Received: 26 June 2019 | Accepted: 30 September 2019

DOI: 10.1111/2041-210X.13318

APPLICATION

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

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[1,3]

[1]Division of Epidemiology and Biostatistics, School of Public Health, University of California, Berkeley, CA, USA

[2]Biophysics Graduate Group, Division of Biological Sciences, College of Letters and Science, University of California, Berkeley, CA, USA

[3]Innovative Genomics Institute, Berkeley, CA, USA

Correspondence
Héctor M. Sánchez C.
Email: sanchez.hmsc@berkeley.edu

John M. Marshall
Email: john.marshall@berkeley.edu

Funding information
Innovative Genomics Institute; UC Irvine Malaria Initiative; Defense Advanced Research Projects Agency, Grant/Award Number: HR0011-17-2-0047

Handling Editor: Nick Golding

Abstract

1. 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 enabled a wide range of gene drive architectures to be realized, creating a need for their population-level and spatial dynamics to be explored.
2. 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 pests, and (c) a landscape module generates the metapopulation model by which insect populations are connected via migration over space.
3. 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. Further documentation and use examples are provided at the project's Github repository.
4. MGDrivE is an open-source R package freely available on CRAN. We intend the package to provide a flexible tool capable of modelling novel inheritance-modifying constructs as they are proposed and become available. The field of gene drive is moving very quickly, and we welcome suggestions for future development.

Sean L. Wu and Jared B. Bennett contributed equally to this work.

KEYWORDS
Aedes aegypti, Anopheles gambiae, inheritance pattern, landscape, life history, mathematical model, population dynamics, R package

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 (Champer, Buchman, & Akbari, 2016; Esvelt, Smidler, Catteruccia, & Church, 2014). 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 (Walker, Griffin, Ferguson, & Ghani, 2016).

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 (Champer et al., 2016; Marshall & Akbari, 2018). 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) (Burt, 2003; Windbichler et al., 2011), (b) remediation systems that could be used to remove effector genes and possibly drive systems from the environment in the event of unwanted consequences (Gantz & Bier, 2014; Marshall & Akbari, 2018), and (c) threshold-dependent systems that may permit confineable and reversible releases (Akbari et al., 2013; Buchman, Ivy, Marshall, Akbari, & Hay, 2018; Marshall & Hay, 2012).

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).For example, Eckhoff (2011) 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. (2009) created Skeeter Buster by extending the CIMSiM model (Focks, Daniels, Haile, & Keesling, 1995) 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 (Legros et al., 2012). The SLiM genetic simulation framework (Haller & Messer, 2017) 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.

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.

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	Populations distributed in space, connected by migration	R package, open source
EMOD (Eckhoff, 2011)	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 (Legros et al., 2012)	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 (Haller & Messer, 2017)	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

2 | DESIGN AND IMPLEMENTATION

The MGDrivE framework is a genetic and spatial extension of the lumped age-class model of mosquito ecology (Hancock & Godfray, 2007) modified and applied by Deredec, Godfray, and Burt (2011) to the spread of homing gene drive systems, and by Marshall, Buchman, Sánchez C., and Akbari (2017) 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 (https://www.r-project.org/) 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.

2.1 | Modules

2.1.1 | 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). 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 (Marshall et al., 2017; Unckless, Clark, & Messer, 2017). 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 having each genotype – for example, genotype-specific: (a) fertility rates, (b) male mating fitness, (c) sex bias at emergence, (d)

