1554-8929. DOI: 10.1021/acschembio.7b00923. URL: https://doi.org/10.1021/acschembio.7b00923.

[140] John M. Marshall and Bruce A. Hay. "Confinement of gene drive systems to local populations: A comparative analysis". In: *Journal of Theoretical Biology* 294 (Feb. 2012), pp. 153–171. ISSN: 0022-5193. URL: https://www.sciencedirect.com/science/article/pii/S0022519311005558.

[141] John M. Marshall and Bruce A. Hay. "GENERAL PRINCIPLES OF SINGLE-CONSTRUCT CHROMOSOMAL GENE DRIVE". In: *Evolution* 66.7 (2012), pp. 2150–2166. DOI: https://doi.org/10.1111/j.1558-5646.2012.01582.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1558-5646.2012.01582.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1558-5646.2012.01582.x.

[142] John M. Marshall et al. "Overcoming evolved resistance to population-suppressing homing-based gene drives". In: *Scientific Reports* 7.1 (June 2017), p. 3776. ISSN: 2045-2322. DOI: 10.1038/s41598-017-02744-7. URL: https://doi.org/10.1038/s41598-017-02744-7.

[143] Antonio Martínez-Ruiz et al. *Global Sensitivity Analysis and Calibration by Differential Evolution Algorithm of HORTSYST Crop Model for Fertigation Management*. 2021. DOI: 10.3390/w13050610. URL: https://doi.org/10.3390/w13050610.

[144] Maciej Maselko et al. "Engineering species-like barriers to sexual reproduction". In: *Nature Communications* 8.1 (Oct. 2017), p. 883. ISSN: 2041-1723. DOI: 10.1038/s41467-017-01007-3. URL: https://doi.org/10.1038/s41467-017-01007-3.

[145] G. R. McFarlane, C. B. A. Whitelaw, and S. G. Lillico. "CRISPR-based gene drives for pest control". In: *Trends Biotechnol.* 36 (2018). DOI: 10.1016/j.tibtech.2017.10.001. URL: https://doi.org/10.1016/j.tibtech.2017.10.001.

[146] K. S. McKim and A. Hayashi-Hagihara. "mei-W68 in Drosophila melanogaster encodes a Spo11 homolog: evidence that the mechanism for initiating meiotic recombination is conserved". In: *Genes Dev.* 12 (1998). DOI: 10.1101/gad.12.18.2932. URL: https://doi.org/10.1101/gad.12.18.2932.

[147] Kathrin Menberg, Yeonsook Heo, and Ruchi Choudhary. "Sensitivity analysis methods for building energy models: Comparing computational costs and extractable information". In: *Energy and Buildings* 133 (Dec. 2016), pp. 433–445. ISSN: 0378-7788. URL: https://www.sciencedirect.com/science/article/pii/S0378778816311112.

[148] Louis Molineaux, G Gramiccia, and World Health Organization. *The Garki project : research on the epidemiology and control of malaria in the Sudan savanna of West Africa / by L. Molineaux and G. Gramiccia*. 1980.

[149] Erin A. Mordecai et al. "Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models". In: *PLOS Neglected Tropical Diseases* 11.4 (Apr. 2017), e0005568. DOI: 10.1371/journal.pntd.0005568. URL: https://doi.org/10.1371/journal.pntd.0005568.

[150] Erin A. Mordecai et al. "Thermal biology of mosquito-borne disease". In: *Ecology Letters* 22.10 (Oct. 2019), pp. 1690–1708. ISSN: 1461-023X. DOI: 10.1111/ele.13335. URL: https://doi.org/10.1111/ele.13335.

[151] Katharine Mullen et al. "DEoptim: An R Package for Global Optimization by Differential Evolution". In: *Journal of Statistical Software* 40.6 (2011), pp. 1–26. DOI: 10.18637/jss.v040.i06.

[152] Simon M Muriu et al. "Larval density dependence in Anopheles gambiae s.s., the major African vector of malaria". en. In: *J. Anim. Ecol.* 82.1 (Jan. 2013), pp. 166–174.

[153] A. Nash et al. "Integral gene drives for population replacement". In: *Biol Open* 8.1 (Jan. 2019). DOI: 10.1242/bio.037762. URL: https://doi.org/10.1242/bio.037762.

