MGDrivE: A modular simulation framework for the spread of gene drives through spatially-explicit mosquito populations

Héctor M. Sánchez C.[1*], Sean L. Wu[1] ‡, Jared B. Bennett[2] ‡, John M. Marshall[1*]

1 Division of Biostatistics and Epidemiology, School of Public Health, University of California, Berkeley, California, United States
2 Biophysics Graduate Group, Division of Biological Sciences, College of Letters and Science, University of California, Berkeley, California, United States

‡These authors contributed equally to this work.
* sanchez.hmsc@berkeley.edu, john.marshall@berkeley.edu

Abstract

Malaria, dengue, Zika, and other mosquito-borne diseases continue to pose a major global health burden through much of the world, despite the widespread distribution of insecticide-based tools and antimalarial drugs. The advent of CRISPR/Cas9-based gene editing and its demonstrated ability to streamline the development of gene drive systems has reignited interest in the application of this technology to the control of mosquitoes and the diseases they transmit. The versatility of this technology has also enabled a wide range of gene drive architectures to be realized, creating a need for their population-level and spatial dynamics to be explored. To this end, we present MGDrivE (Mosquito Gene Drive Explorer): a simulation framework designed to investigate the population dynamics of a variety of gene drive architectures and their spread through spatially-explicit mosquito populations. A key strength of the MGDrivE framework is its modularity: a) a genetic inheritance module accommodates the dynamics of gene drive systems displaying user-defined inheritance patterns, b) a population dynamic module accommodates the life history of a variety of mosquito disease vectors and insect agricultural pest species, and c) a landscape module accommodates the distribution of insect metapopulations connected by migration in space. Example MGDrivE simulations are presented to demonstrate the application of the framework to CRISPR/Cas9-based homing gene drive 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 in both cases. We compare MGDrivE with other genetic simulation packages, and conclude with a discussion of future directions in gene drive modeling.

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 [1–3]. 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. For malaria, recent declines in transmission have been seen following the wide-scale

distribution of bed nets and antimalarial drugs [4]; however, model-based projections suggest that additional tools will be required to eliminate the disease from highly-endemic areas [5]. For dengue, the need for novel vector control strategies is even greater, as the disease is rising in global prevalence and there is currently no cure or vaccine available that is effective against all four serotypes [6]. The recent demonstration of a CRISPR-based gene drive system in Drosophila [7], followed months later by a Zika outbreak in Brazil [8], has prompted development of gene drive technology for *Aedes aegypti*, the primary mosquito vector of Zika, dengue, and Chikungunya, as well as broad development targeting other mosquito species, such as the Anophelines which transmit malaria.

The ease of gene editing afforded by the discovery of CRISPR has also led to significant versatility in terms of the gene drive systems that are now realizable [3, 9]. Prior to the advent of CRISPR, homing endonuclease genes (HEGs) were envisioned to cleave a specific target site lacking the HEG and to be copied to this site by serving as a template for homology-directed repair (HDR), effectively converting a heterozygote into a homozygote and biasing inheritance in favor of the HEG [10]. These dynamics have been demonstrated for a HEG targeting a synthetic target site in the main African malaria vector, *Anopheles gambiae* [11], and steps have also been taken towards engineering an alternative approach in which the HEG is located on the Y chromosome and cleaves the X chromosome in multiple locations, biasing inheritance in its favor as it induces an increasingly male sex bias in the population [12]. A vast range of additional approaches for biasing inheritance are now being proposed, including several threshold-dependent systems that may permit confineable and reversible releases [13–15], and remediation systems that could be used to remove effector genes and possibly entire drive systems from the environment in the event of unwanted consequences [16]. For instance, an ERACR system (Element for the Reversal of the Autocatalytic Chain Reaction) has been proposed that consists of a homing system with a target site corresponding to the original drive system, essentially removing the original drive as it homes into it, and utilizing the Cas9 of the first drive thus also removing this through the homing process [17, 18].

Understanding how these systems are expected to behave in real ecosystems requires a flexible modeling framework that can accommodate a range of inheritance patterns, specific details of the species into which the constructs are to be introduced, and details of the landscape through which spatial spread would occur. 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 mosquito species and other insect species. A key strength of the MGDrivE framework is its modularity. A genetic inheritance module allows the inheritance dynamics of a wide variety of drive systems to be accommodated. An independent population dynamic module allows the life history of a variety of mosquito disease vectors and insect agricultural pests to be accommodated. Thirdly, a landscape module accommodates the distribution of insect metapopulations in space, with movement through the resulting network determined by dispersal kernels. The model can be run in either a deterministic or stochastic form, allowing the chance events that occur at low population or genotype frequencies to be simulated.

What separates MGDrivE from other gene drive modeling frameworks is its ability to simulate a wide array of user-specified inheritance-modifying systems at the population level within a single, computationally efficient framework that also incorporates mosquito life history and landscape ecology. Other frameworks exist that have been designed for more general purposes and applied to specific questions related to gene drive (Table 1) – for instance, Eckhoff *et al.* [19] used the EMOD malaria model to simulate the spread of homing-based gene drive systems through spatial populations

of *An. gambiae*. EMOD is open source and a powerful modeling framework; but significant effort is required from users to redefine genetic control strategies, mosquito life history parameters and landscape details. Magori *et al.* [20] created Skeeter Buster by extending the CIMSiM (container-inhabiting mosquitoes simulation model) model [21] to incorporate genetic inheritance and spatial structure. The Skeeter Buster framework captures the most pertinent mosquito ecology considerations, but is not open source and can only simulate a handful of genetic control strategies [22]. The SLiM genetic simulation framework [23] is capable of modeling the spread of a large variety of user-defined gene drive systems through metapopulations; however, it is not currently capable of accommodating life history ecology and overlapping generations.

