

# SCIENTIFIC REPORTS



OPEN

## Overcoming evolved resistance to population-suppressing homing-based gene drives

John M. Marshall<sup>1</sup>, Anna Buchman<sup>2</sup>, Héctor M. Sánchez C.<sup>3</sup>  & Omar S. Akbari<sup>2</sup>

The recent development of a CRISPR-Cas9-based homing system for the suppression of *Anopheles gambiae* is encouraging; however, with current designs, the slow emergence of homing-resistant alleles is expected to result in suppressed populations rapidly rebounding, as homing-resistant alleles have a significant fitness advantage over functional, population-suppressing homing alleles. To explore this concern, we develop a mathematical model to estimate tolerable rates of homing-resistant allele generation to suppress a wild population of a given size. Our results suggest that, to achieve meaningful population suppression, tolerable rates of resistance allele generation are orders of magnitude smaller than those observed for current designs for CRISPR-Cas9-based homing systems. To remedy this, we theoretically explore a homing system architecture in which guide RNAs (gRNAs) are multiplexed, increasing the effective homing rate and decreasing the effective resistant allele generation rate. Modeling results suggest that the size of the population that can be suppressed increases exponentially with the number of multiplexed gRNAs and that, with four multiplexed gRNAs, a mosquito species could potentially be suppressed on a continental scale. We also demonstrate successful proof-of-principle use of multiplexed ribozyme flanked gRNAs to induce mutations *in vivo* in *Drosophila melanogaster* – a strategy that could readily be adapted to engineer stable, homing-based drives in relevant organisms.

The concept of using homing-based gene drive systems to rapidly invade wild populations and spread effector genes (e.g. conferring pathogen resistance) or to suppress and eliminate populations was first suggested by Burt in 2003<sup>1</sup>. These systems have the remarkable ability to cheat during meiosis, enabling them to rapidly spread into a population even if they confer a fitness cost to their host (2, 3). They achieve this by encoding a sequence-specific nuclease that generates a double-stranded break at one or more specific target loci in a host's genome, directly opposite the drive. To survive, the cell is forced to rapidly repair the DNA break using its endogenous DNA repair machinery. Repair of the break using the homology-directed repair (HDR) pathway, for instance, can result in the drive system being perfectly copied into its competing allele. When this occurs in a germline cell, it effectively converts a heterozygote into a homozygote, allowing the system to circumvent traditional Mendelian inheritance patterns and to drive into a population<sup>2,3</sup>. The first decade following this proposition saw moderate progress in the development of homing-based drive systems in the African malaria vector, *Anopheles gambiae*<sup>2,4,5</sup>. More recently, the development of the CRISPR-Cas9 system has unlocked enormous potential for this technology, with highly efficient drive systems being developed in quick succession to modify laboratory populations of *Drosophila melanogaster*<sup>6</sup>, *Saccharomyces cerevisiae*<sup>7</sup>, the Asian malaria vector, *Anopheles stephensi*<sup>8</sup>, and the main African malaria vector, *An. gambiae*<sup>9</sup>.

The homing-based drive systems developed using CRISPR-Cas9 have a number of highly desirable features: they are relatively straightforward to adapt to new target sequences and to port to other species, and the constructs engineered thus far have extremely high transmission rates, being inherited by 90–99% of the offspring of heterozygous parents<sup>6–9</sup>. However, these systems are not without their shortcomings. Firstly, the CRISPR-Cas9-based constructs engineered to date are associated with high fitness costs, although this is likely not an inherent problem with CRISPR-Cas9 systems, and secondly, the homing process has been shown to be highly error-prone, leading

<sup>1</sup>Divisions of Biostatistics and Epidemiology, School of Public Health, University of California, Berkeley, CA, 94720, USA. <sup>2</sup>Department of Entomology, Center for Disease Vector Research, and Institute for Integrative Genome Biology, University of California, Riverside, CA, 92521, USA. <sup>3</sup>Bioinformatics Research Group, School of Medicine, Tecnológico de Monterrey, Estado de México, 52926, México, USA. Correspondence and requests for materials should be addressed to J.M.M. (email: [john.marshall@berkeley.edu](mailto:john.marshall@berkeley.edu)) or O.S.A. (email: [omar.akbari@ucr.edu](mailto:omar.akbari@ucr.edu))

to the creation of mutant alleles within a few generations<sup>8,9</sup>. A significant number of these mutant alleles consist of small deletions that preserve the reading frame and could represent homing-resistant alleles, i.e. mutant alleles that are not associated with a fitness cost<sup>8,9</sup>. This latter shortcoming is of concern for population suppression strategies because homing-resistant alleles have a strong selective advantage over functional homing alleles that disrupt an essential gene, leading to suppressed populations rapidly rebounding.

Mutant alleles may be generated in several ways. For instance, they can emerge when the cell mends DNA damage at the target site using the non-homologous end-joining (NHEJ) pathway instead of HDR following drive-induced target site cleavage. Mutant alleles may also arise due to incomplete or imperfect copying during HDR. The CRISPR-Cas9 system is particularly vulnerable to this due to its large size – the system consists of promoters, the Cas9 gene, guide RNAs and, depending on the strategy being implemented, multiple effector genes and associated regulatory elements, all of which need to be perfectly copied during HDR to ensure spread into a population. Indeed, for the CRISPR-Cas9 homing construct engineered in *An. gambiae*<sup>9</sup>, incomplete homing or internal deletion events were observed in 47% (15 out of 32) of screened organisms in which an errorless homing event was not observed. These were derived from a minimum of seven originating erroneous homing events, one of which produced a small deletion that preserved the reading frame and may be considered a homing-resistant allele. Homing-resistant alleles may also arise *de novo* via random target site mutagenesis, and some organisms may intrinsically be resistant to homing activity at a given site due to genetic variation within a species.

Therefore, while CRISPR-Cas9-based homing systems have enormous potential for the targeted engineering of populations, significant technical improvements are required if this technology is to be successfully implemented in the field<sup>2,10</sup>. Here, we focus specifically on the issue of homing-resistant allele generation for population suppression homing systems. We largely ignore homing allele fitness costs in this analysis as we consider these to be surmountable through tailored engineering efforts – in one of the constructs engineered thus far, fitness costs seem to result from the transgene being inserted into an eye color gene<sup>8</sup>, and in another, due to the element copying itself to somatic as well as germline cells<sup>9</sup>, both of which we believe to be addressable. The impact of homing-resistant alleles on homing-based population replacement strategies has been described by Noble *et al.*<sup>11</sup> along with a design strategy that selects against the resistant alleles; however, this solution does not apply to the population suppression systems that we explore here.

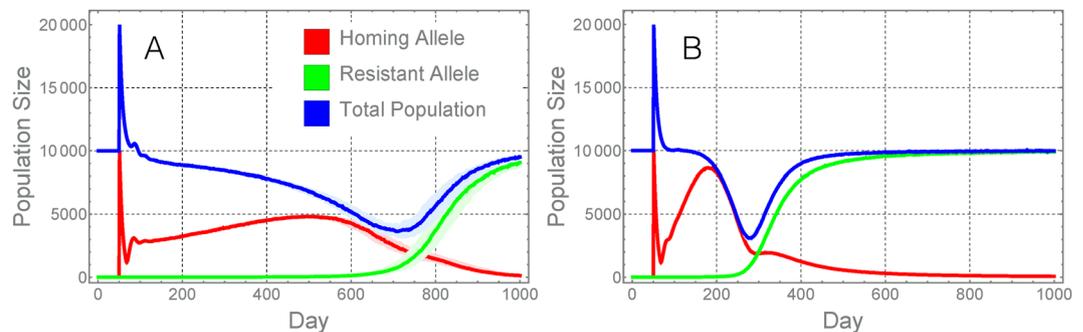
To address the impact of resistant alleles on homing-based population suppression systems, we develop a mathematical model to estimate the maximum tolerable resistant allele generation rates to achieve stable, long-term suppression for populations of various sizes. Our results suggest that, to achieve meaningful population suppression, tolerable rates of resistant allele generation are orders of magnitude lower than those observed for current CRISPR-Cas9-based homing systems. We describe how the required rates can be achieved by targeting multiple locations in a gene through guide RNA (gRNA) multiplexing<sup>2,3,11,12</sup>. Furthermore, we demonstrate successful adaptation of a multiplexed ribozyme-gRNA-ribozyme (RGR) approach previously demonstrated in yeast<sup>13</sup> and in mammalian tissue culture cells<sup>13,14</sup> to *in vivo* mutagenesis of target genes in *Drosophila melanogaster*, and discuss possible designs for, and challenges inherent in, the gRNA multiplexing approach for engineering stable, homing-based suppression gene drive systems. Finally, we explore the scale of population suppression that can be achieved by using this approach.