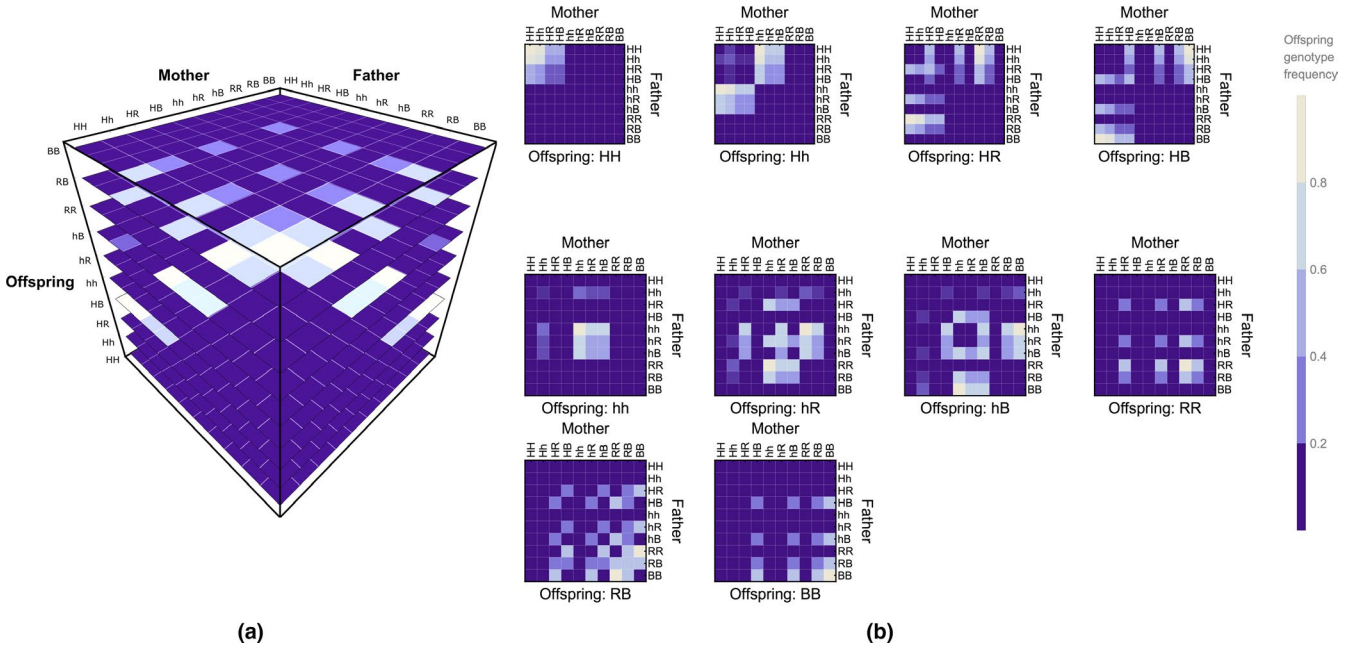

FIGURE 1 Inheritance module. Genetic inheritance is embodied by a three-dimensional tensor referred to as an inheritance cube (left), here depicted for a CRISPR-based homing construct. Maternal and paternal genotypes are depicted on the *x* and *y*-axes and offspring genotypes on the *z*-axis, with slices 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

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 (Gantz et al., 2015; Hammond et al., 2016; Marshall et al., 2017), (c) *Medea* (a maternal toxin linked to a zygotic antidote) (Chen et al., 2007), (d) other toxin-antidote-based underdominant systems such as UD^{MEL} (Akbari et al., 2013; Marshall, Pittman, Buchman, & Hay, 2011), (e) reciprocal chromosomal translocations (Buchman et al., 2018), (f) *Wolbachia* (Hancock, Sinkins, & Godfray, 2011), and (g) the RIDL and pgSIT systems (Kandul et al., 2019; Wise de Valdez et al., 2011) (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 at https://marshalllab.github.io/MGDrivE/docs/reference/.

2.1.2 | Mosquito life history

The mosquito life-history module follows from the lumped age-class model of Hancock and Godfray (2007) adapted by Deredec et al. (2011). In this model (depicted in Figure 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 et al. (2011), 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 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 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.