Table 1. Comparison of spatially-explicit gene drive models.

	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	Metapopulations distributed in space, connected by migration	R package, open source
EMOD [19]	Homing-based gene drive, could be extended with effort	Egg-larva-pupa-adult, density-dependence at larval stage, responsive to environmental variables	Populations arranged on a grid, each representing 1 km^2, connected by migration	Java Script Open Notation (JSON) feeds into executable file, open source
Skeeter Buster [22]	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 [23]	Very flexible, can be user-specified	Discrete generations, no life history at present	Can model either connected metapopulations or cells on a grid	Scripting environment with graphical user interface, open source

In this paper, we describe the key components of the MGDrivE framework – namely, the genetic inheritance, mosquito life history and landscape/metapopulation modules. We then provide a demonstration of 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 modeling.

Design and Implementation

The MGDrivE framework is a genetic and spatial extension of the lumped age-class model of mosquito ecology [24] modified and applied by Deredec *et al.* [25] to the spread of homing gene drive systems, and by Marshall *et al.* [26] 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 then accommodated through a metapopulation structure in which lumped age-class models run in parallel and migrants are exchanged between metapopulations at defined rates. These operations are accommodated by the tensor modeling framework, and full details of this framework are provided in the S1 User Manual.

The core simulation framework is being developed in R (`https://www.r-project.org/`) with certain routines in Rcpp for computational speed. By combining the tensor modeling 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. Within this architecture, each module may be conveniently altered independently of the others. For instance: a) a range of gene drive systems may be explored for a given mosquito species in a given landscape, b) one species may be substituted for another, provided its sequence of life history events is comparable, and c) gene drive spread may be modeled through a range of landscapes, while leaving the rest of the model untouched. We now describe the three distinct modules of the MGDrivE framework – inheritance, life history and spatial structure – in more detail.

Modules

1. Genetic Inheritance The fundamental module for modeling gene drive dynamics is that describing genetic inheritance. In MGDrivE, this is embodied by a three-dimensional tensor referred to as an "inheritance cube" (Figure 1). Each gene drive system has a unique R file containing the three-dimensional inheritance cube. The first and second dimensions of the inheritance cube refer to the maternal and paternal genotypes, respectively, and the third dimension refers to the offspring genotype. The 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 a homing-based drive system, for instance, 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 [26–28]. Input parameters also include those associated with organisms having each genotype – for instance: a) genotype-specific fertility rates, b) male mating fitness, c) sex bias at emergence, d) adult survival rates, and e) male and female pupatory success. These parameters feed into the mosquito life history module, that will be described next, and modify the tensor equations in that module in order to produce the desired biological effect. Finally, a "viability mask" is applied to the offspring genotypes to remove unviable genotypes from the population.

The R function that builds the inheritance cube may receive a number of user-defined input parameters. For a homing-based drive system, for instance, 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 [26–28]. and these parameter values should be used to populate the entries of the inheritance cube. Input parameters also include those associated with organisms having each genotype – for instance, genotype-specific: a) fertility rates, b) male mating fitness, c) sex bias at emergence, d) adult survival rates, and e) male and female pupatory success. These parameter values should then be used to populate the entries

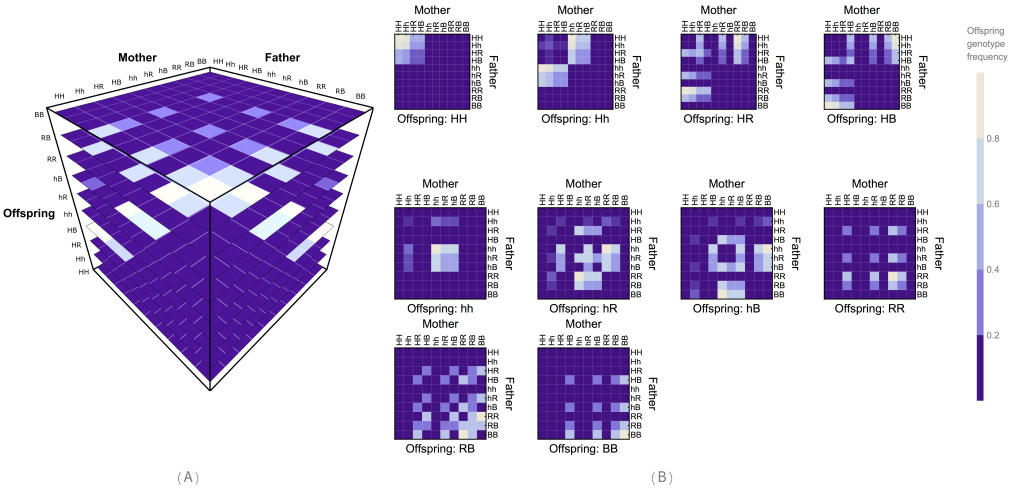

Fig 1. Inheritance module. Genetic inheritance is embodied by a three-dimensional tensor referred to as an inheritance cube (left), here depicted for a CRISPR/Cas9-based homing construct. Maternal and paternal genotypes are depicted on the x and y-axes and offspring genotypes on the z-axis, with slices 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). Offspring genotype frequencies for each parental cross are depicted according to the shading scale (right).