[154] Charleston Noble et al. "Current CRISPR gene drive systems are likely to be highly invasive in wild populations". In: *eLife* 7 (June 2018). eLife 2018;7:e33423, e33423. ISSN: 2050-084X. DOI: 10.7554/eLife.33423. URL: https://doi.org/10.7554/eLife.33423.

[155] Charleston Noble et al. "Daisy-chain gene drives for the alteration of local populations". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 116.17 (Apr. 2019), pp. 8275–8282. DOI: 10.1073/pnas.1716358116. URL: https://doi.org/10.1073/pnas.1716358116.

[156] Charleston Noble et al. "Evolutionary dynamics of CRISPR gene drives". In: *Science Advances* 3.4 (2017). DOI: 10.1126/sciadv.1601964. eprint: https://advances.sciencemag.org/content/3/4/e1601964.full.pdf. URL: https://advances.sciencemag.org/content/3/4/e1601964.

[157] Ace R. North, Austin Burt, and H. Charles J. Godfray. "Modelling the potential of genetic control of malaria mosquitoes at national scale". In: *BMC Biology* 17.1 (Mar. 2019), p. 26. ISSN: 1741-7007. DOI: 10.1186/s12915-019-0645-5. URL: https://doi.org/10.1186/s12915-019-0645-5.

[158] Casper Nyamukondiwa et al. "Thermal biology, population fluctuations and implications of temperature extremes for the management of two globally significant insect pests". In: *Journal of Insect Physiology* 59.12 (2013), pp. 1199–1211. ISSN: 0022-1910. DOI: https://doi.org/10.1016/j.jinsphys.2013.09.004. URL: https://www.sciencedirect.com/science/article/pii/S0022191013002060.

[159] Georg Oberhofer, Tobin Ivy, and Bruce A Hay. "Cleave and Rescue, a novel selfish genetic element and general strategy for gene drive". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 116.13 (Mar. 2019), pp. 6250–6259.

[160] Georg Oberhofer, Tobin Ivy, and Bruce A. Hay. "2-Locus Cleave and Rescue selfish elements harness a recombination rate-dependent generational clock for self limiting gene drive". In: *bioRxiv* (Jan. 2020), p. 2020.07.09.196253. DOI: 10.1101/2020.07.09.196253. URL: https://doi.org/10.1101/2020.07.09.196253.

[161] World Health Organization. *World malaria report 2019*. en. CC BY-NC-SA 3.0 IGO. Geneva: World Health Organization, 2019. ISBN: 9789241565721. URL: https://creativecommons.org/licenses/by-nc-sa/3.0/igo.

[162] Marcelo Otero, Hernán G. Solari, and Nicolás Schweigmann. "A Stochastic Population Dynamics Model for Aedes Aegypti: Formulation and Application to a City with Temperate Climate". In: *Bulletin of Mathematical Biology* 68.8 (Nov. 2006), pp. 1945–1974. ISSN: 1522-9602. DOI: 10.1007/s11538-006-9067-y. URL: https://doi.org/10.1007/s11538-006-9067-y.

[163] Kenneth A. Oye et al. "Regulating gene drives". In: *Science* 345.6197 (Aug. 2014), p. 626. DOI: 10.1126/science.1254287. URL: https://doi.org/10.1126/science.1254287.

[164] Daniel A. Pfeffer et al. "malariaAtlas: an R interface to global malariometric data hosted by the Malaria Atlas Project". In: *Malaria Journal* 17.1 (Oct. 2018), p. 352. ISSN: 1475-2875. DOI: 10.1186/s12936-018-2500-5. URL: https://doi.org/10.1186/s12936-018-2500-5.

[165] T. B. Pham et al. "Experimental population modification of the malaria vector mosquito, Anopheles stephensi". In: *PLoS Genet* 15.12 (Dec. 2019), e1008440.

[166] F. Port et al. "Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila". In: *Proc Natl Acad Sci U S A* 111.29 (July 2014), E2967–2976.

[167] Thomas A. A. Prowse et al. "Dodging silver bullets: good CRISPR gene-drive design is critical for eradicating exotic vertebrates". In: *Proceedings of the Royal Society B: Biological Sciences* 284.1860 (2017), p. 20170799. DOI: 10.1098/rspb.2017.0799. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rspb.2017.0799. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2017.0799.