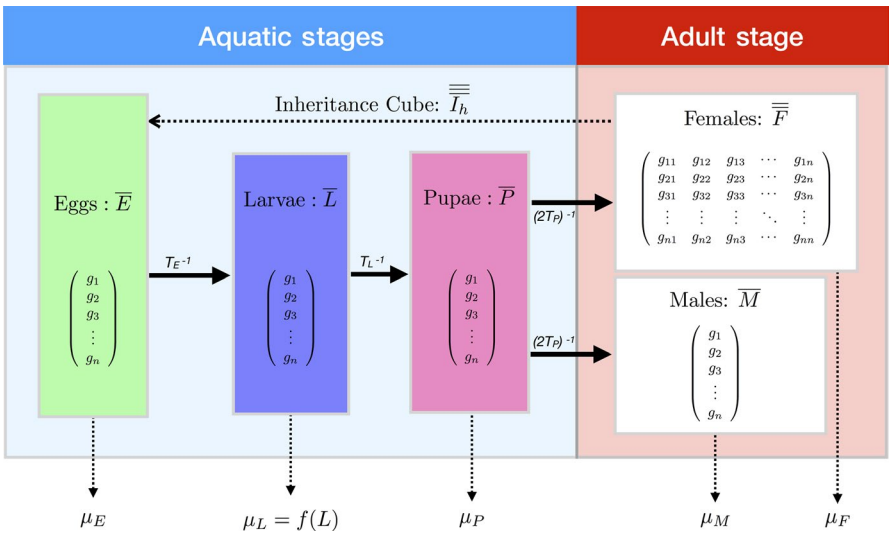

FIGURE 2 Mosquito lifehistory module. Life history is modelled according to an egg (E)-larva (L)-pupa (P)-adult (F for female, M for male) life cycle in which density dependence occurs at the larval stage and autonomous mobility occurs at the adult stage. Genotypes are tracked across all life stages, represented by the subscript $i \in \{1, ..., g\}$. For example, M_i represents the number of adult males having the ith 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

TABLE 2 Life-history module parameter values for three species of interest (at a temperature of 25°)

Parameter	Symbol	*Aedes aegypti*	*Anopheles gambiae*	*Ceratitis capitata*
Egg production per female (day^{-1})	β	20 (Otero, Solari, & Schweigmann, 2006)	32 (Depinay et al., 2004)	20 (Diamantidis, Carey, Nakas, & Papadopoulos, 2011)
Duration of egg stage (days)	T_E	5 (Christophers, 1960)	1 (Depinay et al., 2004)	2 (Diamantidis et al., 2011)
Duration of larval stage (days)	T_L	6 (Christophers, 1960)	13 (Depinay et al., 2004)	6 (Diamantidis et al., 2011)
Duration of pupa stage (days)	T_P	4 (Christophers, 1960)	1 (Depinay et al., 2004)	10 (Diamantidis et al., 2011)
Daily population growth rate (day^{-1})	r_M	1.175 (Simoy, Simoy, & Canziani, 2015)	1.096 (Molineaux & Gramiccia, 1980)	1.031 (Carey, Liedo, & Vaupel, 1995)
Daily mortality risk of adult stage (day^{-1})	μ_F, μ_M	0.090 (Fay, 1964; Focks, Haile, Daniels, & Mount, 1993; Horsfall, 1955)	0.123 (Molineaux & Gramiccia, 1980)	0.100 (Nyamukondiwa, Weldon, Chown, le Roux, & Terblanche, 2013)

2.1.3 | 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 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 (Schmidt, Filipović, Hoffmann, & Rašić, 2018). 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 (Taylor et al., 2001). 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 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 $1/\lambda$ 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 (Schmidt et al., 2018), 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.

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.

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 at https://CRAN.R-project.org/package=MGDrivE, and additional examples and information on Github at https://github.com/MarshallLab/MGDrivE and the package website, https://marshalllab.github.io/MGDrivE/. 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