of the inheritance cube. Input parameters also include those associated with organisms having each genotype – for instance, genotype-specific: a) fertility rates, b) male mating fitness, c) sex bias at emergence, d) adult survival rates, and e) male and female pupatory success. These parameters feed into the mosquito life history module, that will be described next, and modify the tensor equations in that module in order to produce the desired biological effect. 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 [26, 27, 29, 30], c) *Medea* (a maternal toxin linked to a zygotic antidote) [31], d) other toxin-antidote-based underdominant systems such as UD^{MEL} [13, 15, 32], e) reciprocal chromosomal translocations [14, 33], f) *Wolbachia* [34], and g) the RIDL system [35] (release of insects carrying a dominant lethal gene). Details of each of these systems are provided in the S1 User Manual.

2. Mosquito Life History The mosquito life history module follows from the lumped age-class model of Hancock and Godfray [24] adapted by Deredec *et al.* [25]. In this model (depicted in Figure 2), the insect life cycle is divided into four stages – egg (E), larva (L), pupa (P) and adult (M for male and F for female). 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 genetic inheritance module. 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 in turn

determines the degree of additional density-dependent mortality at the larval stage in that patch. Following Deredec *et al.* [25], this is described by an equation of the form: $f(L) = \alpha/(\alpha + L)^{1/T_L}$, 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 the S1 User Manual.

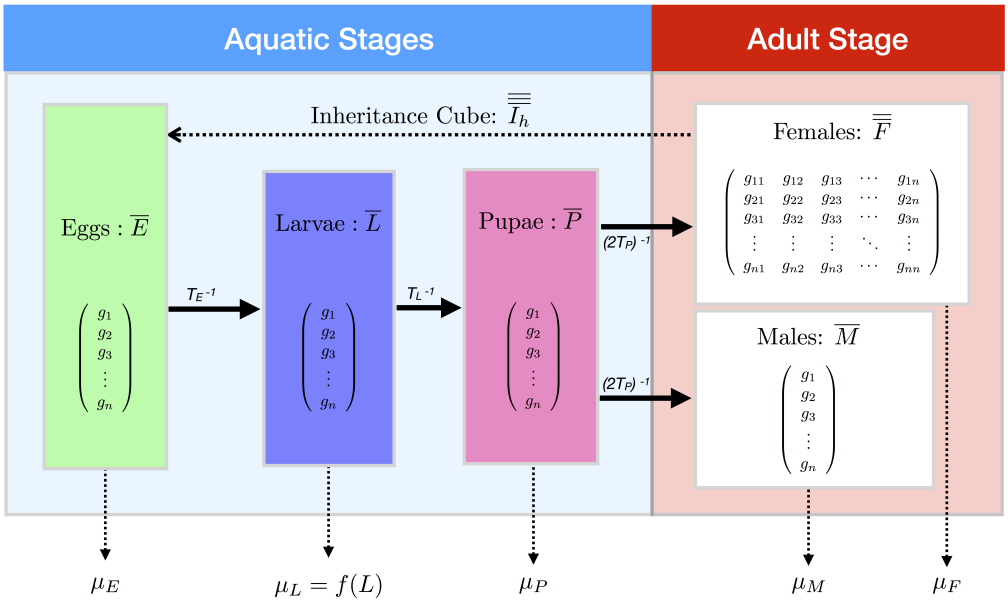

Fig 2. Mosquito life history module. Life history is modeled according to an egg (E)-larva (L)-pupa (P)-adult (M for male, F for female) life cycle in which density dependence occurs at the larval stage and autonomous mobility occurs at the adult stage. Genotypes are tracked across all life stages, where g_1, ..., g_n represent the number of individuals at each life stage having each of n genotypes. Females are modeled as mating once upon emergence and hence 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.

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 the parameter values that describe this module when initializing an MGDrivE simulation. The input variables for this module currently include: 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 2 for three species of interest: a) *An. gambiae*, the main African malaria vector, b) *Ae. aegypti*, the main vector of dengue and Zika virus, and c) *Ceratitis capitata*, a worldwide agricultural crop pest. In some cases, life history parameters will be modified in genotype-specific ways by the gene drive construct, and such modifications are efficiently accommodated within this framework via tensor operations.

Parameter	*Aedes aegypti*	*Anopheles gambiae*	*Ceratitis capitata*
Egg production per female (day^{-1})	20 [36]	32 [37]	20 [38]
Duration of egg stage (days)	5 [39]	1 [37]	2 [38]
Duration of larval stage (days)	6 [39]	13 [37]	6 [38]
Duration of pupa stage (days)	4 [39]	1 [37]	10 [38]
Daily population growth rate (day^{-1})	1.175 [40]	1.096 [41]	1.031 [42]
Daily mortality risk of adult stage (day^{-1})	0.090 [43–45]	0.123 [41]	0.100 [46]

Table 2. Life history module parameter values for three species of interest (at a temperature of 25 Celsius).

