Improved CRISPR-based suppression gene drives mitigate resistance and impose a large reproductive load on laboratory-contained mosquito populations

Andrew Hammond, Kyros Kyrou, Matthew Gribble, Xenia Karlsson, Ioanna Morianou, Roberto Galizi, Andrea Beaghton, Andrea Crisanti* & Tony Nolan*

*corresponding authors

Abstract

CRISPR-based genes drives bias their own inheritance and can be used to modify entire populations of insect vectors of disease as a novel form of sustainable disease control. Gene drives designed to interfere with female fertility can suppress populations of the mosquito vector of malaria, however laboratory demonstrations showed strong unintended fitness costs and high levels of resistant mutations that limited the potential of the first generation of gene drives to spread. We describe three new gene drives designed to restrict spatio-temporal nuclease expression by using novel regulatory sequences. Two of the three new designs dramatically improve fitness and mitigate the creation and selection of resistance. We dissect the relative contributions of germline CRISPR activity versus embryonic CRISPR activity resulting from parental deposition, showing that the improved performance of the new designs is due to tighter germline restriction of the nuclease activity and significantly lower rates of end-joining repair in the embryo. Moreover, we demonstrate in laboratory-contained population experiments that these gene drives show remarkably improved invasion dynamics compared to the first generation drives, resulting in greater than 90% suppression of the reproductive output and a delay in the emergence of target site resistance, even at a loosely constrained target sequence. These results illustrate important considerations for gene drive design and will help expedite the development of gene drives designed to control malaria transmission in Africa.

Introduction

Gene drives and malaria control

Gene drives are genetic elements that have an enormous potential to modify natural populations by biasing their own inheritance, allowing their autonomous spread from a low initial release frequency. The technology has been proposed as a novel method to reduce the harmful impact of pest and vector species, including mosquitoes, ticks and rats. Conventional interventions such as insecticide spraying and the use of insecticide-treated bednets that target the mosquito vector of malaria have been responsible for the halving the world's malaria burden over the last 15 years, yet these gains are being stalled by the emergence of insecticide-resistance. Gene drives have many attractive features for control of malaria in Africa since they do not require substantial infrastructure, they can spread into areas that are difficult to reach with conventional interventions, and the short generation time of mosquitoes would allow a rapid increase in frequency and concomitant alteration of the population. Unlike other genetic modifications, gene drives can spread even if they confer a significant cost to the host organism, since the bias of their inheritance can be so strong as to still result in a net overall increase in frequency each generation.

Driving endonuclease genes

Gene drives based on site-specific endonucleases were first proposed by Austin Burt (Burt, 2003). In principle several classes of endonuclease, providing they have sufficient specificity, can be converted into a driving endonuclease gene (DEG). DEGs, with the CRISPR-Cas9 nuclease being the most favoured in recent years due to the ease with which its specificity can be re-programmed must be

encoded within their own recognition sequence in the genome so that this sequence cannot be cut when the DEG is inside, but homologous chromosomes lacking it are susceptible to cleavage. Therefore, when an unmodified chromosome comes into contact with the DEG-carrying chromosome, it is cut by the DEG. The broken chromosome is usually repaired using the DEG-containing chromosome as a template and, by the process of homology-directed repair, the DEG is copied inside the broken chromosome. The process, called homing, converts a DEG-heterozygote to homozygosity. If this reaction takes place in the germline then inheritance of the DEG is biased because sperm or eggs produced in the germline will inherit the DEG from, either the original DEG-carrying chromosome, or the newly modified DEG chromosome. By targeting the DEG to disrupt a haplosufficient gene that is required for female fertility, it is possible to drive a trait causing population suppression throughout a mosquito population.

Previous limitations of CRISPR-based gene drives

We and others have recently designed several gene drives designed to suppress or modify mosquito populations (Hammond et al., 2016, Gantz and Bier, 2015, Gantz et al., 2015). With any gene drive the force of selection for resistance to the drive will be proportional to the fitness cost imposed by the drive. In the case of population suppression approaches this selection falls on the mosquito - we recently designed several CRISPR-based gene drives targeting female fertility genes for population suppression and demonstrated their spread throughout caged populations of the malaria mosquito, *Anopheles gambiae* (Hammond et al., 2016) however the drive was eventually replaced in the population by mutations that were generated by end-joining repair in the fraction of cleaved chromosomes that were not modified by homing ((Hammond et al., 2017)).