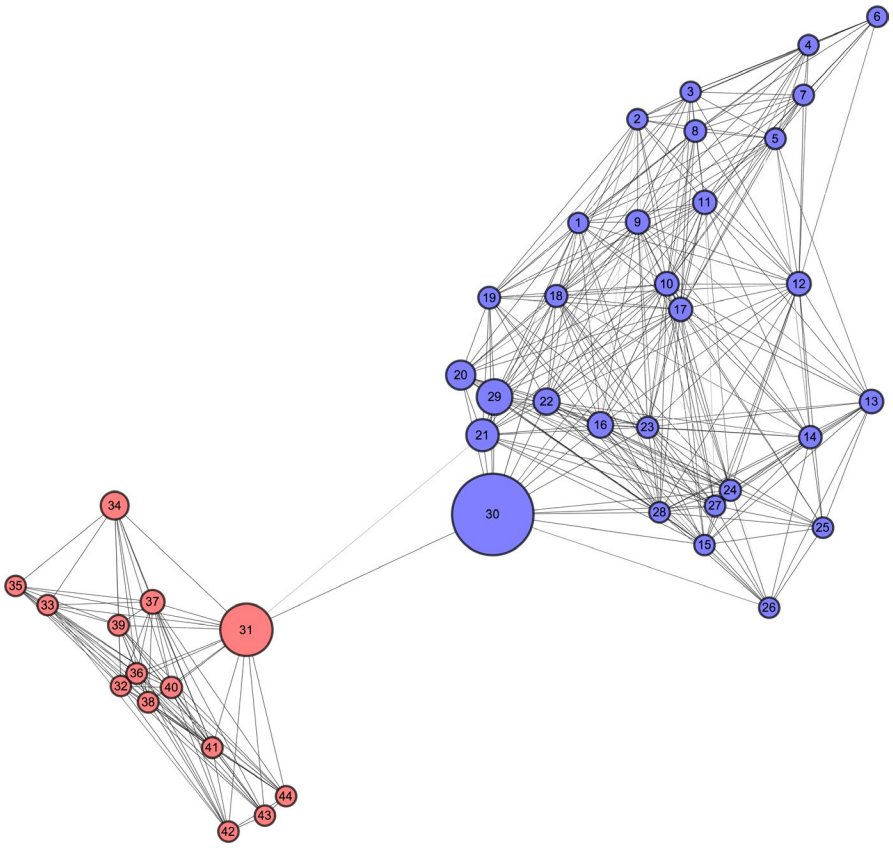

FIGURE 3 Landscape module. Insects are distributed as populations, here depicted by nodes, each having their own coordinates and population size. Movement between populations is derived from a defined dispersal kernel, depicted here by edges between nodes. The example scenario allows both spread within and between communities to be explored. Here, nodes are coloured according to their community (detected by the DBSCAN clustering algorithm, Daszykowski & Walczak, 2010), with sizes proportional to their 'betweenness centrality' – a measure of their connectedness to other nodes in the metapopulation (Freeman, 1978)

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 (Gantz et al., 2015), and (b) disrupting a gene required for female fertility and hence suppressing a population (Hammond et al., 2016). In both cases, we consider a population of *A. 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 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 (Champer et al., 2017) and parameterize the gene drive model based on recently engineered constructs (Gantz et al., 2015; Hammond et al., 2016).

3.1 | Population replacement

We begin by modelling a CRISPR-based homing construct similar to that engineered by Gantz et al. (2015). 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 (Champer et al., 2017). Female fecundity and male mating 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 4, with the full code available at https://github.com/MarshallLab/MGDrivE/tree/master/Examples/. 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 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 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

FIGURE 4 Workflow of an MGDrivE simulation

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).

With our life history and landscape modules defined, we now specify the gene drive system and release strategy we intend to model (Code sample 2). 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 et al. (2015). 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 3).

```
# 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)
```

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}
```

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()
}
```

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.

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 et al., 2016). 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 (Marshall et al., 2017). Lines of code that differ for this system are shown in Code sample 2. While the same inheritance cube applies, specific parameters differ – namely, homing and resistant allele generation rates, and genotype-specific fitness effects.

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,

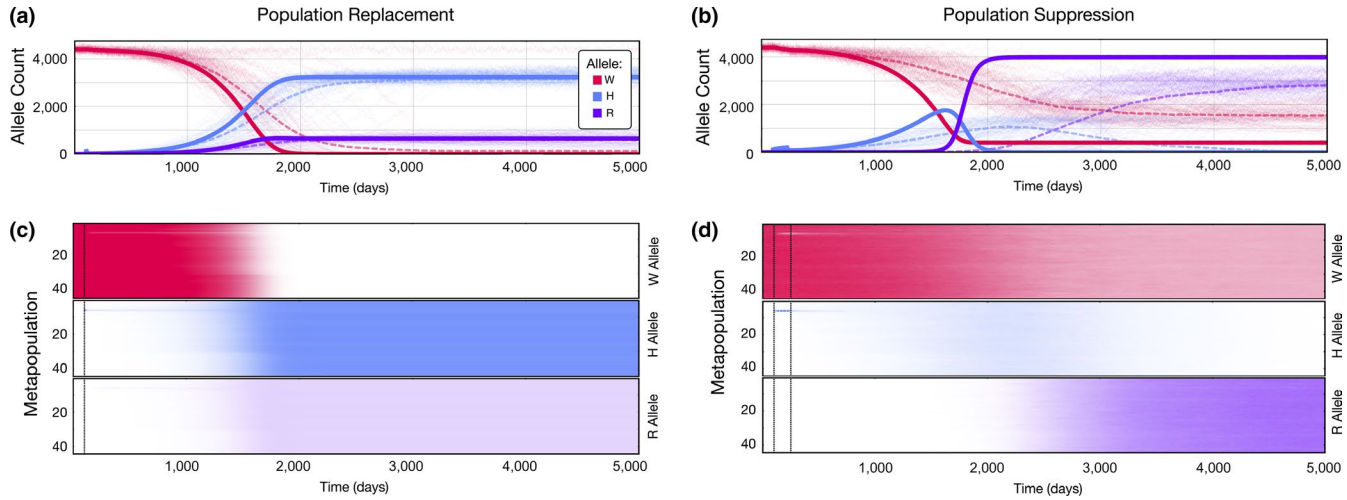

FIGURE 5 Example MGDrivE simulations for CRISPR-based homing constructs. In both cases, an *Aedes aegypti* population is simulated having the bionomic parameters in Table 2 and distributed through the landscape depicted in Figure 3. Deterministic simulations are denoted by solid lines in panels (a) and (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