[168] George Qian and Adam Mahdi. "Sensitivity analysis methods in the biomedical sciences". In: *Mathematical Biosciences* 323 (May 2020), p. 108306. ISSN: 0025-5564. URL: https://www.sciencedirect.com/science/article/pii/S0025556420300018.

[169] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2020. URL: https://www.R-project.org/.

[170] Scott A. Ritchie, Brian L. Montgomery, and Ary A. Hoffmann. "Novel Estimates of Aedes aegypti (Diptera: Culicidae) Population Size and Adult Survival Based on Wolbachia Releases". In: *Journal of Medical Entomology* 50.3 (May 2013), pp. 624–631. ISSN: 0022-2585. DOI: 10.1603/ME12201. URL: https://doi.org/10.1603/ME12201.

[171] Marguerite Robinson et al. "A Model for a Chikungunya Outbreak in a Rural Cambodian Setting: Implications for Disease Control in Uninfected Areas". In: *PLOS Neglected Tropical Diseases* 8.9 (Sept. 2014), e3120. DOI: 10.1371/journal.pntd.0003120. URL: https://doi.org/10.1371/journal.pntd.0003120.

[172] Perran A. Ross, Nancy M. Endersby, and Ary A. Hoffmann. "Costs of Three Wolbachia Infections on the Survival of Aedes aegypti Larvae under Starvation Conditions". In: *PLOS Neglected Tropical Diseases* 10.1 (Jan. 2016), e0004320. DOI: 10.1371/journal.pntd.0004320. URL: https://doi.org/10.1371/journal.pntd.0004320.

[173] Perran A. Ross et al. "Larval Competition Extends Developmental Time and Decreases Adult Size of wMelPop Wolbachia-Infected Aedes aegypti". English. In: *The American Society of Tropical Medicine and Hygiene* 91.1 (July 2014), pp. 198–205. DOI: 10.4269/ajtmh.13-0576. URL: https://doi.org/10.4269/ajtmh.13-0576.

[174] Ronald Ross. "SOME A PRIORI PATHOMETRIC EQUATIONS". In: *British Medical Journal* 1.2830 (Mar. 1915), p. 546. DOI: 10.1136/bmj.1.2830.546. URL: https://doi.org/10.1136/bmj.1.2830.546.

[175] Sir Ronald Ross. *The Prevention of Malaria*. en. 1910.

[176] Andrea Saltelli et al. "Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices". In: *Environmental Modelling & Software* 114 (Apr. 2019), pp. 29–39. ISSN: 1364-8152. URL: https://www.sciencedirect.com/science/article/pii/S1364815218302822.

[177] Héctor M. Sánchez C. et al. "MGDrivE: A modular simulation framework for the spread of gene drives through spatially explicit mosquito populations". In: *Methods in Ecology and Evolution* 11.2 (2020), pp. 229–239. DOI: https://doi.org/10.1111/2041-210X.13318. eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.13318. URL: https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13318.

[178] Héctor M. Sánchez C. et al. "Modeling confinement and reversibility of threshold-dependent gene drive systems in spatially-explicit Aedes aegypti populations". In: *BMC Biology* 18.1 (May 2020), p. 50. ISSN: 1741-7007. DOI: 10.1186/s12915-020-0759-9. URL: https://doi.org/10.1186/s12915-020-0759-9.

[179] Thomas L. Schmidt et al. "Fine-scale landscape genomics helps explain the slow spatial spread of Wolbachia through the Aedes aegypti population in Cairns, Australia". In: *Heredity* 120.5 (May 2018), pp. 386–395. ISSN: 1365-2540. DOI: 10.1038/s41437-017-0039-9. URL: https://doi.org/10.1038/s41437-017-0039-9.

[180] Tom L. Schmidt et al. "Local introduction and heterogeneous spatial spread of dengue-suppressing Wolbachia through an urban population of Aedes aegypti". In: *PLOS Biology* 15.5 (May 2017), e2001894. DOI: 10.1371/journal.pbio.2001894. URL: https://doi.org/10.1371/journal.pbio.2001894.