## Results

The homing population suppression system we explore here is based on that described by Hammond *et al.*<sup>9</sup> in which the CRISPR-Cas9 system is designed to target a gene required for female fertility. This has the effect that females homozygous for the homing allele are infertile; however, heterozygous and wild-type females have at least one functional copy of the fertility gene and are fertile. For sufficiently high homing rates and small fitness costs, this system can spread into a population while it reduces population fertility, eventually leading to a population crash<sup>1</sup>. Hammond *et al.*<sup>9</sup> describe three strains that they engineered with this design. We consider the most successful of these – construct 7280 – for which the transmission rate from heterozygotes was ~99%, and heterozygous females had their fitness reduced by 90.7%. Approximately half (~47%) of those who did not inherit a functional homing allele from a heterozygous parent inherited a copy with errors, and a seventh of the erroneous homing events that produced these alleles generated alleles that could be considered homing-resistant. Although the Hammond *et al.*<sup>9</sup> construct was not proposed as a useable drive system, we consider its parameter values as a basis that future homing-based drive systems could build upon.

**Model framework.** The framework used to model this system is described in the Materials and Methods; but in short, we denote the homing allele as “H”, the wild-type allele as “h”, and the homing-resistant allele (mutant allele with no associated fitness cost) as “R”. HH females are infertile, while all other genotypes are fertile. Hh males and females produce H gametes in the germline at a frequency equal to  $(1 + e)/2$ , where  $e$  denotes the efficiency of homing, or “homing rate”. Hh individuals also produce R gametes in the germline at a frequency equal to  $\rho/2$ , where  $\rho$  denotes the resistant allele generation rate. The remaining gametes produced by Hh individuals are wild-type and are produced at a frequency equal to  $(1 - e - \rho)/2$ . Females heterozygous for the homing allele have their fertility reduced by a fraction,  $s$ , while other genotypes are equally fertile (except for HH females, which are infertile). The crosses describing this system are shown in Supplementary Figure 1.

We use a discrete population, stochastic framework incorporating density-dependence at the larval stage to model this system. Our framework is modified from one previously used to examine the spread of homing endonuclease genes (HEGs) through populations of *An. gambiae*<sup>5</sup>, the main malaria vector and the species in which the CRISPR-Cas9-based constructs were developed by Hammond *et al.*<sup>9</sup>. This framework incorporates the egg, larval, pupal and adult life stages. Generations are overlapping and adult females mate once, retaining the genetic material of the male they mate with for the duration of their adult life. Since we are modeling a population suppression system, a discrete, stochastic model is needed to capture the chance events that happen at low population



**Figure 1.** Predicted population dynamics for the present CRISPR-Cas9-based population suppression homing constructs. Here we model the predicted dynamics of the best-performing construct engineered by Hammond *et al.*<sup>9</sup> in a population of 10,000 adult mosquitoes. The homing rate for this construct is  $\sim 98\%$  and the resistant allele generation rate is  $\sim 0.13\%$ . The model framework is described in the Materials and Methods. In panel (A), the dynamics are shown for the scenario in which females heterozygous for the homing allele have their fertility reduced by 90.7%. In panel (B), the same construct is modeled in the absence of a fertility cost. In both cases, population suppression is moderate and short-lived due to the generation of homing-resistant alleles leading to a population rebound. Red lines represent individuals having at least one copy of the homing allele (i.e. genotypes Hh, HR and HH), green lines represent individuals having at least one copy of the homing-resistant allele (i.e. genotypes hR, HR and RR), and blue lines represent the total population. Solid lines represent the median population size for 100 repetitions of the stochastic model, while shaded regions represent the 25–75% quartile range in these simulations, where this differs significantly from the median.

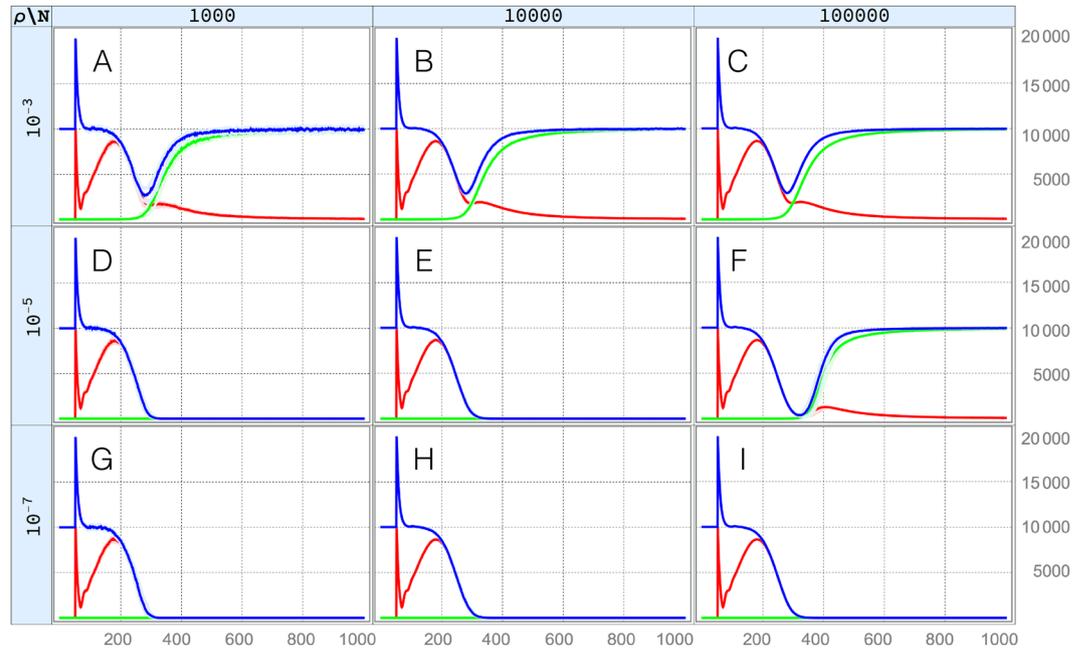
sizes. This also enables the simulation of a population crash. Density-dependence at the larval stage is important to include as it captures the phenomenon in which more larvae survive to emergence when populations are small due to reduced larval competition. To model this, we consider a monotonic increase in larval mortality with larval density, as applied by Deredec *et al.*<sup>5</sup>.

**Expected dynamics of present constructs.** With the modeling framework established, we explore the predicted dynamics of construct 7280, the best-performing construct engineered by Hammond *et al.*<sup>9</sup>, in a population size of  $N = 10,000$  adult mosquitoes. The results described in Fig. 1 correspond to a homing rate of  $e \approx 98\%$  ( $2 \times (99 - 50\%)$ ) and a resistant allele generation rate of  $\rho \approx 0.13\%$  ( $((1 - e) \times 15/32 \times 1/7)$ ), i.e. of the wild-type alleles from heterozygotes that were not converted to homing alleles, 15/32 were mutant alleles and 1/7 of the events that produced these resulted in what may be considered homing-resistant alleles. The scenario in which females heterozygous for the homing allele have their fertility reduced by 90.7% is shown in Fig. 1A. Here, we see that gene drive occurs slowly and population suppression is moderate and transient. The total adult population falls by  $\sim 64\%$  approximately one and three-quarter years following a 1:1 seeding release of HH males to hh males and females; however, this suppression is short-lived – a population reduction of 50% or more is only maintained for about five and a half months before the population rebounds.

Henceforth, let's imagine that the fertility costs of the homing allele in heterozygous females can be prevented through engineering efforts to ensure that the CRISPR-Cas9 system is only expressed in germline cells. This scenario is shown in Fig. 1B. Here, we see that gene drive and population suppression occur more quickly, and the extent of population suppression is slightly greater (the population is suppressed by  $\sim 69\%$  at its peak). However, the duration of suppression is very short – a population reduction of 50% or more is only maintained for about two months. This is far less than what is hoped for gene drive-based population suppression strategies, and is a consequence of the quick emergence of homing-resistant alleles once the gene drive system becomes prevalent in the population, leading to a population rebound.

**Design criteria for population elimination.** The constructs engineered by Hammond *et al.*<sup>9</sup> are clearly inadequate to lead to meaningful population suppression for population sizes of 10,000 adult mosquitoes; but presumably if the homing rate were increased and/or the resistant allele generation rate were decreased, then it may be possible to eliminate a specific population. In Fig. 2, simulations are shown in which homing efficiency is maintained at 98% while the resistant allele generation rate is reduced from 0.1% ( $10^{-3}$ ) to 0.001% ( $10^{-5}$ ) to 0.00001% ( $10^{-7}$ ) for populations of 1,000, 10,000 and 100,000 adult mosquitoes. Here, we see that, for a resistant allele generation rate of  $10^{-3}$  (Fig. 2A–C), which is approximately what was observed for the Hammond *et al.*<sup>9</sup> construct, we do not expect to be able to meaningfully suppress an adult population even as small as 1,000 individuals. However, if we reduce the resistant allele generation rate by two orders of magnitude to  $10^{-5}$  (Fig. 2D–F), then we expect to eliminate populations of sizes 1,000 and 10,000, but not a size of 100,000. As the resistant allele generation rate is further reduced by an additional two orders of magnitude to  $10^{-7}$  (Fig. 2G–I), we expect to eliminate adult populations of all sizes up to 100,000.