3. Landscape The landscape module describes the distribution of mosquito metapopulations in space, with movement through the resulting network determined by dispersal kernels. Metapopulations are randomly mixing populations for which the equations of the lumped age-class model apply. The resolution of the metapopulations (in terms of size) should be chosen according to the dispersal properties of the insect species of interest and the research question being investigated. *Ae. aegypti* mosquitoes, for instance, are thought to be relatively local dispersers, often remaining in the same household for the duration of their lifespan [47]. For modeling the fine-scale spread of gene drive systems in this species, metapopulations the size of households may be appropriate. *An. gambiae* mosquitoes, on the other hand, are thought to display moderate dispersal on the village scale and infrequent inter-village movement [48]. This would suggest villages as an appropriate metapopulation unit; however other levels of aggregation are also possible, in both cases, depending on the level of resolution required from the simulations and the computational power available to the user.

Once the metapopulation size has been decided upon and the metapopulations have been enumerated, MGDrivE accepts a list of coordinates and equilibrium adult population sizes associated with each. In the resulting network structure, nodes represent randomly-mixing metapopulations and edges represent movement of mosquitoes from one metapopulation to any other in the network (Figure 3). Movement between metapopulations is limited to the adult life stage. By default, movement rates between metapopulations are derived from a zero-inflated exponential dispersal kernel, the degree of zero-inflation and mean dispersal distance of which may be defined by the user. That said; the movement kernel may be expanded arbitrarily to account for barriers to movement such as roads [47] and other factors without altering the overarching model structure. Movement rates between nodes are then used to calculate a matrix of node transition probabilities, which is incorporated in the tensor algebraic model formulation described in the S1 User Manual.

Finally, with the inheritance, life history and landscape modules in place, any type of release can be simulated by increasing the number of insects having the released sex and genotype at a specific metapopulation and time. As demonstrated in the following software use example, input variables are provided for: a) release size, b) number of releases, c) time of first release, d) time between releases, e) metapopulation of release, and f) sex and genotype of released insects.

Deterministic vs. Stochastic Simulations

Simulations in MGDrivE 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. For instance, a stochastic model is required to understand the chance of population elimination following releases of insects carrying a population-suppressing homing system in the context of rarely generated

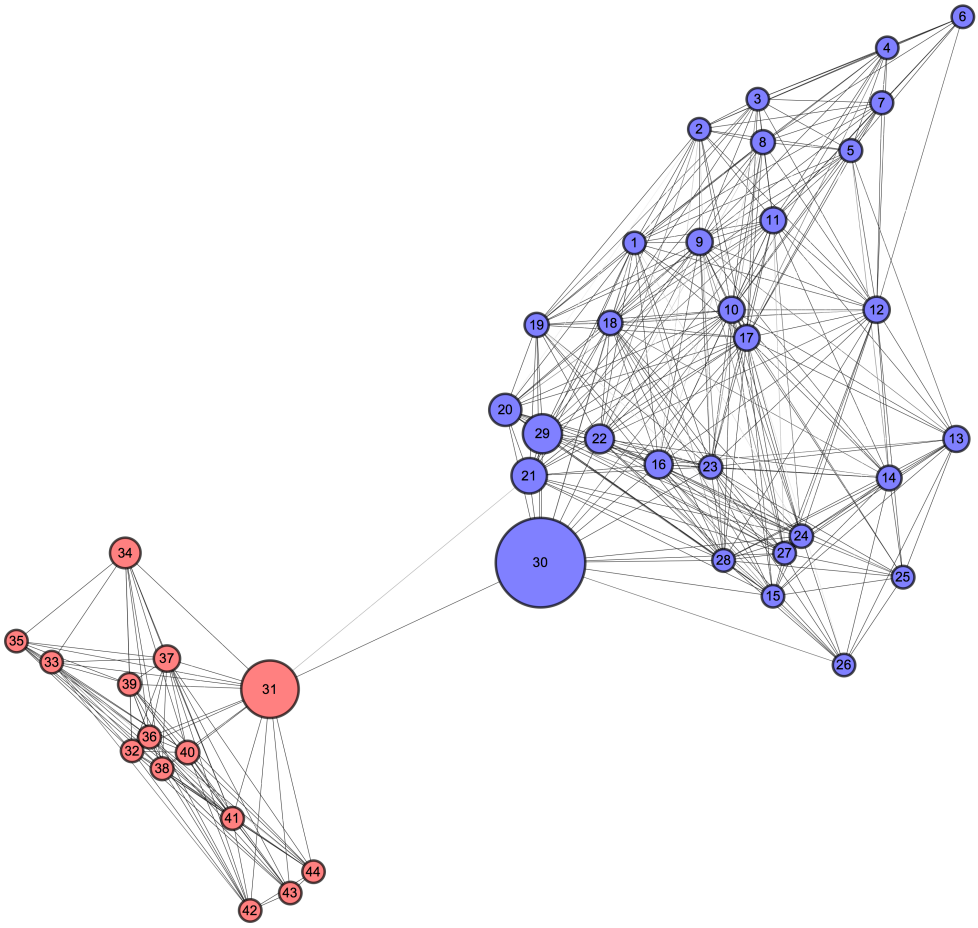

Fig 3. Landscape module. Insects are distributed as metapopulations, here depicted by nodes, each having their own coordinates and population size. Movement between metapopulations is derived from a defined dispersal kernel and is depicted here by edges between nodes. The example "tale of two cities" scenario allows both spread within and between communities to be explored. Here, nodes are colored according to their community (as detected by the DBSCAN clustering algorithm [49]), with sizes proportional to their "betweenness centrality" - a measure of their connectedness to other nodes in terms of number of shortest paths that flow through them [50].

resistant alleles [26]. 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 pattern of the gene drive system, 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.