[181] A. S. Serebrovskii. "On the possibility of a new method for the control of insect pests". In: *Zool Zhurnal* 19 (1940).

[182] M. Seyfarth et al. "Five years following first detection of Anopheles stephensi (Diptera: Culicidae) in Djibouti, Horn of Africa: populations established-malaria emerging". In: *Parasitol Res* 118.3 (Mar. 2019), pp. 725–732.

[183] A. Simoni et al. "A male-biased sex-distorter gene drive for the human malaria vector Anopheles gambiae". In: *Nat Biotechnol* 38.9 (Sept. 2020), pp. 1054–1060.

[184] M.I. Simoy, M.V. Simoy, and G.A. Canziani. "The effect of temperature on the population dynamics of Aedes aegypti". In: *Ecological Modelling* 314 (2015), pp. 100–110. ISSN: 0304-3800. DOI: https://doi.org/10.1016/j.ecolmodel.2015.07.007. URL: https://www.sciencedirect.com/science/article/pii/S0304380015003130.

[185] M. E. Sinka et al. "The dominant Anopheles vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic précis". In: *Parasit Vectors* 4 (May 2011), p. 89.

[186] David L Smith, Jonathan Dushoff, and F. Ellis McKenzie. "The Risk of a Mosquito-Borne Infectionin a Heterogeneous Environment". In: *PLOS Biology* 2.11 (Oct. 2004), null. DOI: 10.1371/journal.pbio.0020368. URL: https://doi.org/10.1371/journal.pbio.0020368.

[187] Karline Soetaert, Jeff Cash, and Francesca Mazzia. "Solving ordinary differential equations in R". In: Solving Differential Equations in R. Springer, 2012, pp. 41–80.

[188] Harald Stenmark. "Rab GTPases as coordinators of vesicle traffic". In: *Nature Reviews Molecular Cell Biology* 10.8 (Aug. 2009), pp. 513–525. ISSN: 1471-0080. DOI: 10.1038/nrm2728. URL: https://doi.org/10.1038/nrm2728.

[189] Steven T Stoddard et al. "House-to-house human movement drives dengue virus transmission". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 110.3 (Jan. 2013), pp. 994–999.

[190] K. Tanaka. "The proteasome: overview of structure and functions". In: *Proc Jpn Acad Ser B Phys Biol Sci* 85.1 (2009), pp. 12–36. DOI: 10.2183/pjab.85.12. URL: https://doi.org/10.2183/pjab.85.12.

[191] C. Taylor et al. "Gene flow among populations of the malaria vector, Anopheles gambiae, in Mali, West Africa". eng. In: *Genetics* 157.2 (Feb. 2001). PMC1461521[pmcid], pp. 743–750. ISSN: 0016-6731. URL: https://pubmed.ncbi.nlm.nih.gov/11156993.

[192] Gerard Terradas et al. "Inherently confinable split-drive systems in Drosophila". In: *Nature Communications* 12.1 (Mar. 2021), p. 1480. ISSN: 2041-1723. DOI: 10.1038/s41467-021-21771-7. URL: https://doi.org/10.1038/s41467-021-21771-7.

[193] "The biology and bionomics of Aedes aegypti in the laboratory." In: *Mosquito News* 24 (1964). Author(s) - Fay, R. W., pp. 300–308. URL: https://www.biodiversitylibrary.org/part/129100.

[194] Frédéric Tripet, Guimogo Dolo, and Gregory C. Lanzaro. "Multilevel Analyses of Genetic Differentiation in Anopheles gambiae s.s. Reveal Patterns of Gene Flow Important for Malaria-Fighting Mosquito Projects". In: *Genetics* 169.1 (Jan. 2005), pp. 313–324. ISSN: 1943-2631. DOI: 10.1534/genetics.104.026534. URL: https://doi.org/10.1534/genetics.104.026534.

[195] M. Turelli and A. A. Hoffmann. "Microbe-induced cytoplasmic incompatibility as a mechanism for introducing transgenes into arthropod populations". In: *Insect Mol Biol* 8.2 (May 1999), pp. 243–255.