In determining the propensity for target site resistance to arise, two issues are paramount – 1) functional constraint at the target site that will determine the degree of variation that can be tolerated there; 2) the relative contribution of the error-prone end-joining repair pathway to introduce mutation at the target site during repair of the broken chromosome. The target site sequence described in Hammond et al. (2016) is poorly conserved, suggesting little functional constraint, and therefore is particularly prone to accommodating resistant alleles that restore function to the target gene ('r1' alleles) (Supplementary Figure 1) (Hammond et al., 2017). Moreover, cleavage by maternally deposited nuclease in the embryo is believed to be the major source of these mutations in current gene drive designs (Gantz et al., 2015, Hammond et al., 2017, Champer et al., 2017) and we have previously observed high levels of embryonic end-joining following strong maternal deposition using the *vas2* promoter (Hammond et al., 2017). Additionally, using the same promoter, unrestricted activity of the endonuclease outside of the mosquito germline caused a strong unintended fitness cost in females harbouring a single copy of the gene drive, due to conversion of the soma to the homozygosity for a null allele. These effects retarded the invasive potential of the drive and increased the selection pressure for resistant mutations (Hammond et al., 2017).

Given the limitations of the promoter in previous gene drive constructs we decided to test a suite of novel germline promoters for their ability to restrict nuclease activity to the germline. We deliberately tested their activity at the previously validated female fertility locus that was prone to resistance, in order that we could quantify any reduction in the creation or selection of resistant mutations that resulted from reducing embryonic end-joining and improving female fertility, respectively.

Results

Choice of germline promoters

To find alternatives to the *vas2* promoter, we investigated three mosquito genes, AGAP006241 (*zero population growth*, *zpg*), AGAP006098 (*nanos*, *nos*) and AGAP007365 (*exuperantia*, *exu*), that are expressed in the germline of males and female *Anopheles gambiae* (Baker et al. 2011) and may show reduced somatic expression or deposition into the embryo.

The gene *zero population growth* is specifically expressed in the germline of male and female fruitflies, where it plays a role in forming gap junctions between the developing germline and cyst cells (Tazuke et al., 2002). Crucially, deposited mRNA concentrates at the germ plasm, due to regulatory signals present on the untranslated regions (Tazuke et al., 2002, Rangan et al., 2009). In *Anopheles gambiae*, the ortholog of *zpg* is AGAP006241 and has not been used to direct transgene expression however it is known to play an essential role in male and female gonad development (Magnusson et al., 2011). *Exuperantia* and *nanos* are maternal effect genes in Drosophila that are transcribed in the oocyte and deposited into the early embryo where they eventually localise to the germ plasm (Wang and Lehmann, 1991, Berleth et al., 1988). Nanos mRNA is trafficked to the pole plasm by the action of several other maternal effect genes, and it is later incorporated into the pole cells that eventually form the germline (Wang et al., 1994). Regulatory sequences on the 3' UTR of *nanos* further restrict translation of maternal mRNAs to the germline (Rangan et al., 2009). The promoter region of *exuperantia* from the tiger mosquito, *Aedes aegytpi*, has been used to drive robust expression in both male and female germlines (Akbari et al., 2014). In contrast to *zpg* and *exu*, several reports have suggested that *nos* is specific to the female germline, however use of the promoter in *Aedes aegypti* and *Anopheles gambiae* have found low level expression in males that was attributed to incomplete recapitulation of endogenous gene expression (Adelman et al., 2007, Meredith et al., 2013, Macias et al., 2017).

Generation of new gene drives designed to restrict spatiotemporal expression to the mosquito germline

The p165 *CRISPR*^{h} vector used in Hammond et al. (2016) was modified to contain Cas9 under control regulatory sequences flanking upstream and downstream of AGAP006241 (*zpg*), AGAP006098 (*nos*) and AGAP007365 (*nos*) to create vectors p173, p174 and p175, respectively. Each vector was secondarily modified to contain a gRNA targeted to AGAP007280 (**Figure 1**). gRNA-modified *CRISPR*^{h} vectors were inserted by recombinase-mediated cassette exchange into a pre-existing "*hdrGFP* docking site" at the target site in AGAP007280 (Hammond et al., 2016). G1 transformants were identified by a change in fluorescent signal form GFP in the docking line to DsRed linked the gene drive. The new gene drive strains were named *nos-CRISPR*^{h}, *zpg-CRISPR*^{h}, *exu-CRISPR*^{h} (*nos, zpg* and *exu* respectively).