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 5, 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 5a and b and allele frequency (depicted as colour density) in Figure 5c and d, with time on the horizontal axis. For population replacement (Figure 5a 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 5a).

For population suppression (Figure 5b 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 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 5 are provided at https://github.com/MarshallLab/MGDrivE/tree/master/Examples.

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 (Champer et al., 2017), (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 (Noble, Olejarz, Esvelt, Church, & Nowak, 2017), and multiplexing schemes in which a single gene is targeted at multiple locations to reduce the rate of resistant allele formation (Prowse et al., 2017).

ACKNOWLEDGEMENTS

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

AUTHORS' CONTRIBUTIONS

H.M.S.C. and J.M.M. conceived the project. H.M.S.C. led MGDrivE development and S.L.W. and J.B.B. contributed substantially to core development. H.M.S.C. and J.M.M. wrote the first draft of the manuscript. J.B.B. and S.L.W. wrote the vignettes. All authors revised the manuscript and approved for publication.

ORCID

Héctor M. Sánchez C. https://orcid.org/0000-0001-7378-8853
John M. Marshall https://orcid.org/0000-0003-0603-7341

DATA AVAILABILITY STATEMENT

MGDrivE version 1.1.0 is available on CRAN at https://CRAN.R-project.org/package=MGDrivE. Additional examples and plotting scripts are available on Github at https://github.com/MarshallLab/MGDrivE, and the package website, https://marshalllab.github.io/MGDrivE/. The source code is available under the GPL3 License and is free for other groups to modify and extend as needed (https://doi.org/10.5281/zenodo.3479781). Mathematical details of the model formulation are available in Data S1, and documentation of all MGDrivE functions, including vignettes, are available at the project's Github repository at https://marshalllab.github.io/MGDrivE/docs/reference/. To run the software, we recommend using R version 3.4.4 or higher.

REFERENCES

Akbari, O. S., Matzen, K. D., Marshall, J. M., Huang, H., Ward, C. M., & Hay, B. A. (2013). A synthetic gene drive system for local, reversible modification and suppression of insect populations. *Current Biology*, *23*(8), 671–677. https://doi.org/10.1016/j.cub.2013.02.059

Buchman, A. B., Ivy, T., Marshall, J. M., Akbari, O. S., & Hay, B. A. (2018). Engineered reciprocal chromosome translocations drive high threshold, reversible population replacement in drosophila. *ACS Synthetic Biology*, *7*, 1359–1370. https://doi.org/10.1021/acssynbio.7b00451

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

Carey, J. R., Liedo, P., & Vaupel, J. W. (1995). Mortality dynamics of density in the mediterranean fruit fly. *Experimental Gerontology*, *30*(6), 605–629. https://doi.org/10.1016/0531-5565(95)00013-5

Champer, J., Buchman, A., & Akbari, O. S. (2016). Cheating evolution: Engineering gene drives to manipulate the fate of wild populations. *Nature Reviews Genetics*, *17*(3), 146–159. https://doi.org/10.1038/nrg.2015.34

Champer, J., Reeves, R., Oh, S. Y., Liu, C., Liu, J., Clark, A. G., & Messer, P. W. (2017). Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLoS Genetics*, *13*(7), 1–18. https://doi.org/10.1371/journal.pgen.1006796

Chen, C. H., Huang, H., Ward, C. M., Su, J. T., Schaeffer, L. V., Guo, M., & Hay, B. A. (2007). Replacement in drosophila. *Science*, *316*, 597–600.

Christophers, S. R. (1960). *Aedes aegypti the Yellow Fever Mosquito* (1st ed.). London: Cambridge University Press.

Daszykowski, M., & Walczak, B. (2010). Density-based clustering methods. *Comprehensive Chemometrics*, *2*, 635–654.