[196] Robert L Unckless, Andrew G Clark, and Philipp W Messer. "Evolution of Resistance Against CRISPR/Cas9 Gene Drive". In: *Genetics* 205.2 (Feb. 2017), pp. 827–841. ISSN: 1943-2631. DOI: 10.1534/genetics.116.197285. eprint: https://academic.oup.com/genetics/article-pdf/205/2/827/35553514/genetics0827.pdf. URL: https://doi.org/10.1534/genetics.116.197285.

[197] M. A. Vazquez-Cruz et al. "Global sensitivity analysis by means of EFAST and Sobol' methods and calibration of reduced state-variable TOMGRO model using genetic algorithms". In: *Computers and Electronics in Agriculture* 100 (Jan. 2014), pp. 1–12. ISSN: 0168-1699. URL: https://www.sciencedirect.com/science/article/pii/S0168169913002482.

[198] Patrick G. T. Walker et al. "Estimating the most efficient allocation of interventions to achieve reductions in ¡em¿Plasmodium falciparum¡/em¿ malaria burden and transmission in Africa: a modelling study". In: *The Lancet Global Health* 4.7 (July 2016), e474–e484. ISSN: 2214-109X. DOI: 10.1016/S2214-109X(16)30073-0. URL: https://doi.org/10.1016/S2214-109X(16)30073-0.

[199] T. Walker et al. "The wMel Wolbachia strain blocks dengue and invades caged Aedes aegypti populations". In: *Nature* 476.7361 (Aug. 2011), pp. 450–453. ISSN: 1476-4687. DOI: 10.1038/nature10355. URL: https://doi.org/10.1038/nature10355.

[200] Catherine M. Ward et al. "MEDEA SELFISH GENETIC ELEMENTS AS TOOLS FOR ALTERING TRAITS OF WILD POPULATIONS: A THEORETICAL ANALYSIS". In: *Evolution* 65.4 (Apr. 2011), pp. 1149–1162. ISSN: 0014-3820. DOI: 10.1111/j.1558-5646.2010.01186.x. URL: https://doi.org/10.1111/j.1558-5646.2010.01186.x.

[201] Helen J Wearing and Pejman Rohani. "Ecological and immunological determinants of dengue epidemics". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 103.31 (Aug. 2006), pp. 11802–11807.

[202] Michael T White et al. "Modelling the impact of vector control interventions on Anopheles gambiae population dynamics". en. In: *Parasit. Vectors* 4 (July 2011), p. 153.

[203] Darren J. Wilkinson. *Stochastic modelling for systems biology*. CRC press, 2018. ISBN: 135100090X.

[204] Nikolai Windbichler et al. "A synthetic homing endonuclease-based gene drive system in the human malaria mosquito". In: *Nature* 473.7346 (May 2011), pp. 212–215. ISSN: 1476-4687. DOI: 10.1038/nature09937. URL: https://doi.org/10.1038/nature09937.

[205] Megan R. Wise de Valdez et al. "Genetic elimination of dengue vector mosquitoes". In: *Proceedings of the National Academy of Sciences* 108.12 (2011), pp. 4772–4775. ISSN: 0027-8424. DOI: 10.1073/pnas.1019295108. eprint: https://www.pnas.org/content/108/12/4772.full.pdf. URL: https://www.pnas.org/content/108/12/4772.

[206] Sewall Wright. "On the Probability of Fixation of Reciprocal Translocations". In: *The American Naturalist* 75.761 (Nov. 1941), pp. 513–522. ISSN: 0003-0147. DOI: 10.1086/280996. URL: https://doi.org/10.1086/280996.

[207] Sean L. Wu et al. "MGDrivE 2: A simulation framework for gene drive systems incorporating seasonality and epidemiological dynamics". In: *PLOS Computational Biology* 17.5 (May 2021), e1009030. DOI: 10.1371/journal.pcbi.1009030. URL: https://doi.org/10.1371/journal.pcbi.1009030.

[208] Xiang-Ru Shannon Xu et al. "Active Genetic Neutralizing Elements for Halting or Deleting Gene Drives". en. In: *Mol. Cell* (Sept. 2020).

[209] Martin P Zeidler et al. "Temperature-sensitive control of protein activity by conditionally splicing inteins". en. In: *Nat. Biotechnol.* 22.7 (July 2004), pp. 871–876.