Zpg and *nos* promoters of drive high levels of homing in the germline

Phenotypic assays were performed to measure simultaneously the fecundity and transmission rates associated with each of three drives (**Figure 2**). To assess the level of homing, drive heterozygotes were crossed to wild-type, allowed to lay individually, and their progeny scored for the presence of DsRed linked to the construct. Maternally- or paternally-deposited Cas9 can cause resistant mutations in the embryo that may reduce the rate of homing in the next generation(Champer et al., 2018, Champer et al., 2017, Gantz et al., 2015, Hammond et al., 2017). To test the magnitude of any parental effects, we separated male and female drive heterozygotes according to whether they

inherited the drive from their mother or father and scored inheritance of the drive in their progeny (**Figure 2**). In the case of the *zpg-CRISPR*^{h} gene drive we observed transmission rates that exceeded 90.6%+/-1.8% s.e.m.) in males and 97.8%+/-0.6% in females, falling only slightly below previously observed rates for *vas2-CRISPR*^{h} (99.6% in males and 97.7% in females) (Hammond et al., 2016). We saw no significant maternal or paternal effect with this construct. On the contrary with the *nos-CRISPR*^{h} construct, though homing rates were still high, we saw a reduction in homing performance in individuals that inherited the drive allele from their mother, suggesting that *nos::Cas9* is maternally deposited – transmission rates in males or females receiving the allele from their father were 99.1%+/-0.3% and 99.6%+/-0.3%, respectively, compared to 83.6%%+/-5.0% and 85.2%+/-5.0% when receiving it from their mother. *exu-CRISPR*^{h} showed only modest homing rates in males (65%+/-2.0% transmission rate) and no homing in females, with no obvious parental effect.

New gene drive constructs confer significantly less fecundity costs in heterozygous females than previous constructs

Fertility assays were performed to measure the larval output in individual crosses of drive heterozygotes to wild-type (**Figure 2**). All new drives showed a marked improvement in relative fecundity - where *vas-CRISPR*^{h} females showed approximately 8.4% relative female fecundity compared to wild type, the relative fecundity of *zpg-CRISPR*^{h} (50-58.3%), *nos-CRISPR*^{h} (40.2-55.9%), and *exu-CRISPR*^{h} (75.5-77.4%) females were much improved.

Embryonic end-joining is reduced in gene drives containing *zpg*, *nos* and *exu* promoters

To specifically examine the contribution of deposited nuclease to the generation of resistant alleles, we designed a screen such that the wild-type target allele in the embryo is only exposed to a paternal or maternal dose of the nuclease in the absence of a genetically encoded drive construct. To do this, we crossed individuals that had a gene drive allele balanced against a resistant R1 allele containing a 6bp deletion at the target site (203-GAGGAG) and screened for the offspring that did not inherit the gene drive allele (**Figure 3**). Two replicates were performed for each cross. As expected, maternal *vas2-CRISPR*^{h} generated high levels of novel end-joining mutations (21.3-24.8%) compared to just a small fraction in paternally exposed targets (1.8-3.5%). In contrast, *zpg-CRISPR*^{h} did not appear to generate any new target site mutations (0.5-0.9% vs 0.7% in the wild type control). Surprisingly, we observed a marked increase in the parental R1 allele in all crosses with *vasa* or *zpg* gene drives that was not observed in the control cross and may reflect embryonic 'homing' of the R1 allele into the wild type target site following CRISPR-mediated from deposited nuclease. To explain the apparent discrepancy in the ratio of end-joining to homing events caused nuclease deposited by *vas2-CRISPR*^{h} compared to *zpg-CRISPR*^{h}, we hypothesise that *vasa* causes Cas9 to be deposited as a protein whereas *zpg* may predominantly deposit Cas9 mRNA that is subsequently trafficked to the germ plasm whereby translated Cas9 cleaves within a spatio-temporal window that is biased towards homology-directed repair.

zpg-CRISPR^h spreads to close to fixation in caged releases and exerts a large reproductive load on the population

Given its improved performance in terms of germline restriction and propensity to generate a lower rate of end joining mutations at the target locus we investigated the potential for the *zpg-CRISPR*^{h} gene drive to spread throughout naïve populations of malaria mosquitoes. Two replicate cages were initiated with either 10% or 50% of drive heterozygotes, and monitored for 16 generations.

Remarkably, the drive spread to more than 97% of the population in all four replicates (**Figure 4**) and had achieved complete population modification in one of the two 50% release cages after just four generations. In all four releases, the drive sustained more than 95% frequency for at least 3 generations before its spread was reversed by the gradual selection of drive-resistant alleles. Notably, we observed similar dynamics of spread whether released at 50% or 10%, demonstrating that initial release frequency has little impact upon the potential to spread. These results are all the more surprising when compared to *vas2-CRISPR*^{h} targeted to the exact same locus at *AGAP007280* (Figure 4a). Here, the spread of the drive was slower and resistance arose before it reached 80% frequency in the population (Hammond et al., 2016).