This trend of being able to eliminate populations of larger size with smaller resistant allele generation rates is intuitive as, in a larger population, there are more opportunities for error-prone homing events to occur, leading to the emergence of homing-resistant alleles. These resistant alleles will quickly be selected for, reversing any prior population suppression. The design target for the resistant allele generation rate will therefore be determined by the population size we wish to eliminate.



**Figure 2.** Homing and resistant allele trajectories for a range of population sizes and resistant allele generation rates. Here, we model a population suppression homing construct with a homing rate of 98% and no fertility cost. In panels (A–C) the resistant allele generation rate is 0.1% ( $10^{-3}$ ), in panels (D–F) it is 0.0001% ( $10^{-5}$ ), and in panels (G–I) it is 0.00001% ( $10^{-7}$ ). In the leftmost panels (A,D and G), a population of 1,000 is modeled, in the middle panels (B,E and H), it is 10,000, and in the rightmost panels (C,F and I), it is 100,000. Red lines represent individuals having at least one copy of the homing allele, green lines represent individuals having at least one copy of the homing-resistant allele, and blue lines represent the total population. Solid lines represent the median value obtained from 100 repetitions of the stochastic model, while shaded regions represent the 25–75% quartile range, where this differs significantly from the median. As the resistant allele generation rate is reduced, we expect to eliminate populations of larger sizes.

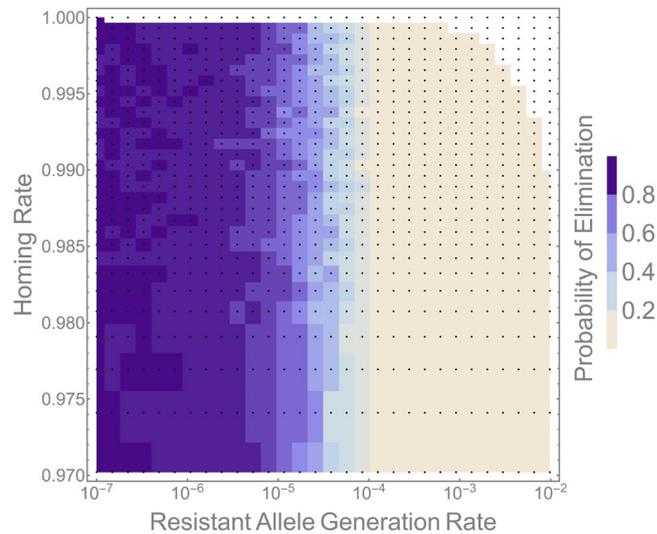
The homing rate, however, does not factor into these design considerations, at least for already high homing rates. Figure 3 shows the probability of population elimination as homing efficiency is varied between 97% and 99.99% and the resistant allele generation rate is varied between  $10^{-2}$  and  $10^{-7}$  for a population of 10,000 adult mosquitoes. The elimination probability is calculated as the proportion of simulations (from a total of 100 per parameter set) in which the *An. gambiae* population was eliminated within 950 days of a 1:1 release of HH males to hh males and females. Here we see that, for a population size of 10,000 adults, population elimination is highly likely for resistant allele generation rates smaller than  $10^{-5}$ , and is unlikely for rates above  $10^{-4}$ . There is a critical rate between these two values at which the population is equally likely to either rebound or be eliminated and, interestingly, these dynamics are independent of the homing rate for  $e > 98\%$ .

**Dependence of the design criteria for  $\rho$  on  $N$ .** The independence of elimination probability and homing rate (for already high homing rates) means that we can focus our attention on achieving a resistant allele generation rate,  $\rho$ , small enough such that population elimination is likely for a given population size,  $N$ . Stochastic simulations become highly computationally intensive as population size increases, and so we seek a relationship between  $N$  and the corresponding resistant allele generation rate,  $\rho$ , for which we can be 90% sure of achieving population elimination (or sure with some other degree of certainty). To this end, Fig. 4A depicts elimination probability as a function of  $\rho$  as we vary  $N$  between 1,000 and 100,000. The familiar case of a population size of 10,000 is shown in light gold and indicates that elimination is ~10% likely for a  $\rho$  value of  $10^{-4}$  and ~80% likely for a  $\rho$  value of  $10^{-5}$ . As the population size increases from 1,000 to 100,000, we see that the  $\rho$  value required to achieve an elimination probability of 90% or higher becomes significantly smaller.

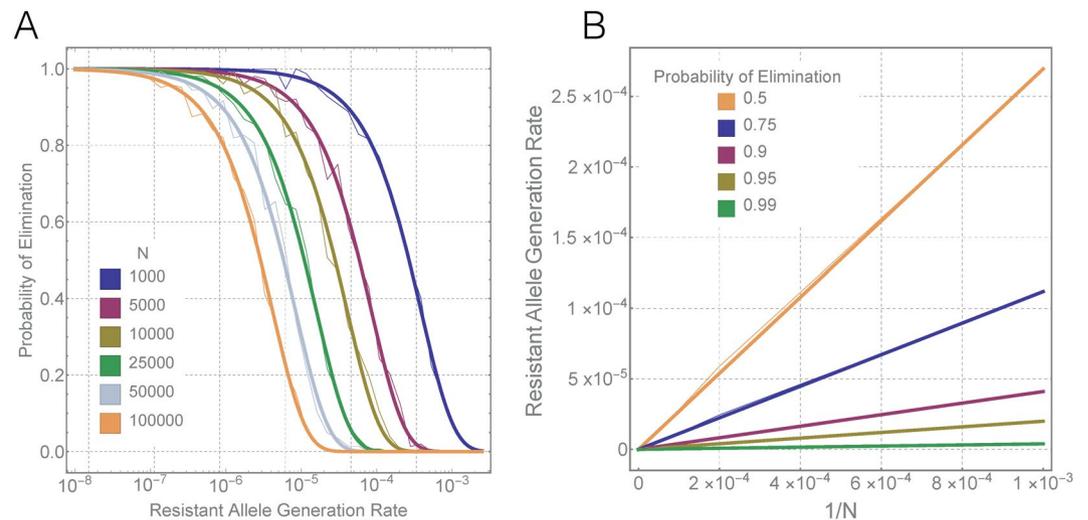
The form of the relationship between  $N$  and  $\rho_x$ , the resistant allele generation rate leading to an elimination probability of  $x$ , is depicted in Fig. 4B for selected elimination probabilities. Fortunately for our ability to extrapolate to larger population sizes, there is a linear relationship between  $1/N$  and  $\rho_x$  for all elimination probabilities investigated, i.e.:

$$\rho_x = c_x/N. \quad (1)$$

Here,  $c_x$  represents the slope from Fig. 4B corresponding to the elimination probability,  $x$ . To be 90% sure of population elimination,  $c_x$  is 0.0410, to be 95% sure,  $c_x$  is 0.0199, and to be 99% sure,  $c_x$  is 0.00391. This means that, to be 90% sure of eliminating a population of size 1,000,  $\rho$  should be less than  $4.1 \times 10^{-5}$  (0.0041%), and to be 90% sure of eliminating a population of size 100,000,  $\rho$  should be less than  $4.1 \times 10^{-7}$  (0.000041%). To be 95% sure of



**Figure 3.** Dependence of population elimination probability on homing rate and resistant allele generation rate. Here, we model a population suppression homing construct in a population of 10,000 adult mosquitoes. Each pixel represents a combination of homing and resistant allele generation rates for which the simulation was run. Pixel shadings represent the proportion of 100 simulations in which population elimination was achieved within 950 days of a seeding 1:1 release of HH males to hh males and females. Both rate parameters were sampled logarithmically to gain higher resolution at high homing rates and low resistant allele generation rates. The white region represents impossible combinations of rate parameters (the rates would sum to  $>1$ ). Population elimination probability is independent of the homing rate (for already high homing rates) and critically dependent on the resistant allele generation rate.



**Figure 4.** Relationship between population size and the resistant allele generation rate required for a given population elimination probability. **(A)** Elimination probability as a function of resistant allele generation rate for a range of population sizes,  $N$ , between 1,000 and 100,000. Sigmoidal curves are fitted to data points covering 30 resistant allele generation rates sampled logarithmically between  $10^{-2}$  and  $10^{-7}$ . **(B)** Linear relationship between  $1/N$  and the resistant allele generation rate leading to a given probability of population elimination. Values of  $1/N$  are as shown in panel A, and resistant allele generation rates are inferred from the sigmoid curves. Faint lines in both panels represent interpolation between simulated data points while solid lines represent fitted linear relationships. There is a clear linear relationship between  $1/N$  and the resistant allele generation rate leading to a given elimination probability.

eliminating populations of these sizes,  $\rho$  should be about half that predicted for a 90% chance of elimination, and for a 99% chance of elimination,  $\rho$  should be an order of magnitude smaller than that predicted for a 90% chance of elimination. These estimates are derived for a mosquito population growth rate of  $R_M = 9.1$  per generation, based on mosquito count data from the Garki project in Nigeria<sup>15</sup>. We repeated these analyses for low, medium

and high population growth rates ( $R_M = 2, 6$  and  $12$  per generation, respectively) and found that, while the slope,  $c_p$ , changes, the linear relationship in Equation 1 still holds (Supplementary Figure 2).