[210] J. Zhang et al. "PEAR: a fast and accurate Illumina Paired-End reAd mergeR". In: *Bioinformatics* 30.5 (Mar. 2014), pp. 614–620.

Appendix A

Supplement to "Trans-Complementing Gene Drive Removes Cas9 after Population Modification: A Modeling Analysis"

Parameter (unit)	**Symbol**	**Value**	**Reference**
Egg production per female (day^{-1})	β	20	[162]
Duration of egg stage (day)	T_E	2	[53]
Duration of larval stage (day)	T_L	5	[53]
Duration of pupal stage (day)	T_P	1	[53]
Population growth rate (day^{-1})	r	1.175	[184]
Mortality rate of adult stage (day^{-1})	μ_M	0.09	[55, 96, 193]

Table A.1: **Life History Parameters**

Parameter	Value	Reference
f^c	0.075	[123]
f^{Ar}, f^{Br}, f^{Cr}	0 to 1 by 0.025	
$f^{Arw}, f^{Brw}, f^{Crw}$	0.05	[129]
Cleavage rate (male and female)	1.00	[129]
HDR rate, Cas9 locus (male and female)	0.90	[129]
HDR rate, effector locus (male and female)	0.99	[129]
in-frame NHEJ rate, Cas9 locus (male and female)	0.33	[123]
in-frame NHEJ rate, effector locus (male and female)	0.33	[123]

Table A.2: **Gene Drive Implementation Parameters**
Fitness cost labels are from figure 3.6.