A population suppression gene drive is designed to exert a considerable reproductive load on the population as it increases in frequency and sterile homozygous females are produced. To investigate this, we recorded the total egg output of the cage over time. From generation 4 onwards, we counted the number of eggs produced each generation to measure the level of population suppression (**Figure 4b**). Egg production was suppressed by an average of 92% in each cage at or shortly after the peak in drive frequency, representing a reduction from more than 15,000 eggs to under 1,200 eggs, with the greatest suppression seen in the two cages initiated from a 10% release. At its lowest point, one cage produced just 643 eggs. Each generation was maintained using 600 hatched larvae, rather than reducing the adult population size in proportion to the level of egg suppression, and so a reduction in adult population could only be observed when egg production was reduced by more than 96%. Less favourable environmental conditions and predation would likely allow wild populations to be eliminated with the high levels of egg suppression observed in these caged experiments.

A novel single generation resistance assay designed to deplete R2 non-functional mutations and reveal R1 functional resistant mutations

Several strategies have been proposed that aim to mitigate these resistant mutations and include: multiplexing the nuclease (Burt, 2003, Esvelt et al., 2014); using alternative nucleases (Hammond and Galizi, 2017); targeting highly conserved and constrained sequences (Burt, 2003, Hammond and Galizi, 2017); and targeting sites that bias non-resistant mutations during repair by microhomology-mediated end-joining (Hammond and Galizi, 2017). To help compare these strategies before initiating a lengthy caged release experiment, we have developed a novel assay to enrich and test putative resistant mutations in a single generation (**Figure 5**).

The assay is designed to first recover a high proportion of potential mutant alleles, and then test which of these are resistant mutations that can rescue against the gene drive allele that would otherwise confer sterility. Initially, we crossed large numbers of drive heterozygotes (F_O) to each other and used COPAS fluorescence based sorter to enrich for the rare fraction of F_1 progeny that were heterozygous for the drive. These F_1 heterozygotes contain one copy of the gene drive and one non-drive chromosome that was exposed to the nuclease and could contain a resistant mutation. Target site mutations that are resistant to cleavage can be divided into two types: those that do not restore function to the target gene (R2) and those (R1) that, at least partially, restore function to the target and may be subject to selection. To identify resistant mutations amongst these individuals, F_1 female heterozygotes were crossed to wild-type and allowed to give F_2 progeny. Because F_1 female heterozygotes contain a null copy the AGAP007280 fertility gene that is disrupted by the gene drive, they can only produce progeny if the non-drive chromosome contains a wild-type or resistant mutation. As such, we expected R2 mutations to be depleted in the F_2 progeny of these females. To measure this effect and characterise R1 and R2 mutations, we performed amplicon sequencing across the target site in pooled samples of F1 heterozygous females and their F_2 progeny. In the F_1,

we identified 14 mutant alleles, including a partial gene drive allele that likely resulted from an incomplete homing event, present in at least one replicate experiment at 1% frequency or more amongst modified, non-drive alleles. The nature and frequency of these mutations are similar to those previously generated by *vas2-CRISPR*^{h} at the target site in AGAP007280, and as before (Hammond et al. 2017), the most frequent mutations can be explained by micro-homology mediated end-joining (MMEJ) or incomplete homing (**Figure 5**). In each replicate, we observed a strong stratification towards known resistant mutations in the F_2 that was concomitant with at least 90% depletion of all frame-shift (R2) mutations, including near complete loss of the incomplete homing event.

Concluding statement

Building suppression gene drives for mosquito control that are fit for purpose will likely require a combination of the following features: optimal target sites that due to functional constraint cannot easily support variation; the targeting of multiple sites by the same gene drive construct, akin to combination therapy with drugs to slow resistance; fine tuning of the expression of the nuclease that serves as the gene drive's 'motor' in order that the drive shows the most efficient invasion dynamics. For this last aspect we have shown here that simple changes to the promoter controlling nuclease activity can lead to drastic improvements in the speed of invasion due to two principle reasons: 1) less propensity to induce end-joining repair in the germline or precursors thereof and 2) improved fecundity of 'carrier' females that contain a single copy of the gene drive and transmit it in a super-Mendelian fashion to their offspring.