Depinay, J. M., Mbogo, C. M., Killeen, G., Knols, B., Beier, J., Carlson, J., … McKenzie, F. E. (2004). A simulation model of African Anopheles ecology and population dynamics for the analysis of malaria transmission. *Malaria Journal*, *3*, 29.

Deredec, A., Godfray, H. C. J., & Burt, A. (2011). Requirements for effective malaria control with homing endonuclease genes. *Proceedings of the National Academy of Sciences*, *108*(43), E874–E880. https://doi.org/10.1073/pnas.1110717108

Diamantidis, A. D., Carey, J. R., Nakas, C. T., & Papadopoulos, N. T. (2011). Population-specific demography and invasion potential in medfly. *Ecology and Evolution*, *1*(4), 479–488. https://doi.org/10.1002/ece3.33

Eckhoff, P. A. (2011). A malaria transmission-directed model of mosquito life cycle and ecology. *Malaria Journal*, *10*(1), 303. https://doi.org/10.1186/1475-2875-10-303

Esvelt, K. M., Smidler, A. L., Catteruccia, F., & Church, G. M. (2014). Concerning RNA-guided gene drives for the alteration of wild populations. *eLife*, *3*, e03401. https://doi.org/10.7554/eLife.03401

Fay, R. W. (1964). The biology and bionomics of *Aedes aegypti* in the laboratory. *Mosquito News*, *24*(3), 300–308.

Focks, D. A., Daniels, E., Haile, D. G., & Keesling, J. E. (1995). 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*, *53*(5), 489–506. https://doi.org/10.4269/ajtmh.1995.53.489

Focks, D. A., Haile, D. G., Daniels, E., & Mount, G. A. (1993). Dynamic life table model for *Aedes aegypti* (Diptera: Culicidae): Analysis of the literature and model development. *Journal of Medical Entomology*, *30*(6), 1003–1017. https://doi.org/10.1093/jmedent/30.6.1003

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239. https://doi.org/10.1016/0378-8733(78)90021-7

Gantz, V., & Bier, E. (2014). The mutagenic chain reaction: A method for converting heterzygous to homozygous mutations. *Human Molecular Genetics*, *348*(March), 1–7.

Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E., & James, A. A. (2015). Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proceedings of the National Academy of Sciences*, *112*(49), E6736–E6743. https://doi.org/10.1073/pnas.1521077112

Haller, B. C., & Messer, P. W. (2017). SLiM 2: Flexible, interactive forward genetic simulations. *Molecular Biology and Evolution*, *34*(1), 230–240. https://doi.org/10.1093/molbev/msw211

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

Hancock, P. A., & Godfray, H. C. J. (2007). Application of the lumped age-class technique to studying the dynamics of malaria-mosquito-human interactions. *Malaria Journal*, *6*, 98. https://doi.org/10.1186/1475-2875-6-98

Hancock, P. A., Sinkins, S. P., & Godfray, H. C. J. (2011). Population dynamic models of the spread of Wolbachia. *The American Naturalist*, *177*(3), 323–333. https://doi.org/10.1086/658121

Horsfall, W. E. (1955). *Mosquitoes: Their bionomics and relation to disease*. New York, NY: The Eonald Press Co.

Kandul, N. P., Liu, J., Sánchez C., H. M., Wu, S. L., Marshall, J. M., & Akbari, O. S. (2019). Transforming insect population control with precision guided sterile males with demonstration inflies. *Nature Communications*, *10*. https://doi.org/10.1038/s41467-018-07964-7

Legros, M., Xu, C., Okamoto, K., Scott, T. W., Morrison, A. C., Lloyd, A. L., & Gould, F. (2012). Assessing the feasibility of controlling *Aedes aegypti* with transgenic methods: A model-based evaluation. *PLoS ONE*, *7*(12), e52235. https://doi.org/10.1371/journal.pone.0052235

Magori, K., Legros, M., Puente, M. E., Focks, D. A., Scott, T. W., Lloyd, A. L., & Gould, F. (2009). Skeeter Buster: A stochastic, spatially explicit modeling tool for studying *Aedes aegypti* population replacement and population suppression strategies. *PLoS Neglected Tropical Diseases*, *3*(9), e508. https://doi.org/10.1371/journal.pntd.0000508

Marshall, J. M., & Akbari, O. S. (2018). Can CRISPR-based gene drive be confined in the wild? a question for molecular and population biology. *ACS Chemical Biology*, *13*(2), 424–430. https://doi.org/10.1021/acschembio.7b00923