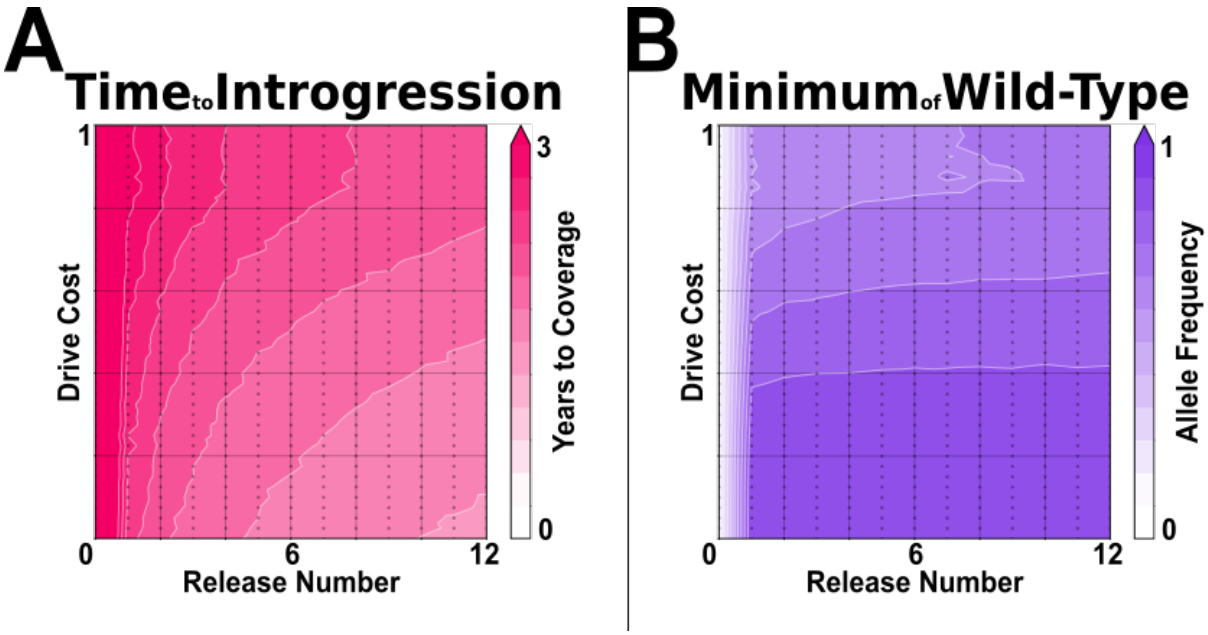

Figure A.1: **Supplement to Figure 3.7**
This is an extension of figure 3.7, exploring two additional metrics - the time until the effector carrier frequency reaches a pre-determined level, and the fraction of wild-type alleles left in the population after dropout of Cas9. All parameters are the same as panels **B** and **C** of the main-text figure.
(**A**) As complete introgression is not required for epidemiological impact, we define a 75% carrier frequency as the lowest metric for epidemiological impact, and measure the amount of time it takes for the effector gene to reach that frequency in the female population. More releases with a lower drive cost reach the threshold sooner.
(**B**) Figure 3.7a demonstrates two outcomes at the Cas9 locus - complete replacement with non-costly resistance alleles, or partial replacement with wild-type alleles remaining. This plot quantifies the maximum amount of non-wild-type alleles, in the whole population, over the simulation period, with 100% being complete replacement and 0% remaining completely wild-type. At low drive cost and a larger amount of releases, the most effective parameter range in panel **A**, there is complete replacement of the wild-type allele. However, at higher fitness cost, especially at low numbers of releases, a small fraction of wild-type alleles remain in the population after the Cas9 has dropped out.
Simulation parameters are provided in supplemental tables A.1 and A.2

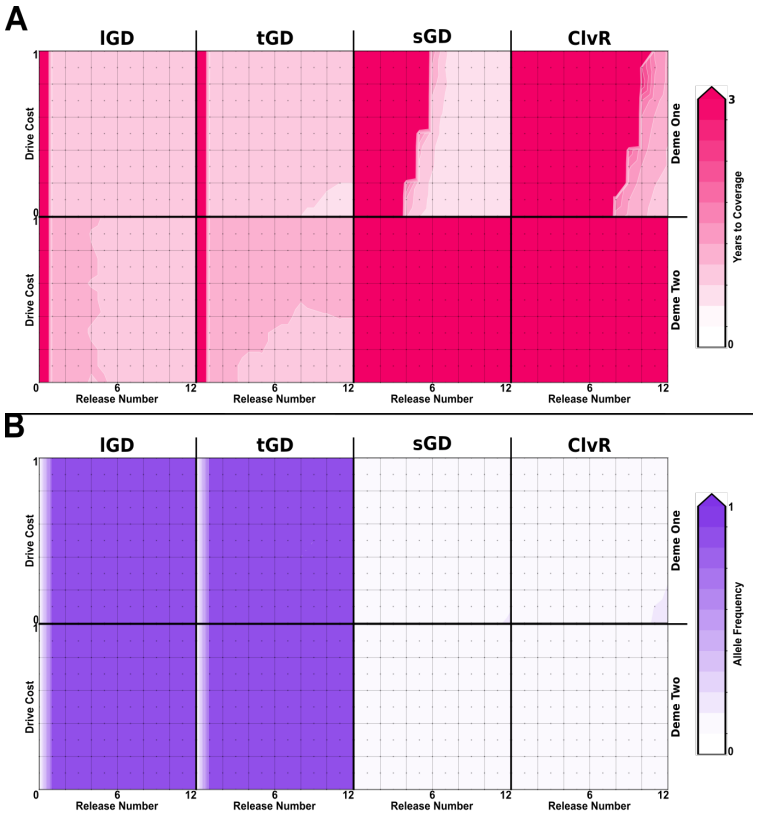

Figure A.2: **Supplement to Figure 3.8**
This is an extension of figure 3.8 and similar to figure A.1. It explores the same two added metrics - time to threshold and fraction of wild-type alleles remaining - while comparing the four designs described in figure 3.6 in a two deme setup. All parameters are the same as panels **B** and **C** of the main-text figure.
(**A**) This panel compares our four designs for the rate at which they reach 75% carrier frequency in the female population. Linked and trans-complementing gene drives are extremely fast, but also spread through both populations. Split drive and ClvR are much slower at small numbers of releases, but never reach the threshold in the second population.
(**B**) Similar to figure A.1b, we measure the maximum amount of non-wild-type alleles in the populations, for all four drives and in both populations. The results for tGD differ from the panmictic case (Fig. A.1b), potentially due to the added mixing from two populations. The tGD system behaves most similarly to a linked system here, while split drive and ClvR, neither of which drive at the Cas9 locus, remain at low allele frequencies in all conditions.
Simulation parameters are provided in supplemental tables A.1 and A.2