Our data on the relative contributions of end-joining repair and homology-directed repair in the germline and in the early embryo will be particularly important for the modelling of expected performance of gene drive constructs. Likewise, the propensity for maternal and paternal deposition of the nuclease under control of the promoters of the drive constructs is an important variable that we have shown can vary greatly. There is also the possibility that both the end-joining/HDR ratio and the magnitude of parental effect might additionally be locus dependent to some extent(Champer et al., 2018). In this regard our development of a single generation assay for large scale screening to determine the rate at which functionally resistant alleles can be formed could be applied to find, relatively rapidly, the best promoter:target site combinations.

The improvements conferred by the *zpg* promoter meant that even though its target site was poorly conserved the gene drive was able to reach a high frequency in the population while at the same time imposing a very strong reproductive load due to disruption of the haplosufficient female fertility gene, before resistant alleles eventually became selected. Combined with a functionally constrained target site, the expected increase in efficacy afforded by such a gene drive should be considerable.

Methods

Amplification of promoter and terminator sequences

The published *Anopheles gambiae* genome sequence provided in Vectorbase (Giraldo-Calderon et al, 2015) was used as a reference to design primers in order to amplify the promoters and terminators of the three *Anopheles gambiae* genes: AGAP006098 (*nanos*), AGAP006241 (*zero population growth*) and AGAP007365 (*exuperantia*). Using the primers provided in Supplementary Table 1 we performed PCRs on 40 ng of genomic material extracted from wild type mosquitoes of the G3 strain using the

Wizard Genomic DNA purification kit (Promega). The primers were modified to contain suitable Gibson assembly overhangs (underlined) for subsequent vector assembly. Promoter and terminator fragments were 2092 bp and 601 bp for *nos*, 1074 bp and 1034 bp for *zpg*, and 849 and 1173 bp for *exu*, respectively. The sequences of all regulatory fragments can be found in Supplementary Table 2.

Generation of CRISPR^h drive constructs

We modified available template plasmids used previously in Hammond *et al.* (2016)[2] to replace and test alternative promoters and terminators for expressing the Cas9 protein in the germline of the mosquito. p16501, which was used in that study carried a human optimised Cas9 (hCas9) under the control of the *vas22* promoter and terminator, an RFP cassette under the control of the neuronal *3xP3* promoter and a U6:sgRNA cassette targeting the AGAP007280 gene in *Anopheles gambiae*.

The hCas9 fragment and backbone (sequence containing 3xP3::RFP and a U6::gRNA cassette), were excised from plasmid p16501 using the restriction enzymes XhoI+PacI and AscI+AgeI respectively. Gel electrophoresis fragments were then re-assembled with PCR amplified promoter and terminator sequences of *zpg*, *nos* or *exu* by Gibson assembly to create new *CRISPR^h* vectors named p17301 *(nos*), p17401 *(zpg)* and p17501 *(exu).*

Transformation of drive constructs into genome at AGAP007280

CRISPR^h constructs containing Cas9 under control of the *zpg*, *nos* and *exu* promoters were inserted into an *hdrGFP* docking site previously generated at the target site in AGAP007280 (Hammond et al., 2016). Briefly, *Anopheles gambiae* mosquitoes of the *hdrGFP-7280* strain were reared under standard conditions of 80% relative humidity and 28°C, and freshly laid embryos used for microinjections as described before (Fuchs et al., 2013). Recombinase-mediated cassette exchange (RCME) reactions were performed by injecting each of the new *CRISPR^h* constructs into embryos of the *hdrGFP* docking line that was previously generated at the target site in AGAP007280 (Hammond et al., 2016). For each construct, embryos were injected with solution containing *CRISPR^h* (400ng/µl) and a *vas2::integrase* helper plasmid (400ng/µl) (Volohonsky et al., 2015). Surviving G_0 larvae were crossed to wild type transformants identified by a change from GFP (present in the *hdrGFP* docking site) to DsRed linked the *CRISPR^h* construct that should indicate successful RCME.

Phenotypic assays to measure fertility and rates of homing

Heterozygous *CRISPR^h*/+ mosquitoes from each of the three new lines *zpg-CRISPR^h*, *nos-CRISPR^h*, *zpg-CRISPR^h*, were mated to an equal number of wild type mosquitoes for 5 days in reciprocal male and female crosses. Females were blood fed on anesthetized mice on the sixth day and after 3 days, a minimum of 40 were allowed to lay individually into a 25-ml cup filled with water and lined with filter paper. The entire larval progeny of each individual was counted and a minimum of 50 larvae were screened to determine the frequency of the DsRed that is linked to the *CRISPR^h* allele by using a Nikon inverted fluorescence microscope (Eclipse TE200). Females that failed to give progeny and had no evidence of sperm in their spermathecae were excluded from the analysis. Statistical differences between genotypes were assessed using the Kruskal-Wallis test.