Two Example MGDrivE Simulations

To demonstrate how the MGDrivE framework can be used to initialize and run a simulation of a gene drive system through a network of connected metapopulations, we describe the application of the package to two CRISPR/Cas9-based homing gene drive

strategies: a) driving a disease-refractory gene into a population [7], and b) disrupting a gene required for female fertility and hence suppressing a population [30]. In both cases, we consider a population of *Ae. aegypti* mosquitoes having the bionomic parameters provided in Table 2 and distributed through the network landscape depicted in Figure 3. To demonstrate the functionality of the MGDrivE package, we model the population replacement strategy (i.e. replacing the population with a disease-refractory one) using the deterministic implementation, and model the population suppression strategy using the stochastic implementation. The stochastic implementation is more relevant to population suppression as it can capture rare resistant allele generation and the possibility of population extinction. In both cases, we include the generation of in-frame and out-of-frame or otherwise costly resistant alleles [28, 51] and parameterize the gene drive model based on recently engineered constructs [7, 30].

1. Population Replacement We begin by modeling a CRISPR/Cas9-based homing construct similar to that engineered by Gantz et al. [7]. 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 [51]. Female fecundity and male mating fitness are reduced by 25% per H or R allele and by 50% per B allele.

Fig 4. Workflow of an MGDrivE simulation.

The code for this simulation (Code samples 1-3) can be entered directly in R, and the details of the various functions used are described in the S1 User Manual. The

general workflow for the simulation is shown in Figure 4. We begin by loading the MGDrivE package in R and choosing the working and output directories. The output directory should be a dedicated directory for MGDrivE simulation output, to avoid interfering with other files. We then choose between the deterministic and stochastic implementation of the model framework – in this case the deterministic version. Next, we specify the bionomic parameters of the species we are modeling – in this case, *Ae. aegypti*, whose default life history parameters are provided in Table 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 metapopulations in Figure 3. A function is then applied that computes daily movement rates between each of the metapopulations 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 metapopulations; however in this case, we assume these are identical across all metapopulations and provide a single population size (Code sample 1).

```
# 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=calc_HurdleExpKernel(
  distancesMatrix,
  MGDrivE::kernels$exp_rat,
  calculateZeroInflation(
    lifespanNonMigratoryProbability,
    bioParameters$muAd
  )
)
patchPops=rep(50,sitesNumber)
```

Code sample 1. Loading the package and setting up the life history and landscape modules.

With our life history and landscape modules defined and parameterized, we now specify the gene drive system and release strategy we intend to model (Code sample 2). We use a pre-specified inheritance cube function, "Cube_HomingDrive()", that models the inheritance pattern of a homing-based gene drive system. The input options for this function can be seen by typing "?Cube_HomingDrive()" at the command prompt. We specify the sex-specific homing rates, resistant allele generation rates, and genotype-specific fitness effects as described earlier based on the construct engineered

by Gantz et al. [7]. 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 then incorporated into a vector also specifying the inheritance cube and the sex and genotype of the released insects. Finally, the metapopulations in which the release takes place are specified. With the simulation framework now fully specified, the model is now ready to run (Code sample 3).

```
# SET UP INHERITANCE / GENE DRIVE #################
### A. Replacement Drive
sH=sR=.25
sB=.50
eM=0.9
eF=0.5
driveCube=Cube_HomingDrive(
  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
eM=eF=0.999
driveCube=Cube_HomingDrive(
  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}
```