The linear nature of the relationship between  $1/N$  and  $\rho_x$  is understandable since each mosquito presents an opportunity for a resistant allele to emerge and prevent population elimination; however, it is reasonable to question whether population structure has a significant effect on the tolerable resistant allele generation rate. Supplementary Figure 3 explores the relationship between  $1/N$  and  $\rho_x$  for randomly mixing populations of sizes 10,000–50,000 and for 1–5 randomly mixing populations each of size 10,000 that exchange migrants with the other populations at a rate of 1% per adult mosquito per generation. The linear relationship between  $1/N$  and  $\rho_x$  is unchanged by the presence of population structure, which supports the intuition that each mosquito presents an opportunity for resistant allele emergence independent of population structure.

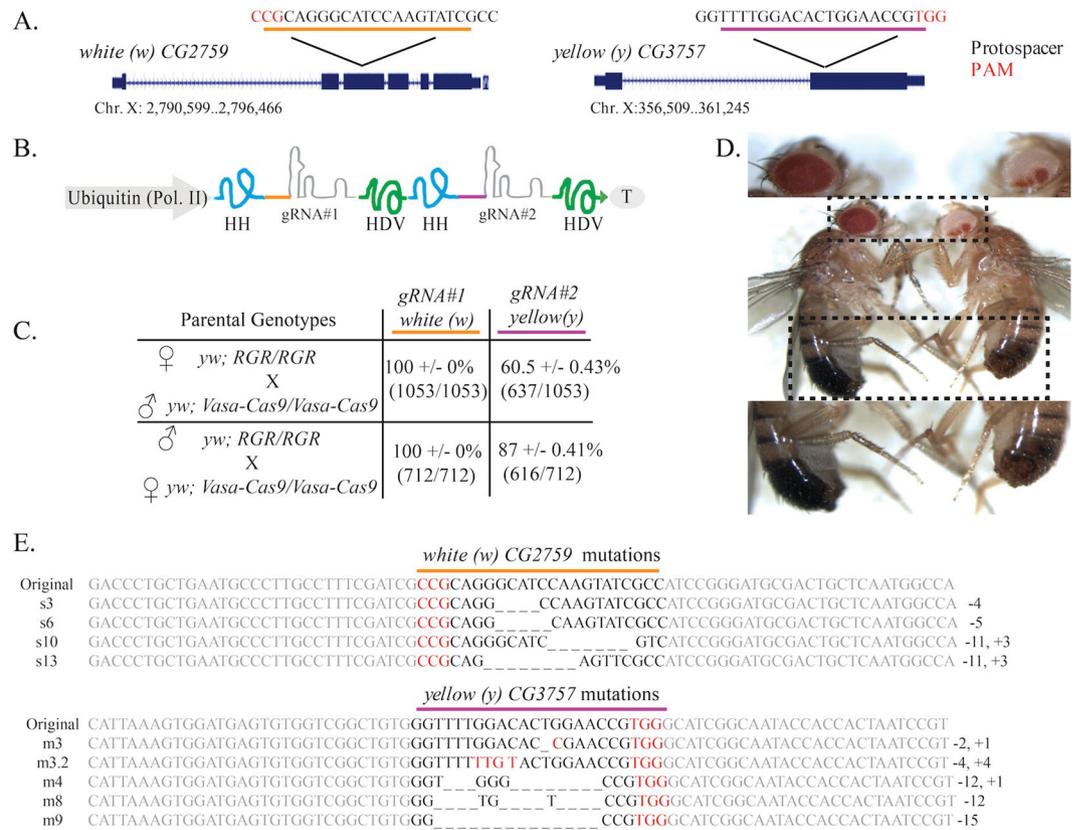
The population size that we wish to eliminate will vary depending on our goals; but one proposition for homing-based gene drive has been to eliminate a disease vector species such as *An. gambiae* on a continental scale. Assuming there are about ten times as many adult *An. gambiae* mosquitoes on the African continent as there are people, this suggests a population of  $\sim 10$  billion ( $10^{10}$ ). The effective population size for a given elimination probability will be smaller than this due to population structure details and seasonal fluctuations in *An. gambiae* population size<sup>16</sup>; however, we consider this larger population size to provide conservative estimates of required  $\rho$  values. To be 90% sure that resistant alleles will not prevent elimination of a population this size,  $\rho$  should be less than  $4.1 \times 10^{-12}$  (0.0000000004%). To be 95% sure,  $\rho$  should be less than  $2.0 \times 10^{-12}$ , and to be 99% sure,  $\rho$  should be less than  $3.9 \times 10^{-13}$ .

**Multiplexing gRNAs.** The  $\rho$  values required to prevent resistant alleles from interfering with the elimination of an *An. gambiae* population on the scale of the African continent are vanishingly small; but interestingly, the  $\rho$  value required to have a 90% chance of suppressing a population of just 1,000 adult mosquitoes is already several orders of magnitude smaller than that observed for the best-performing construct of Hammond *et al.*<sup>9</sup>. Given the inevitability of the evolution of homing-resistant alleles, mitigating their impact is imperative to creating functional and stable homing-based population suppression gene drive systems.

A promising strategy for achieving this feat is the multiplexing of gRNAs in the gene drive to target multiple sequences. This idea has been previously proposed to increase the stability of drives<sup>2, 3, 11, 17</sup>; however, to date only one multiplexing strategy has been demonstrated to function in a whole-animal model<sup>12</sup>. Therefore, to further expand the toolbox for multiplexing gRNAs in whole animals, here we test the effectiveness of a technique previously demonstrated in yeast<sup>13</sup> and in mammalian tissue culture cells<sup>14</sup> that relies on flanking gRNAs with self-cleaving ribozymes, known as the ribozyme-gRNA-ribozyme (RGR) approach *in vivo* in *Drosophila melanogaster*<sup>13</sup>. To validate the ability of this technique to induce mutations *in vivo* in *Drosophila*, we generated plasmid OA-16 that contains two multiplexed RGRs, the first targeting the *white* gene and the second targeting the *yellow* gene at target sequences previously validated<sup>18</sup>, driven by a single polymerase-2 ubiquitin promoter<sup>19</sup> (Fig. 5A,B). The plasmid also contains a *white* gene that could be targeted by the *white* gRNA and used for detecting transgenic individuals bearing the OA-16 construct.

This plasmid was integrated site-specifically into the *Drosophila attP* line BSC (Bloomington Stock Center) 24486. Generated transgenic males and females were individually mated to females and males of transgenic line BSC 51324 (*vasa-Cas9*), and the progeny of the resulting crosses were scored (Fig. 5C). As expected, transformant flies bearing the OA-16 plasmid had no mutations in the *white* or *yellow* genes in the absence of Cas9 (Fig. 5D, left). All the scored offspring (712 from OA-16 male/Cas9 female crosses and 1053 from OA-16 female/Cas9 male crosses) had completely white or significantly variegated eyes (Fig. 5D), indicating that the first of the two RGRs had a relative somatic mutation efficiency near 100%, at least as measured by the proportion of individuals in which mutation events occurred. Although precise data were not compiled, mutation-bearing individuals gave rise to completely mutated, i.e., white-eyed, progeny at very high frequency, indicating functional germline transmission which is important for development of gene drives. Additionally, 87%  $\pm$  0.41 (616/712) of the offspring of OA-16 male/Cas9 female crosses and 60.5%  $\pm$  0.43 (637/1053) of the offspring from OA-16 female/Cas9 male crosses had a predominantly yellow cuticle (Fig. 5D, right), indicating that the second RGR was also functional, albeit with a significantly lower efficiency than the first. It should be noted that the efficiency of the second gRNA may be underestimated, as small mutant patches of yellow cuticle tissue are generally more difficult to detect than white patches of ommatidia facets in the compound eye. The presence of mutations was confirmed by sequencing of PCR products that span the cleavage site (Fig. 5E). Together, these data conclusively provide a proof-of-principle for the feasibility of the RGR approach as a method for multiplexing gRNAs in whole animals.