Caged Experiments

The cage trials were performed following the same principle described before in Hammond et al. (2016). Briefly, heterozygous *zpg-CRISPR^h* that had inherited the drive from a female parent were mixed with age-matched wild type at L1 at 10% or 50% frequency of heterozygotes. At the pupal stage, 600 were selected to initiate replicate cages for each initial release frequency. Adult mosquitoes were left to mate for 5 days before they were blood fed on anesthetized mice. Two days after, the mosquitoes were left to lay in a 300 ml egg bowl filled with water and lined with filter

paper. Each generation, all eggs were allowed two days to hatch and 600 randomly selected larvae were screened to determine the transgenic rate by presence of DsRed and then used to seed the next generation. From generation 4 onwards, adults were blood-fed a second time and the entire egg output photographed and counted using JMicroVision V1.27. Larvae were reared in 2L trays in 500ml of water, allowing a density of 200 larvae per tray. After recovering progeny, the entire adult population was collected and entire samples from generation 0, 2, 5, and 8 were used for pooled amplicon sequence analysis.

Pooled amplicon sequencing

Pooled amplicon sequencing was performed essentially as described before in Hammond et al. (2017). Genomic DNA was mass extracted from pooled samples of mosquitoes using the Wizard Genomic DNA purification kit (Promega), and 90 ng of each used for PCR using KAPA HiFi HotStart Ready Mix PCR kit (Kapa Biosystems) in 50 ul reactions. For single generation resistance experiments and caged release experiments, a 332 bp locus spanning the target site was amplified using primers designed to include the Illumina Nextera Transposase Adapters (underlined), 7280-Illumina-F (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGAGAAGGTAAATGCGCCAC) and 7280-Illumina-R (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCGCTTCTACACTCGCTTCT).

For deposition testing experiments, a 196 bp locus spanning the target site was amplified using primers were designed to include the Illumina partial adapters (underlined), Illumina-AmpEZ-7280-F1 (ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTTAACTGTCTTGGTGGTGAGG) and Illumina-AmpEZ-7280-R1 (GACTGGAGTTCAGACGTGTGCTCTTCCGATCTCACGCTTAACGTCGTCGTTTC). PCR reactions were amplified for 20 cycles and subsequently processed and sequenced by PoloGGB or Genewiz using an Illumina MiSeq instrument.

Analysis of pooled amplicon sequencing

Pooled amplicon sequencing reads were analysed using CRISPResso software v1.0.8 (Pinello et al., 2016) using an average read quality threshold of 30. Insertions and deletions were included if they altered a window of 20 bp surrounding the cleavage site that was chosen on the basis of previously observed mutations at this locus (Hammond et al., 2017). Allele frequencies were calculated by summing individual insertion or deletion events across all haplotypes on which they were found. A large insertion event, representing incomplete homing of *CRISPR*^{h}, was found to occur outside of this window and its combined frequency across several haplotypes was calculated and included in the final frequency tables.

Single generation enrichment of resistant mutations

Male and female drive heterozygotes (F0) were crossed to each other, allowed to lay *en masse,* and their progeny screened for putative drive heterozygosity (F1) on the basis of low DsRed expression using COPAS (complex object parametric analyser and sorter), as in Marois et al. (2012)**.** A total of 246 COPAS-selected putative heterozygote females were crossed to an equal number of wild type males in four replicate cages (36, 70, 64 and 70 individuals in each replicate), allowed to lay, and their F2 progeny collected at L1. Genomic DNA was then mass extracted from entire samples of each replicate at F1 and F2 using the Wizard Genomic DNA purification kit (Promega) and used for pooled amplicon sequencing analysis as described before. Mutant alleles were included for analysis if over 1% frequency amongst non-reference (wild-type) alleles in any replicate of the F1 generation.

Deposition Testing

F1 heterozygotes containing a gene drive and resistant allele (GD/R1) were generated by crossing *zpg-CRISPR*^{h} or *vas2-CRISPR*^{h} to a resistant strain that is homozygous for the 203-GAGGAG R1 allele

at the target site in AGAP007280. At least 40 F1 heterozygotes were group mated to wild-type in reciprocal male and female crosses, and allowed to lay *en masse*. F2 progeny were screened for the absence of DsRed that is linked to the *CRISPR*^{h} allele and pooled together for mass genomic DNA extraction and pooled amplicon sequencing as described elsewhere.