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=Network.Parameters(
    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()
}
```

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.

2. Population Suppression As a second example, we demonstrate the application of the MGDrivE package to model a population suppression homing construct similar to that engineered by Hammond et al. [30]. 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 *An. gambiae*, although with a higher homing efficiency that could be achieved through guide RNA multiplexing [26]. Lines of code that differ for this system are shown in Code sample 2. We choose the stochastic implementation of the model framework this time, and while the same inheritance cube function applies, it's parameters differ – namely, homing and resistant allele generation rates, and genotype-specific fitness effects.

Output Analysis

In the current version of MGDrivE, simulation results are output as CSV files, which enables the user to analyze results in any platform of their choice – R, Python, Mathematica, etc. What the user decides to plot will depend on the number of possible genotypes, whether the male-to-female ratio is altered, whether the population is suppressed, and the spatial structure of the landscape through which drive occurs. If the number of genotypes is large, for instance, then allele abundance may provide a more manageable output than that of genotypes.

In Figure 5, we display a potential visualization scheme produced in Mathematica for the population replacement and suppression simulations described above (additionally, videos for both simulations running in the spatial networks can be accessed in the supplementary information: S1 Video and S2 Video). As there are four alleles for both systems (the homing allele, H, the wild-type allele, h, and the two resistant alleles, R and B), we depict their abundance in the Figures 5A and 5B and their frequency in Figures 5C and 5D, with time on the horizontal axis and metapopulation number on the vertical axis. For population replacement (Figures 5A and 5C), we see the gene drive system (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. For population suppression (Figures 5B and 5D), 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. Also visible in Figure 5 is the slightly extended time it takes for both homing systems to spread through the second population cluster visible in the metapopulation landscape depicted in Figure 3.

Availability and Future Directions

As of the date of publication, we are releasing MGDrivE version 1.0 ("Rise and Shine"), available at our permanent github repository at:
`https://github.com/MarshallLab/MGDrivE`. The source code is available under the GPL3 License and free for other groups to modify and extend as needed. The S1 User Manual, including documentation of all MGDrivE functions and mathematical details of the model formulation are available at the project's github repository. To run the software, we recommend using R version 3.4.4 or higher.

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 extremely quickly, especially since the discovery of CRISPR-based gene editing, and we intend the MGDrivE package to provide a flexible tool capable of modeling novel inheritance-modifying constructs as they are proposed and become available. Future functionality that we intend to incorporate into the software includes: a) "shadow drive", in which the Cas9 enzyme is passed on to the offspring even if the gene expressing it is not [51], b) life history models encompassing a more diverse range of insect disease vectors and agricultural pests, and c) populations that vary in size periodically or in response to environmental input variables. We are also developing a corresponding individual-based model that is capable of modeling 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 [27], and multiplexing schemes in

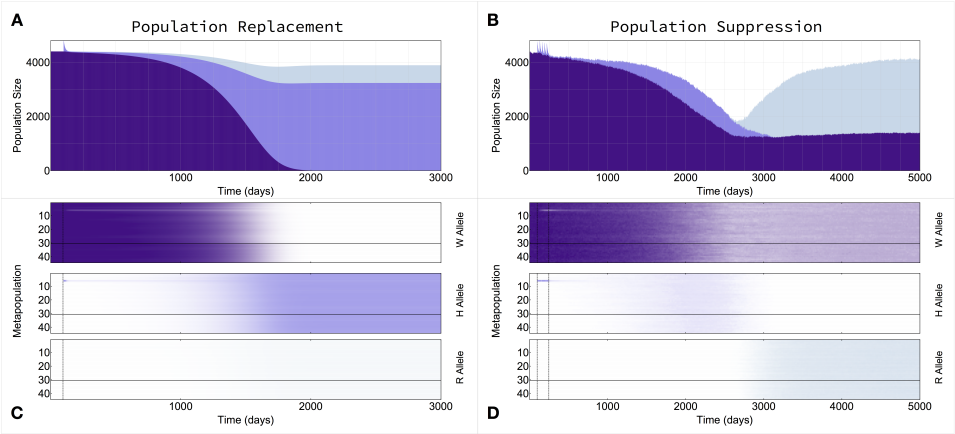

Fig 5. Example MGDrivE simulations for CRISPR-based homing constructs. In both cases, an Aedes aegypti population is simulated having the bionomic parameters in Table 2 and distributed through the landscape depicted in Figure 3. A. A population replacement homing construct that drives a disease-refractory 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 (light purple) is driven into the population, the wild-type allele (dark purple) is eliminated, and the in-frame resistant allele (light blue) 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 y HH females at node 6 are modeled. As the homing allele (light purple) is driven into the population, it suppresses the population due to its impact on female fertility. Eventually, an in-frame resistant allele (light blue) emerges and leads the population to rebound due to its selective advantage over both wild-type and homing alleles. C-D. Here, population frequencies of the wild-type, homing and in-frame resistant alleles are shown in each metapopulation over time for the population replacement construct (panel C) and 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 and solid horizontal lines represent the division between population clusters.

which a single gene is targeted at multiple locations with separate guide RNAs to reduce the rate of resistant allele formation [52].

Supporting information

S1 Video. Population replacement use example. A visualization of the homing-based population replacement simulation.