**Design requirements for multiplex number.** This demonstration of the multiplexing of gRNAs is encouraging for the same being achieved in insect species that transmit human diseases, such as *An. gambiae*. Multiplexing is expected to increase the effective homing rate, as only one of several target sites must have a functional copy of the homing allele to be capable of homing. That said, as depicted in Fig. 3, the probability of population elimination is independent of the homing rate for  $e > 98\%$ , but is highly dependent on the resistant allele generation rate,  $\rho$ . We derive the effective resistant allele generation rate for two and three multiplexed gRNAs in Supplementary Text S1 and find that this is approximately equal to  $\rho^2$  and  $\rho^3$ , respectively. This logically follows since resistant allele generation in the presence of multiplexing requires all gRNA target sites to have a homing-resistant allele. In Supplementary Text S1 we show that, although homing-resistant alleles may accumulate in a composite allele with multiple target sites, partially resistant composite alleles are rarely generated and are frequently converted to homing alleles in subsequent generations. The rate of completely resistant composite alleles emerging is therefore approximately equal to the rate of resistant alleles emerging at all target sites at once, i.e.  $\rho^m$  for a multiplex number of  $m$ .



**Figure 5.** RGR/Cas9-induced mutations at the yellow and white loci. **(A)** Schematic of the white and yellow genes showing the gRNA target sites. Exons are shown as blue boxes, the gRNA target site locations are indicated by black lines, and the gRNA target site sequences (with black letters indicating protospacer sequences and red letters indicating PAM sequences) are underlined in yellow for white, purple for yellow. **(B)** Schematic of the OA-16 construct utilized in generating mutations. The first and second gRNAs (targeting white and yellow, respectively) are shown in grey. Each gRNA has a hammerhead ribozyme 5' (shown in blue) and an HDV ribozyme 3' (shown in green). The gRNAs are driven by a single *Drosophila* ubiquitin polII promoter. **(C)** Crossing scheme used to generate mutants, and obtained results. Individual male and female flies homozygous for the OA-16 construct were crossed to individual female and male flies, respectively, of a homozygous vasa-Cas9 line. Progeny were scored for eye and body color. Percentages correspond to number of flies out of total cross progeny (+/- SEM) exhibiting a mutation for each gRNA. **(D)** The white and yellow tissues of an OA-16 homozygous male fly with no exposure to Cas9 are un-mutated (left), while the white and yellow tissues of a fly generated by crossing OA-16 homozygotes to vasa-Cas9 homozygotes show mosaic expression (right). **(E)** Examples of sequences of CRISPR/Cas-induced mutations in white (top) and yellow (bottom). The first line in each alignment represents wild-type sequence, and subsequent lines show individual mutant clones.

The effective resistant allele generation rate therefore becomes exponentially smaller as the number of multiplexed gRNAs increases. For a baseline  $\rho$  value of 0.13%, the effective  $\rho$  value becomes  $\sim 1.7 \times 10^{-6}$  for a multiplex number of two, and  $\sim 2.2 \times 10^{-9}$  for a multiplex number of three. Following on from Equation 1, for a baseline resistant allele generation rate of  $\rho$  at a single site, the multiplex number,  $m$ , we must achieve to have a chance  $x$  of eliminating a population of size  $N$  is given by:

$$m > (\log c_x - \log N) / \log \rho. \quad (2)$$

This means that, to have a 90% chance of eliminating a population of size 10,000, we require a multiplex number of two; for a population of size 1 million, we require a multiplex number of three; and for a population of size 10 billion, we require a multiplex number of four. An encouraging property of these predictions is that, as multiplex number increases linearly, the population size that we can eliminate increases exponentially. A modest additional increase in multiplex number can also lead to a much higher chance of eliminating a population of the same size. For instance, a multiplex number of five is predicted to provide a > 99% chance of eliminating an *An. gambiae* population on the scale of the African continent. Important spatial factors are not considered here.

In the multiplexing experiments demonstrating the RGR approach in *D. melanogaster*, a reduced efficiency was observed at the second target site as compared to the first, for reasons that are unclear (but that may or may not be tied to position within the RGR array<sup>12</sup>). Presumably this is something that could be solved through subsequent engineering efforts; however, mathematical analysis described in Supplementary Text S1 shows that, at least for the two-gRNA system, a reduced cleavage rate at one site doesn't significantly alter the effective resistant

allele generation rate overall. This is because a reduced cleavage rate at one site also suggests a reduced resistant allele generation rate at this site, as both H and R alleles are generated through the same cleavage and repair mechanism. However, this reduction in the resistant allele generation rate is compensated for by the increased accumulation of partially-resistant composite alleles and their subsequent development into completely resistant composite alleles. The net effect is that, even if one of two multiplexed gRNAs displays a reduced cleavage rate, the benefits of multiplexing in terms of reduced resistant allele generation are very similar.

## Discussion

The possibility of using gene drive systems to suppress and potentially eliminate wild populations has provoked intense interest over the last decade<sup>2,3,20–22</sup>. This excitement has recently been fueled by significant developments in genetic engineering, and by the CRISPR revolution, which has enabled scientists to develop homing-based gene drive systems targeting a range of sites in any genome with relative ease. In terms of population suppression, recessive lethal and sterility genes are of interest as targets because a homing system targeting these genes can potentially spread to fixation and eliminate the population in the process, even when introduced beginning with a single drive-containing organism<sup>4</sup>. While this excitement is warranted, it is highly relevant to determine how the evolution of homing-resistant alleles could interfere with population suppression drive strategies, and to determine design criteria to increase stability of the drive for the likely elimination of populations of a given size.

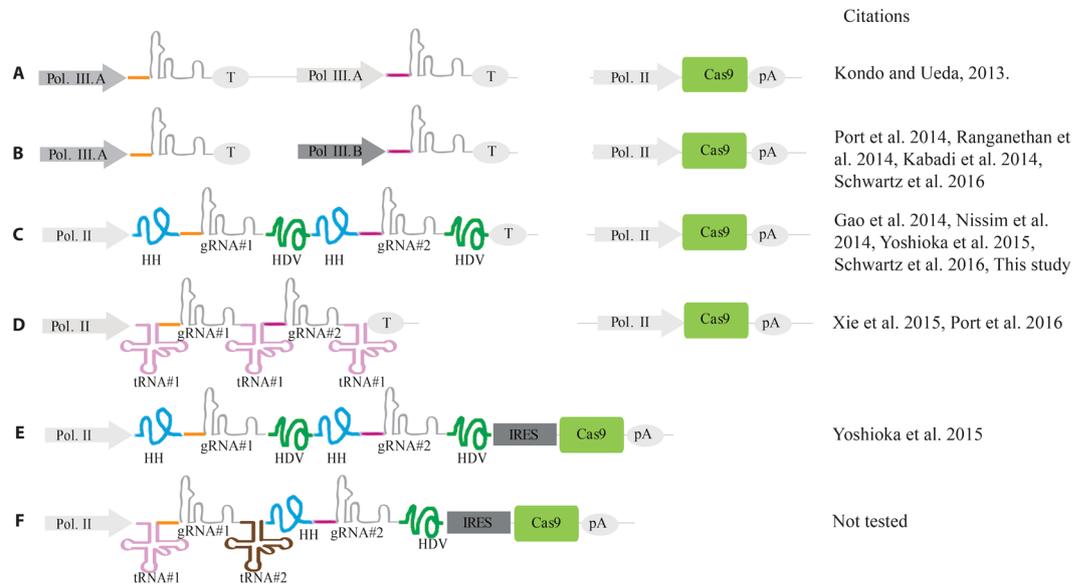
To address this important question, we developed a mathematical model to describe the spread of a population-suppressing homing allele through a population of *An. gambiae*, and the impact that a homing-resistant allele could have on these dynamics. Homing-resistant alleles may originate through several mechanisms: a) *de novo* mutations, which occur independently from the drive; b) pre-existing natural variation in the population, which may be minimized through intelligent selection of homing recognition sites; and c) in response to the drive, either by errors introduced during HDR or by the cell mending DNA damage via the NHEJ pathway. While the former two mechanisms are important, we have focused this study on the latter (i.e. resistant allele formation in response to the drive), as this is expected to occur at a significantly higher frequency than the other two mechanisms<sup>23</sup>. Firstly, *de novo* mutations occur at a rate of  $\sim 10^{-9}$ – $10^{-8}$  per individual per generation in species similar to that of interest<sup>24,25</sup> – a rate that is many orders of magnitude smaller than resistant allele generation in response to the drive. Secondly, the frequency of resistance alleles in the population prior to the introduction of the drive system may be predicted based on mutation-selection-drift balance<sup>26</sup>. Unckless *et al.*<sup>23</sup> have shown that pre-existing resistant alleles are likely when the effective population size exceeds  $10^6$ , a population size that would be several orders of magnitude higher for a composite allele. Furthermore, prior screening of target sites in wild populations could be used to identify pre-existing resistant alleles at frequencies less than  $10^{-3}$ , and at frequencies several orders of magnitude smaller for composite alleles. All of this highlights the significant relative importance of resistant allele generation in response to the drive.