Ethics statement

All animal work was conducted according to UK Home Office Regulations and approved under Home Office License PPL 70/6453.

Figure Legends

Figure 1. Target site and design of new *CRISPR*^{h} gene drives designed to express Cas9 under regulation of *zpg*, *nos* and *exu* germline promoters (**a**) The haplosufficient female fertility gene, AGAP007280, and its target site in exon 6 (highlighted in grey), showing the protospacer-adjacent motif (highlighted in teal) and cleavage site (red dashed line). (**b**) *CRISPR*^{h} alleles were inserted at the target in AGAP007280 using ϕC31-recombinase mediated cassette exchange (RCME). Each *CRISPR*^{h} RCME vector was designed to contain *Cas9* under transcriptional control of the *nos*, *zpg* or *exu* germline promoter, a gRNA targeted to AGAP007280 under the control of the ubiquitous *U6* PolIII promoter, and a *3xP3::DsRed* marker.

Figure 2. Gene drives designed to express Cas9 under regulation of *zpg*, *nos* and *exu* germline promoters show high rates of biased transmission and substantially improved fertility compared with the *vas2* promoter. Phenotypic assays were performed to measure fertility and transmission rates for each of three drives. The larval output was determined for individual drive heterozygotes crossed to wild-type (left), and their progeny scored for the presence of DsRed linked to the construct (right). Males and females were further separated by whether they had inherited the *CRISPR*^{h} construct from either a male or female parent. For example, ♂→♀ denotes progeny and transmission rates of a heterozygous *CRISPR*^{h} female that had inherited the drive allele from a heterozygous male. The average progeny count and transmission rate is also shown (± s.e.m.). High levels of homing were observed in the germline of *zpg-CRISPR*^{h} and *nos-CRISPR*^{h} males and females, however the *exu* promoter generated only moderate levels of homing in the germline of males but not females. Counts of hatched larvae for the individual crosses revealed improvements in the fertility of heterozygous females containing *CRISPR*^{h} alleles based upon *zpg*, *nos* and *exu* promoters compared to the *vas2* promoter. In each case, the average number of hatched larvae improved relative to wild-type controls, or equivalent *CRISPR*^{h} heterozygous males (whereby no fertility cost is anticipated). Phenotypic assays were performed on G2 and G3 for *zpg*, G3 and G4 for *nos*, and ~G15 for *exu*. ♀* denotes *vas2-CRISPR*^{h} females that were heterozygous with a resistance (R1) allele, these were used because heterozygous *vas2-CRISPR*^{h} females are usually sterile.

Figure 3. Deposited nucleases promotes embryonic HDR but not end-joining when expressed under zpg. (a) *zpg-CRISPR*^{h} or *vas2-CRISPR*^{h} were crossed to a homozygous resistant strain (R1 203-GAGGAG) to generate F1 heterozygotes containing a gene drive and resistant allele (GD/R1). F1 heterozygotes and R1 homozygotes (control) were crossed to wild-type and their non-drive F2 progeny analysed by pooled amplicon sequencing across the target site in AGAP007280. **(b)** A wide range of end-joining mutations were detected in the progeny of male (1.8-3.5%) and female (21.3-24.8%) *vas2-CRISPR*^{h}, but novel alleles present in *zpg-CRISPR*^{h} samples could not be distinguished from noise found in the control. Both *zpg-CRISPR*^{h} and *vas2-CRISPR*^{h} showed evidence of embryonic

HDR because the original R1 mutation (203-GAGGAG) was more frequent than predicted by Mendelian inheritance (red dashed line) that was observed in the control cross.

Figure 4. *zpg-CRISPR*^{h} drives spread throughout entire caged populations of the malaria mosquito and cause a substantial reduction in reproductive output. (a) Equal numbers of *CRISPR*^{h}*/+* and WT individuals were used to initiate replicate caged populations, and the frequency of drive-modified mosquitoes was recorded each generation by screening larval progeny for the presence of DsRed linked to the *CRISPR*^{h} construct. Solid lines show results from two replicate cages for *zpg* (black) and previous results for *vas2* (grey) (Hammond et al. 2017). Deterministic predictions are shown for *zpg* (black dashed line) and *vas2* (grey dashed line) based on observed parameter values for homing in males (*zpg* = 83.6%, *vas2* = 98.4%), homing in females (*zpg* = 93.4%, *vas2* = 98.4%), heterozygous female fitness (*zpg* = 50-58.3%, *vas2* = 9.3%), homozygous females completely sterile, and assuming no fitness cost in males. **(b)** A lower release rate of 10% *CRISPR*^{h}*/+* was used to initiate two further replicate populations in which the frequency of drive-modified mosquitoes (solid line) and counts of the entire egg progeny (dashed line) were recorded each generation.