S2 Video. Population suppression use example. A visualization of the

homing-based population suppression simulation.

S1 User Manual. MGDrivE's version 1.0 documentation. This documentation covers the functionality, equations, and limitations of the package up to the release of the version 1.0 of the software. For the most current version, visit our website: `https://marshalllab.github.io/MGDrivE/`

Acknowledgments

The authors would like to thanks Drs. Omar Akbari, Ethan Bier and Anthony James for discussions on gene drive architectures and molecular biological considerations, and Drs. Gregory Lanzaro, Yoosook Lee and Gordana Rašić and Ms. Partow Imani for discussions on mosquito ecology, life history and dispersal behavior. This work was supported by a DARPA Safe Genes Program Grant (HR0011-17-2-0047) awarded to JMM and funds from the UC Irvine Malaria Initiative and Innovative Genomics Institute, UC Berkeley/UCSF awarded to JMM.

References

1. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. Science. 2014;346(6213). doi:10.1126/science.1258096.
2. Esvelt KM, Smidler AL, Catteruccia F, Church GM. Concerning RNA-guided gene drives for the alteration of wild populations. eLife. 2014;3:e03401. doi:10.7554/eLife.03401.
3. Champer J, Buchman A, Akbari OS. Cheating evolution: engineering gene drives to manipulate the fate of wild populations. Nature Reviews Genetics. 2016;17(3):146–159. doi:10.1038/nrg.2015.34.
4. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. Nature. 2015;526(7572):207–211. doi:10.1038/nature15535.
5. Walker PGT, Griffin JT, Ferguson NM, Ghani AC. Estimating the most efficient allocation of interventions to achieve reductions in Plasmodium falciparum malaria burden and transmission in Africa: A modelling study. The Lancet Global Health. 2016;4(7):e474–e484. doi:10.1016/S2214-109X(16)30073-0.
6. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature. 2013;496(7446):504–7. doi:10.1038/nature12060.
7. Gantz VM, Jasinskiene N, Tatarenkova O, Fazekas A, Macias VM, Bier E, et al. Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. Proceedings of the National Academy of Sciences. 2015;112(49):E6736–E6743. doi:10.1073/pnas.1521077112.
8. Lessler J, Chaisson LH, Kucirka LM, Bi Q, Grantz K, Salje H, et al. Assessing the global threat from Zika virus. Science. 2016;353(6300):aaf8160–aaf8160. doi:10.1126/science.aaf8160.
9. Marshall JM, Akbari OS. Can CRISPR-Based Gene Drive Be Confined in the Wild? A Question for Molecular and Population Biology. ACS Chemical Biology. 2018;13(2):424–430. doi:10.1021/acschembio.7b00923.

10. Burt A. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. Proceedings of the Royal Society B: Biological Sciences. 2003;270(1518):921–928. doi:10.1098/rspb.2002.2319.

11. Windbichler N, Menichelli M, Papathanos PA, Thyme SB, Li H, Ulge UY, et al. A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. Nature. 2011;473(7346):212–215. doi:10.1038/nature09937.

12. Galizi R, Doyle LA, Menichelli M, Bernardini F, Deredec A, Burt A, et al. A synthetic sex ratio distortion system for the control of the human malaria mosquito. Nature Communications. 2014;5:1–8. doi:10.1038/ncomms4977.

13. Akbari OS, Matzen KD, Marshall JM, Huang H, Ward CM, Hay Ba. A synthetic gene drive system for local, reversible modification and suppression of insect populations. Current biology. 2013;23(8):671–7. doi:10.1016/j.cub.2013.02.059.

14. Buchman AB, Ivy T, Marshall JM, Akbari OS, Hay BA. Engineered Reciprocal Chromosome Translocations Drive High Threshold, Reversible Population Replacement in Drosophila. ACS Synthetic Biology. 2018; p. acssynbio.7b00451. doi:10.1021/acssynbio.7b00451.

15. Marshall JM, Hay BA. General principles of single-construct chromosomal gene drive. Evolution. 2012;66(7):2150–2166. doi:10.1111/j.1558-5646.2012.01582.x.

16. Marshall JM, Akbari OS. Can CRISPR-Based Gene Drive Be Confined in the Wild? A Question for Molecular and Population Biology. ACS Chemical Biology. 2018;13(2):424–430. doi:10.1021/acschembio.7b00923.

17. Gantz V, Bier E. The mutagenic chain reaction: A method for converting heterzygous to homozygous mutations. Human molecular genetics. 2014;348(March):1–7. doi:10.1093/hmg/ddu125.

18. Vella MR, Gunning CE, Lloyd AL, Gould F. Evaluating strategies for reversing CRISPR-Cas9 gene drives. Scientific Reports. 2017;7(1):1–8. doi:10.1038/s41598-017-10633-2.

19. Eckhoff Pa. A malaria transmission-directed model of mosquito life cycle and ecology. Malaria Journal. 2011;10(1):303. doi:10.1186/1475-2875-10-303.

20. Magori K, Legros M, Puente ME, Focks DA, Scott TW, Lloyd AL, et al. Skeeter Buster: a stochastic, spatially explicit modeling tool for studying Aedes aegypti population replacement and population suppression strategies. PLoS neglected tropical diseases. 2009;3(9):e508. doi:10.1371/journal.pntd.0000508.

21. Focks DA, Daniels E, Haile DG, Keesling JE. A simulation model of the epidemiology of urban dengue fever: literature analysis, model development, preliminary validation, and samples of simulation results. The American journal of tropical medicine and hygiene. 1995;53(5):489–506.

22. Legros M, Xu C, Okamoto K, Scott TW, Morrison AC, Lloyd AL, et al. Assessing the Feasibility of Controlling Aedes aegypti with Transgenic Methods: A Model-Based Evaluation. PLoS ONE. 2012;7(12). doi:10.1371/journal.pone.0052235.

23. Haller BC, Messer PW. SLiM 2: Flexible, interactive forward genetic simulations. Molecular Biology and Evolution. 2017;34(1):230–240. doi:10.1093/molbev/msw211.