We discover that, despite promising experimental data reporting extremely high rates of homing in the germline for recently engineered CRISPR-Cas9-based homing systems<sup>6,8,9</sup>, population suppression is expected to be moderate and short-lived with current construct architectures due to the generation of homing-resistant alleles. To minimize chances of a population rebound, current homing rates are adequate; however, reducing the resistant allele generation rate is crucial. For example, to be 95% sure that resistant alleles will not interfere with suppression of a population of *An. gambiae* mosquitoes on the scale of the African continent, the resistant allele generation rate should be less than  $\sim 2 \times 10^{-12}$  per homing event – about nine orders of magnitude smaller than current resistant allele generation rates<sup>9</sup>.

While it might be near impossible to achieve resistant allele generation rates this low with a single gRNA recognizing an exclusive target site, one strategy to mitigate the impact of homing-resistant alleles is to multiplex gRNAs in the drive system<sup>2,3,9,17</sup>. By multiplexing gRNAs to target multiple locations within an essential gene, each site is required to be homing-resistant in order for the composite allele to have the homing-resistant phenotype. Our results suggest that the effective resistant allele generation rate becomes exponentially smaller as the number of multiplexed gRNAs increases, and that a multiplex number of four may be sufficient to have a 90% chance of eliminating an *An. gambiae* population on the scale of the African continent.

Several approaches to multiplexing gRNAs have been described, including the use of different polIII promoters such as U6:1-U6:3<sup>27</sup>, HPI<sup>28,29</sup>, 7SK<sup>28</sup>, or tRNA promoters<sup>30</sup> to promote expression of individual gRNAs (Fig. 6A,B). While these strategies are effective, they are limited by the fact that most polIII promoters do not drive temporal and/or tissue-specific expression, which may incur increased fitness costs to the organism due to ubiquitous and continuous gRNA expression. These strategies also require an individual promoter element for each gRNA, thereby increasing the overall size of the drive and possibly introducing repetitive elements. Repetitive DNA sequences have reduced stability<sup>31</sup> and previous attempts to build drives with zinc-finger nucleases and TALENs have indeed demonstrated that larger and more repetitive the drive systems are less evolutionarily stable. Therefore, it is essential to minimize the size and repetitiveness of the drive<sup>32</sup>. To circumvent the need to express each gRNA from a different polIII promoter, gRNAs can be flanked with self-cleaving ribozymes<sup>13,14,30,33</sup> or tRNAs<sup>12,27,30,34</sup>, which can allow the use of a single temporal and tissue-specific polII promoter to drive expression of an array of flanked gRNAs, thus reducing the overall drive size and repetitiveness of the drive element.

To date, only the tRNA-flanked gRNA multiplex approach has been validated in a whole animal model<sup>12</sup>. Here, we report that the adaptation of another multiplex RGR approach, previously demonstrated in yeast<sup>13</sup> and in mammalian tissue culture cells<sup>14</sup>, functions efficiently in *D. melanogaster*, at least with respect to the ability to be processed somatically and induce mutations via NHEJ. We observe highly efficient somatic mutation rates (as measured by the percentage of individuals possessing mutations) approaching 100% for the first gRNA and of 60%–86% for the second gRNA, depending on whether Cas9 is either maternally or paternally inherited (Fig. 5), and a high frequency of mutation transmission to the offspring of mutant individuals. While these observations do not directly indicate that the RGR approach will effectively function to induce HDR in the germline,



**Figure 6.** Schematic of various proposed strategies for multiplexing gRNAs. **(A)** A gRNA multiplexing scheme where the same polIII promoter drives each of two gRNAs, and a polII-driven Cas9 is provided as a separate transgene. **(B)** A multiplexing scheme where two different polIII promoter drive each of two gRNAs, and a polII-driven Cas9 is provided as a separate transgene. **(C)** A multiplexing scheme where each of two gRNAs are surrounded by a 5' HH ribozyme and a 3' HDV ribozyme, and a polII-driven Cas9 is provided as a separate transgene. **(D)** A multiplexing scheme where the two gRNAs are surrounded by copies of the same tRNA (with a tRNA 5' of the first gRNA, between gRNAs 1 and 2, and 3' of the second gRNA), and a polII-driven Cas9 is provided as a separate transgene. **(E)** A multiplexing scheme where each of two gRNAs are surrounded by a 5' HH ribozyme and a 3' HDV ribozyme, as in **(C)**, but the Cas9 is located on the same transgene, 3' of the gRNAs and preceded by an IRES. **(F)** A proposed multiplexing scheme where the first of two gRNAs is surrounded by two different tRNAs, the second gRNA is flanked by the HH and HDV ribozymes (as in **(C)** and **(E)**), and the Cas9 is located on the same transgene, 3' of the gRNAs and preceded by an IRES (as in **(E)**). Grey triangles represent polIII promoters; blue triangles are polII promoters; terminators (T) and polyA signals (pA) are shown in grey ovals; Cas9 is represented as a green rectangle; internal ribosomal entry sequences (IRES) are grey sequences; gRNA scaffolds are shown as grey lines, with red and purple connecting lines representing two different gRNAs; the hammerhead (HH) and HDV ribozymes are shown as blue and green lines, respectively; and two different tRNAs (tRNA<sup>Gly</sup> and a non-specific tRNA) are shown as pink and brown lines, respectively.

they at least indicate that ribozyme-flanked gRNAs are processed *in vivo* and can direct efficient mutagenesis. Importantly, while we use the same two ribozymes to flank each gRNA (Figs 5B and 6C), the RGR approach may be expanded in the future to incorporate different ribozymes than the two tested here<sup>35,36</sup> to reduce repetitiveness, and the same strategy could be applied to the use of tRNAs other than the tRNA<sup>Gly</sup> (Fig. 6D) that was shown to work by Port *et al.*<sup>12</sup>. Furthermore, as recently demonstrated by Yoshioka *et al.*<sup>14</sup> in mammalian cells, the RGR approach could also be optimized to allow for expression of both the multiplexed gRNAs and CRISPR-Cas9 components from a single polII promoter, further reducing drive size and repetitiveness (Fig. 6E). Utilization of this single promoter-gRNA-CRISPR/Cas9 strategy combined with both the tRNA and RGR approaches may yield an optimal drive element design, both in terms of efficiency and stability (Fig. 6F). Finally, it may be important to utilize improved gRNA backbones to further reduce repetitiveness of the gRNAs<sup>37</sup>. Overall, the above design considerations may offer opportunities to engineer compact, evolutionarily stable gene drive cassettes; however, these ideas are largely untested and the use of multiplexed gRNAs in a functional gene drive system remains to be demonstrated.

Several assumptions have been made in the modeling portion of this study. Most noteworthy is the description of an *An. gambiae* population on the scale of the African continent as randomly mixing. Clearly, the study of gene drive in *An. gambiae* at anything beyond the village scale will require an understanding of population structure, and in fact, even at the village scale there are population considerations regarding gene flow within the *An. gambiae* species complex<sup>38,39</sup>. By ignoring population structure, the model described here cannot be used to predict the spatial pattern or timescale of gene drive<sup>40</sup>. However, even in the presence of population structure, each target site on a chromosome represents an opportunity for homing resistance to emerge and, in this sense, we expect there to be some validity to predictions regarding the resistant allele generation rate required to ensure that a population rebound unlikely (Supplementary Figure 3). Recent work by Eckhoff *et al.*<sup>41</sup> has modeled homing-based gene drive systems in a spatially-explicit manner with seasonality based on Namawala, Tanzania and the Garki District, Nigeria. For a seasonal population size of  $10^3$ – $10^5$  mosquitoes, a critical NHEJ rate of  $10^{-5}$  was observed, which is in agreement with our model<sup>41</sup>. One potential impact of population structure is the emergence of resistance to one guide RNA in one population and to another guide RNA in another population rarely combining due to low rates of gene flow. However, the suggestion in Supplementary Text S1 that partially resistant composite

alleles are rare due to their low rate of emergence and quick conversion to functional homing alleles would argue against this concern.

At the molecular level, several assumptions have been made regarding the dynamics of multiplexed gRNAs. These have been modeled as independently-acting homing systems; however, the use of multiplexed gRNAs in a functional gene drive system has yet to be demonstrated and hence these dynamics will be elucidated in future drive experiments. Potential problems may arise from sequence repetitiveness in the drive element if identical gRNA backbones and promoters are used (Fig. 6), creating the possibility of recombination between identical sequences<sup>12</sup> and thus reducing the overall evolutionary stability of the system<sup>42</sup>. Furthermore, it is not clear how the cleavage and homing rates will vary as multiplex number is increased substantially. If the gRNA target sites are far away from each other (e.g., >1–5 kb), then it is theoretically possible that multiplexing may not be an effective strategy as homology arms may not be able to effectively direct HDR (Supplementary Figure 4A–D), while if they are close together (e.g., <1 kb), multiplexing may increase homing effectiveness (Supplementary Figure 4E–H), as homology arms may be able to effectively direct HDR, although this hypothesis remains to be demonstrated. Interestingly, a reduction in the homing rate associated with one of the gRNAs may not interfere with our design criteria for population suppression; however, it is important that we have a good quantitative understanding of the underlying molecular dynamics of multiplexed gRNAs to make accurate model predictions.