Marshall, J. M., Buchman, A., Sánchez C., H. M., & Akbari, O. S. (2017). Overcoming evolved resistance to population-suppressing homing-based gene drives. *Scientific Reports*, *7*(1), 1–46. https://doi.org/10.1038/s41598-017-02744-7

Marshall, J. M., & Hay, B. A. (2012). General principles of single-construct chromosomal gene drive. *Evolution*, *66*(7), 2150–2166. https://doi.org/10.1111/j.1558-5646.2012.01582.x

Marshall, J. M., Pittman, G. W., Buchman, A. B., & Hay, B. A. (2011). Semele: A killer-male, rescue-female system for suppression and replacement of insect disease vector populations. *Genetics*, *187*(2), 535–551. https://doi.org/10.1534/genetics.110.124479

Molineaux, L., & Gramiccia, G. (1980). *The Garki Project. Research on the epidemiology and control of malaria in the Sudan Savanna of West Africa*. Geneva: World Health Organization.

Noble, C., Olejarz, J., Esvelt, K. M., Church, G. M., & Nowak, M. A. (2017). Evolutionary dynamics of CRISPR gene drives. *Science Advances*, *3*(4), 3–10. https://doi.org/10.1126/sciadv.1601964

Nyamukondiwa, C., Weldon, C. W., Chown, S. L., le Roux, P. C., & Terblanche, J. S. (2013). Thermal biology, population fluctuations and implications of temperature extremes for the management of two globally significant insect pests. *Journal of Insect Physiology*, *59*(12), 1199–1211. https://doi.org/10.1016/j.jinsphys.2013.09.004

Otero, M., Solari, H. G., & Schweigmann, N. (2006). A stochastic population dynamics model for *Aedes aegypti*: Formulation and application to a city with temperate climate. *Bulletin of Mathematical Biology*, *68*(8), 1945–1974. https://doi.org/10.1007/s11538-006-9067-y

Prowse, T. A. A., Cassey, P., Ross, J. V., Pfitzner, C., Wittmann, T. A., & Thomas, P. (2017). Dodging silver bullets: Good CRISPR gene-drive design is critical for eradicating exotic vertebrates. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1860), 20170799. https://doi.org/10.1098/rspb.2017.0799

Schmidt, T. L., Filipović, I., Hoffmann, A. A., & Rašić, G. (2018). Fine-scale landscape genomics helps explain the slow spatial spread of Wolbachia through the *Aedes aegypti* population in Cairns, Australia. *Heredity*, *120*(5), 386–395. https://doi.org/10.1038/s41437-017-0039-9

Simoy, M. I., Simoy, M. V., & Canziani, G. A. (2015). The effect of temperature on the population dynamics of *Aedes aegypti*. *Ecological Modelling*, *314*, 100–110. https://doi.org/10.1016/j.ecolmodel.2015.07.007

Taylor, C., Touré, Y. T., Carnahan, J., Norris, D. E., Dolo, G., Traoré, S. F., … Lanzaro, G. C. (2001). Gene flow among populations of the malaria vector, *Anopheles gambiae*, in Mali, West Africa. *Genetics*, *157*(2), 743–750.

Unckless, R. L., Clark, A. G., & Messer, P. W. (2017). Evolution of resistance against CRISPR/Cas9 gene drive. *Genetics*, *205*(2), 827–841. https://doi.org/10.1534/genetics.116.197285

Walker, P. G. T., Griffin, J. T., Ferguson, N. M., & Ghani, A. C. (2016). 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*, *4*(7), e474–e484. https://doi.org/10.1016/S2214-109X(16)30073-0

Windbichler, N., Menichelli, M., Papathanos, P. A., Thyme, S. B., Li, H., Ulge, U. Y., … Crisanti, A. (2011). A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature*, *473*(7346), 212–215. https://doi.org/10.1038/nature09937

Wise de Valdez, M. R., Nimmo, D., Betz, J., Gong, H.-F., James, A. A., Alphey, L., & Black, W. C. (2011). Genetic elimination of dengue vector mosquitoes. *Proceedings of the National Academy of Sciences*, *108*(12), 4772–4775. https://doi.org/10.1073/pnas.1019295108

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: 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. *Methods Ecol Evol*. 2019;00:1–11. https://doi.org/10.1111/2041-210X.13318