24. Hancock PA, Godfray HCJ. Application of the lumped age-class technique to studying the dynamics of malaria-mosquito-human interactions. Malaria journal. 2007;6:98. doi:10.1186/1475-2875-6-98.

25. Deredec A, Godfray HCJ, Burt A. Requirements for effective malaria control with homing endonuclease genes. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(43):E874–80. doi:10.1073/pnas.1110717108.

26. Marshall J, Buchman A, C HMS, Akbari OS. Overcoming evolved resistance to population-suppressing homing-based gene drives. Nature Scientific Reports. 2017;(June 2017):1–46. doi:https://doi.org/10.1101/088427.

27. Noble C, Olejarz J, Esvelt KM, Church GM, Nowak MA. Evolutionary dynamics of CRISPR gene drives. Science Advances. 2017;3(4):3–10. doi:10.1126/sciadv.1601964.

28. Unckless RL, Clark AG, Messer PW. Evolution of resistance against CRISPR/Cas9 gene drive. Genetics. 2017;205(2):827–841. doi:10.1534/genetics.116.197285.

29. Deredec A, Burt A, Godfray HCJ. The population genetics of using homing endonuclease genes in vector and pest management. Genetics. 2008;179(4):2013–26. doi:10.1534/genetics.108.089037.

30. Hammond A, Galizi R, Kyrou K, Simoni A, Siniscalchi C, Katsanos D, et al. A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. Nature Biotechnology. 2016;34(1):78–83. doi:10.1038/nbt.3439.

31. Chen Ch, Schaeffer LV, Guo M, Hay Ba. Replacement in Drosophila. Science. 2007;597(2007):597–600. doi:10.1126/science.

32. Marshall JM, Pittman GW, Buchman AB, Hay Ba. Semele: a killer-male, rescue-female system for suppression and replacement of insect disease vector populations. Genetics. 2011;187(2):535–51. doi:10.1534/genetics.110.124479.

33. Curtis CF. Possible use of translocations to fix desirable genes in insect pest populations. Nature. 1968;218(5139):368–369. doi:10.1038/218368a0.

34. Hancock PA, Sinkins SP, Godfray HCJ. Population Dynamic Models of the Spread of Wolbachia. The American Naturalist. 2011;177(3):323–333. doi:10.1086/658121.

35. Wise de Valdez R M, Nimmo D, Betz J, Gong HF, James AA, Alphey L, et al. Genetic elimination of dengue vector mosquitoes. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(12):4772–5. doi:10.1073/pnas.1019295108.

36. Otero M, Solari HG, Schweigmann N. A stochastic population dynamics model for Aedes aegypti: formulation and application to a city with temperate climate. Bulletin of mathematical biology. 2006;68(8):1945–74. doi:10.1007/s11538-006-9067-y.

37. Depinay JMO, Mbogo CM, Killeen G, Knols B, Beier J, Carlson J, et al. A simulation model of African Anopheles ecology and population dynamics for the analysis of malaria transmission. Malaria journal. 2004;3:29. doi:10.1186/1475-2875-3-29.

38. Diamantidis AD, Carey JR, Nakas CT, Papadopoulos NT. Population-specific demography and invasion potential in medfly. Ecology and Evolution. 2011;1(4):479–488. doi:10.1002/ece3.33.

39. Christophers SR. Aedes aegypti the Yellow Fever Mosquito. 1st ed. London: Cambridge University Press; 1960.

40. Simoy MI, Simoy MV, Canziani GA. The effect of temperature on the population dynamics of Aedes aegypti. Ecological Modelling. 2015;314:100–110. doi:10.1016/j.ecolmodel.2015.07.007.

41. Molineaux L, Gramiccia G. The Garki Project. Research on the epidemiology and control of malaria in the Sudan Savanna of West Africa; 1980.

42. Carey JR, Liedo P, Vaupel JW. Mortality dynamics of density in the mediterranean fruit fly. Experimental Gerontology. 1995;30(6):605–629. doi:10.1016/0531-5565(95)00013-5.

43. Focks DA, Haile DG, Daniels E, Mount GA. Dynamic life table model for Aedes aegypti (Diptera: Culicidae): analysis of the literature and model development. Journal of Medical Entomology. 1993;30(6):1003–17. doi:10.1093/jmedent/30.6.1003.

44. Horsfall WE. Mosquitoes. Their Bionomics and Relation to Disease. The Eonald Press Co.; 1955.

45. Fay RW. The biology and bionomics of Aedes aegypti in the laboratory. Mosquito News. 1964;24(3):300–308.

46. Nyamukondiwa C, Weldon CW, Chown SL, le Roux PC, Terblanche JS. Thermal biology, population fluctuations and implications of temperature extremes for the management of two globally significant insect pests. Journal of Insect Physiology. 2013;59(12):1199–1211. doi:10.1016/j.jinsphys.2013.09.004.

47. Schmidt TL, Filipović I, Hoffmann AA, Rašić G. Fine-scale landscape genomics helps explain the slow spatial spread of Wolbachia through the Aedes aegypti population in Cairns, Australia. Heredity. 2018; p. 1–10. doi:10.1038/s41437-017-0039-9.

48. Taylor C, Touré YT, Carnahan J, Norris DE, Dolo G, Traoré SF, et al. Gene flow among populations of the malaria vector, Anopheles gambiae, in Mali, West Africa. Genetics. 2001;157(2):743–50.

49. Daszykowski M, Walczak B. Density-Based Clustering Methods. Comprehensive Chemometrics. 2010;2:635–654. doi:10.1016/B978-044452701-1.00067-3.

50. Freeman LC. Centrality in social networks conceptual clarification. Social Networks. 1978;1(3):215–239. doi:10.1016/0378-8733(78)90021-7.

51. Champer J, Reeves R, Oh SY, Liu C, Liu J, Clark AG, et al. Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. PLoS Genetics. 2017;13(7):1–18. doi:10.1371/journal.pgen.1006796.

52. Prowse TAA, Cassey P, Ross JV, Pfitzner C, Wittmann TA, Thomas P. Dodging silver bullets: good CRISPR gene-drive design is critical for eradicating exotic vertebrates. Proceedings of the Royal Society B: Biological Sciences. 2017;284(1860):20170799. doi:10.1098/rspb.2017.0799.