In conclusion, multiplexing gRNAs appears to be a highly effective strategy by which to reduce the effective resistant allele generation rate and hence to enable the elimination of large populations. Due to the exponential decrease in resistant allele generation with increasing multiplex number, only a modest number of gRNAs are needed to achieve population suppression potentially on a continental scale. These approaches need to be tested in relevant organisms to accurately describe their dynamics and to confirm their utility for population suppression strategies. Future studies should address additional sources of resistant alleles, such as *de novo* mutations and naturally-occurring genetic variation. Additional strategies for overcoming resistance should also be explored, for instance, engineering successive gene drive systems, each designed to target different essential genes, and releasing these one after the other<sup>1</sup>. While both this approach and gRNA multiplexing may be effective for overcoming resistance, neither has been demonstrated. Given how quickly this field is advancing, understanding strategies such as these should be of high priority so that the full potential of homing-based population suppression drives can be properly evaluated.

## Materials and Methods

**Modeling CRISPR-Cas9 population genetics.** To characterize the basic dynamics of the autosomal CRISPR-Cas9-based gene drive system targeting a gene required for female fertility<sup>9</sup>, we represent the CRISPR homing construct as a single autosomal allele, “H”, with a corresponding wild-type allele, “h”. We denote homing-resistant alleles (mutant alleles with no associated fitness cost) as “R”. The CRISPR construct creates a bias among gametes of both parental heterozygotes towards gametes having the CRISPR allele. We consider the case where the homing rate is the same among Hh males and females and denote this as  $e$ . We also consider a resistant allele generation rate that is identical among Hh males and females and denote this as  $\rho$ . The homing rate,  $e$ , is the proportion of h gametes in heterozygotes that become H gametes due to the act of homing, and hence the proportion of H gametes arising from heterozygotes of both sexes is equal to  $(1 + e)/2$ . The resistant allele generation rate,  $\rho$ , is the proportion of h gametes that become R gametes due to errors introduced during the DNA breakage and repair process, and hence the proportion of R gametes arising from heterozygotes of both sexes is equal to  $\rho/2$ . The remaining gametes arising from heterozygotes are wild-type, h. This leads to an increase in the frequency of the H allele in the population and, since HH females are infertile, there is potential for a population crash to occur under permissive conditions. However, since the R allele represents resistance to homing and hence resistance to the spread of the H allele, it has a selective advantage following emergence and is expected to reverse the effects of population suppression. The crosses describing this system are shown in Supplementary Figure 1 and effective resistant allele generation rates for higher multiplex numbers are derived in Supplementary Text S1.

**Modeling *An. gambiae* population dynamics.** Using *An. gambiae* as a case study, we adapt the modeling framework of Deredec *et al.*<sup>4</sup> to describe the spread of the CRISPR and homing-resistant alleles through a discrete, density-dependent population with time steps of one day. In this model, the mosquito life cycle is divided into four life stages – egg, larva, pupa and adult (both male and female adults are modeled). The daily, density-independent mortality rates for the juvenile stages are assumed to be identical and are chosen for consistency with the population growth rate in the absence of density-dependent mortality, while the duration of these stages differ. Additional density-dependent mortality occurs at the larval stage, and we use a density-dependent equation of the form,  $F(L) = T_L \sqrt{\alpha / (\alpha + L)}$ , where  $L$  is the number of larvae,  $T_L$  is the duration of the larval stage, and  $\alpha$  is a parameter determining the strength of density-dependence which is chosen to produce the desired equilibrium density of adult mosquitoes in the population. The form of the density-dependence equation is taken from Deredec *et al.*<sup>4</sup>, and may be thought of as a discrete version of the logistic growth model. It has the desirable feature that, while density-independent life parameters determine the population growth rate, a single parameter,  $\alpha$ , determines the equilibrium density of adults (the value of this parameter may be determined from Equation S55, Supplementary Text S1). Adult males mate throughout their lifetime, while adult females mate only once, soon after that they emerge. Fecundity rates differ per genotype, with wild-type females laying  $\beta$  eggs per day, females heterozygous for the homing allele laying  $\beta(1 - s)$  eggs per day, HH females being infertile, and females of all other genotypes laying  $\beta$  eggs per day. Here,  $s$  represents the fractional reduction in fertility of females heterozygous for the homing allele. Initial estimates for these and other parameter values are provided in Supplementary Table 1. Equations describing this system are provided in Supplementary Text S1.

We use a stochastic implementation of this model to capture the random effects at low population sizes, for instance when the CRISPR-Cas9 system is causing significant population suppression. We assume that the number of eggs produced per day by females follows a Poisson distribution, the number of eggs having each genotype follows a multinomial distribution, and all survival/death events follow a Bernoulli distribution. Finally, female mate choice follows a binomial distribution with probabilities given by the relative frequency of each male genotype in the population.

**Construct Assembly.** Gibson enzymatic assembly (EA) cloning method was used for all cloning<sup>43</sup>. To generate plasmid OA-16, components were cloned into the multiple cloning site (MCS) of a commonly used plasmid in the lab for *D. melanogaster* transformation that contains the *white* gene as a marker and an attB-docking site. Specifically, the *Drosophila* ubiquitin promoter<sup>19</sup> was amplified from *D. melanogaster* genomic DNA using primers OA16-1 and OA16-2, and the SV40 3'UTR fragment was amplified from template pMos-3xP3-DsRed-attp (addgene plasmid #52904) using primers OA16-3 and OA16-4. The two RGRs were generated via sequential PCRs using primers OA16-5 and OA16-6 for the first PCR and OA16-5 and OA16-7 for the *white* RGR, and primers OA16-8 and OA16-5 for the first PCR and OA16-9 and OA16-10 for the *yellow* RGR. The construct was assembled in one step: the *D. melanogaster* attB stock plasmid was digested with AscI and XbaI, and the ubiquitin promoter, *white* RGR, *yellow* RGR, and the SV40 3'UTR were cloned in via EA cloning. A list of primer sequences used in the above construct assembly can be found in Supplementary Table 2.

**Fly Culture and Strains.** Fly husbandry and crosses were performed under standard conditions at 25°C. Rainbow Transgenics (Camarillo, CA) carried out all of the fly injections. The OA-16 construct was integrated into Bloomington Stock Center (BSC) fly strain 86Fa (BSC #24485:  $y^1 M\{vas-int.Dm\}ZH-2A^w$ ;  $M\{3xP3-RFP.attP\}ZH-68E$ ), and fly stock BSC#51324 ( $w[1118]$ ;  $PBac\{y[+mDint2]=vas-Cas9\}VK00027$ ) was used as the source of *vasa*-Cas9. For balancing chromosomes, fly stocks BSC#39631 ( $w[*]$ ;  $wg[Sp-1]/CyO$ ;  $P\{ry[+7.2]=neoFRT\}82B\ln[SS6]/TM6C$ ,  $Sb[1]$ ) and BSC#2555 ( $CyO/sna[ScO]$ ) were used. Homozygous stocks were first generated for 86Fa-OA-16 flies via use of balancer flies. Then, single homozygous female virgins and males were crossed out in triplicate to single male and female virgins, respectively, from the *vasa*-Cas9 line. The offspring (1765 in total) were scored for body color and eye color. The standard error of the mean (SEM) was calculated for each cross type and each phenotype using standard procedures.

**Sequencing to confirm mutations.** Genomic DNA was extracted from single mutant flies using Qiagen DNeasy Blood and Tissue Kit (Hilden, Germany), and PCRs were set up using standard protocols to amplify regions of the *white* (primer set OA16-S1/OA16-S2) and *yellow* (primer set OA16-S3/OA16-S4) genes that span the cleavage site. Sequencing was performed by Source Bioscience (Nottingham, UK). Primer sequences can be found in Supplementary Table 2.

## References

- Burt, A. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc. Biol. Sci.* **270**, 921–928, doi:10.1098/rspb.2002.2319 (2003).
- Champer, J., Buchman, A. & Akbari, O. S. Cheating evolution: engineering gene drives to manipulate the fate of wild populations. *Nat. Rev. Genet.* **17**, 146–159, doi:10.1038/nrg.2015.34 (2016).
- Esvelt, K. M., Smidler, A. L., Catteruccia, F. & Church, G. M. Concerning RNA-guided gene drives for the alteration of wild populations. *Elife* **3**, e03401, doi:10.7554/eLife.03401 (2014).
- Deredec, A., Burt, A. & Godfray, H. C. J. The population genetics of using homing endonuclease genes in vector and pest management. *Genetics* **179**, 2013–2026, doi:10.1534/genetics.108.089037 (2008).
- Deredec, A., Godfray, H. C. J. & Burt, A. Requirements for effective malaria control with homing endonuclease genes. *Proc. Natl. Acad. Sci. USA* **108**, E874–80, doi:10.1073/pnas.1110717108 (2011).
- Gantz, V. M. & Bier, E. Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science* **348**, 442–444, doi:10.1126/science.aaa5945 (2015).
- DiCarlo, J. E., Chavez, A., Dietz, S. L., Esvelt, K. M. & Church, G. M. Safeguarding CRISPR-Cas9 gene drives in yeast. *Nat. Biotechnol.* **33**, 1250–1255, doi:10.1038/nbt.3412 (2015).
- Gantz, V. M. *et al.* Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proc. Natl. Acad. Sci. USA* **112**, E6736–43, doi:10.1073/pnas.1521077112 (2015).
- Hammond, A. *et al.* A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nat. Biotechnol.* **34**, 78–83, doi:10.1038/nbt.3439 (2016).
- Alphey, L. Genetic control of mosquitoes. *N. Biotechnol.* (2016).
- Noble, C., Olejarz, J., Esvelt, K., Church, G. & Nowak, M. Evolutionary dynamics of CRISPR gene drives. *Sci. Advances* doi:10.1126/sciadv.1601964 (2017).
- Port, F. & Bullock, S. L. Augmenting CRISPR applications in *Drosophila* with tRNA-flanked sgRNAs. *Nat. Methods* doi:10.1038/nmeth.3972 (2016).
- Gao, Y. & Zhao, Y. Self-processing of ribozyme-flanked RNAs into guide RNAs *in vitro* and *in vivo* for CRISPR-mediated genome editing. *J. Integr. Plant Biol.* **56**, 343–349, doi:10.1111/jipb.v56.4 (2014).
- Yoshioka, S., Fujii, W., Ogawa, T., Sugiura, K. & Naito, K. Development of a mono-promoter-driven CRISPR/Cas9 system in mammalian cells. *Sci. Rep.* **5**, 18341, doi:10.1038/srep18341 (2015).
- Bruce-Chwatt, L. J. The Garki Project. Research on the epidemiology and control of malaria in the Sudan savanna of West Africa. *Trans. R. Soc. Trop. Med. Hyg.* **75**, 190–191, doi:10.1016/0035-9203(81)90085-7 (1981).
- Takken, W. & Scott, T. W. *Ecological Aspects for Application of Genetically Modified Mosquitoes*. (Springer Science & Business Media, 2003).
- Galizi, R. *et al.* A CRISPR-Cas9 sex-ratio distortion system for genetic control. *Sci. Rep.* **6**, 31139, doi:10.1038/srep31139 (2016).
- Bassett, A. R., Tibbit, C., Ponting, C. P. & Liu, J.-L. Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. *Cell Rep.* **4**, 220–228, doi:10.1016/j.celrep.2013.06.020 (2013).
- Akbari, O. S., Oliver, D., Eyer, K. & Pai, C.-Y. An Entry/Gateway<sup>®</sup> cloning system for general expression of genes with molecular tags in *Drosophila melanogaster*. *BMC Cell Biol.* **10**, 1–11, doi:10.1186/1471-2121-10-8 (2009).
- Sinkins, S. P. & Gould, F. Gene drive systems for insect disease vectors. *Nat. Rev. Genet.* **7**, 427–435, doi:10.1038/nrg1870 (2006).
- Burt, A. Heritable strategies for controlling insect vectors of disease. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130432–20130432, doi:10.1098/rstb.2013.0432 (2014).

22. Gould, F. & Fred, G. Broadening the Application of Evolutionarily Based Genetic Pest Management. *Evolution* **62**, 500–510, doi:10.1111/j.1558-5646.2007.00298.x (2008).
23. Unckless, R. L., Clark, A. G. & Messer, P. W. Evolution of resistance against CRISPR/Cas9 gene drive. *bioRxiv* 058438, doi:10.1101/058438 (2016).
24. Haag-Liautard, C. *et al.* Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature* **445**, 82–85, doi:10.1038/nature05388 (2007).
25. Keightley, P. D. *et al.* Estimation of the Spontaneous Mutation Rate in *Heliconius melpomene*. *Mol. Biol. Evol.* **32**, 239–243, doi:10.1093/molbev/msu302 (2014).
26. Hermisson, J. Soft Sweeps: Molecular Population Genetics of Adaptation From Standing Genetic Variation. *Genetics* **169**, 2335–2352, doi:10.1534/genetics.104.036947 (2005).
27. Port, F., Chen, H.-M., Lee, T. & Bullock, S. L. Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **111**, E2967–76, doi:10.1073/pnas.1405500111 (2014).
28. Kabadi, A. M., Ousterout, D. G., Hilton, I. B. & Gersbach, C. A. Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic Acids Res.* **42**, e147–e147, doi:10.1093/nar/gku749 (2014).
29. Ranganathan, V., Wahlin, K., Maruotti, J. & Zack, D. J. Expansion of the CRISPR–Cas9 genome targeting space through the use of H1 promoter-expressed guide RNAs. *Nat. Commun.* **5**, 4516, doi:10.1038/ncomms5516 (2014).
30. Schwartz, C. M., Hussain, M. S., Blenner, M. & Wheelodon, I. Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR–Cas9-Mediated Genome Editing in *Yarrowia lipolytica*. *ACS Synth. Biol.* **5**, 356–359, doi:10.1021/acssynbio.5b00162 (2016).
31. Bzymek, M. & Lovett, S. T. Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. *Proc. Natl. Acad. Sci. USA* **98**, 8319–8325, doi:10.1073/pnas.111008398 (2001).
32. Simoni, A. *et al.* Development of synthetic selfish elements based on modular nucleases in *Drosophila melanogaster*. *Nucleic Acids Res.* **42**, 7461–7472, doi:10.1093/nar/gku387 (2014).
33. Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P. & Lu, T. K. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Mol. Cell* **54**, 698–710, doi:10.1016/j.molcel.2014.04.022 (2014).
34. Xie, K., Minkenberg, B. & Yang, Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci. USA* **112**, 3570–3575, doi:10.1073/pnas.1420294112 (2015).
35. Roth, A. *et al.* A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nat. Chem. Biol.* **10**, 56–60, doi:10.1038/nchembio.1386 (2014).
36. Wilson, T. J. & Lilley, D. M. J. Do the hairpin and VS ribozymes share a common catalytic mechanism based on general acid-base catalysis? A critical assessment of available experimental data. *RNA* **17**, 213–221, doi:10.1261/rna.2473711 (2011).
37. Thyme, S. B., Akhmetova, L., Montague, T. G., Valen, E. & Schier, A. F. Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nat. Commun.* **7**, 11750, doi:10.1038/ncomms11750 (2016).
38. Taylor, C. *et al.* Gene flow among populations of the malaria vector, *Anopheles gambiae*, in Mali, West Africa. *Genetics* **157**, 743–750 (2001).
39. Tripet, F., Dolo, G. & Lanzaro, G. C. Multilevel analyses of genetic differentiation in *Anopheles gambiae* s.s. reveal patterns of gene flow important for malaria-fighting mosquito projects. *Genetics* **169**, 313–324, doi:10.1534/genetics.104.026534 (2005).
40. North, A., Burt, A., Godfray, H. C. J. & Buckley, Y. Modelling the spatial spread of a homing endonuclease gene in a mosquito population. *J. Appl. Ecol.* **50**, 1216–1225, doi:10.1111/1365-2664.12133 (2013).
41. Eckhoff, P. A., Wenger, E. A., Godfray, H. C. J. & Burt, A. Impact of mosquito gene drive on malaria elimination in a computational model with explicit spatial and temporal dynamics. *Proc. Natl. Acad. Sci. USA* **114**, E255–E264, doi:10.1073/pnas.1611064114 (2017).
42. Jack, B. R. *et al.* Predicting the Genetic Stability of Engineered DNA Sequences with the EFM Calculator. *ACS Synth. Biol.* **4**, 939–943, doi:10.1021/acssynbio.5b00068 (2015).
43. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345, doi:10.1038/nmeth.1318 (2009).

## Acknowledgements

This work was supported by generous University of California, Riverside (UCR) laboratory start-up funds awarded to O.S.A., a private donation by [www.MaxMind.com](http://www.MaxMind.com) awarded to O.S.A., a US National Institutes of Health (NIH) K22 grant (5K22AI113060-02) and an NIH-R21 grant (R21AI123937-01) awarded to O.S.A., a UC MEXUS grant (CN-15-47) awarded to J.M.M., and a gift from The Parker Foundation to the University of California, San Francisco, Global Health Group Malaria Elimination Initiative.

## Author Contributions

J.M.M. and H.M.S. developed mathematical models with input from O.S.A. O.S.A. and A.B. developed and tested the ribozyme multiplexing technique in *Drosophila melanogaster*. All authors contributed to the writing of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-02744-7

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017