Figure 5. A single generation assay to enrich drive-resistant mutations. (a) A minimum of 100 males and 100 females heterozygous for *zpg-CRISPR*^{h} (F_0) were crossed to each other and allowed to lay *en masse*. High rates of homing in the F_0 mean that only a small fraction of their progeny will contain a target site that is unmodified or mutated by end-joining. To enrich for these individuals, F_1 progeny were screened for putative drive heterozygosity on the basis of low DsRed expression using COPAS (complex object parametric analyser and sorter). Females that are heterozygous for a drive allele balanced across an end-joining mutation will be fertile only if that mutation is functionally resistant to the drive (i.e. an R1 mutation). Selected F_1 females were group mated to wild-type, allowed to lay *en masse,* and their F_2 progeny collected. To reveal putative R1 resistant mutations that are enriched in the F_2, the nature and frequency of target site alleles was analysed by pooled amplicon sequencing in COPAS-selected F_1 females and their F_2 progeny. **(b)** We identified 14 mutant alleles present in at least one replicate at 1% frequency or more amongst modified, non-drive alleles. Highlighted in red are the in-frame mutations. The F_2 show a substantial stratification towards several in-frame mutations, including the major R1 resistant mutations previously selected after caged release of *vas2-CRISPR*^{h} (203-GAG, 203-GAGGAG, and 202-TGAGGA).

References

ADELMAN, Z. N., JASINSKIENE, N., ONAL, S., JUHN, J., ASHIKYAN, A., SALAMPESSY, M., MACCAULEY, T. & JAMES, A. A. 2007. nanos gene control DNA mediates developmentally regulated transposition in the yellow fever mosquito Aedes aegypti. *Proc Natl Acad Sci U S A,* 104**,** 9970-5.

AKBARI, O. S., PAPATHANOS, P. A., SANDLER, J. E., KENNEDY, K. & HAY, B. A. 2014. Identification of germline transcriptional regulatory elements in Aedes aegypti. *Sci Rep,* 4**,** 3954.

BERLETH, T., BURRI, M., THOMA, G., BOPP, D., RICHSTEIN, S., FRIGERIO, G., NOLL, M. & NUSSLEIN-VOLHARD, C. 1988. The role of localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo. *EMBO J,* 7**,** 1749-56.

BURT, A. 2003. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc Biol Sci,* 270**,** 921-8.

CHAMPER, J., LIU, J., OH, S. Y., REEVES, R., LUTHRA, A., OAKES, N., CLARK, A. G. & MESSER, P. W. 2018. Reducing resistance allele formation in CRISPR gene drive. *Proc Natl Acad Sci U S A,* 115**,** 5522-5527.

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 Genet,* 13**,** e1006796.

FUCHS, S., NOLAN, T. & CRISANTI, A. 2013. Mosquito transgenic technologies to reduce Plasmodium transmission. *Methods Mol Biol,* 923**,** 601-22.

GANTZ, V. M. & BIER, E. 2015. Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science,* 348**,** 442-4.

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. *Proc Natl Acad Sci U S A,* 112**,** E6736-43.

HAMMOND, A., GALIZI, R., KYROU, K., SIMONI, A., SINISCALCHI, C., KATSANOS, D., GRIBBLE, M., BAKER, D., MAROIS, E., RUSSELL, S., BURT, A., WINDBICHLER, N., CRISANTI, A. & NOLAN, T. 2016. A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nat Biotechnol,* 34**,** 78-83.

HAMMOND, A. M., KYROU, K., BRUTTINI, M., NORTH, A., GALIZI, R., KARLSSON, X., KRANJC, N., CARPI, F. M., D'AURIZIO, R., CRISANTI, A. & NOLAN, T. 2017. The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLoS Genet,* 13**,** e1007039.

MACIAS, V. M., JIMENEZ, A. J., BURINI-KOJIN, B., PLEDGER, D., JASINSKIENE, N., PHONG, C. H., CHU, K., FAZEKAS, A., MARTIN, K., MARINOTTI, O. & JAMES, A. A. 2017. nanos-Driven expression of piggyBac transposase induces mobilization of a synthetic autonomous transposon in the malaria vector mosquito, Anopheles stephensi. *Insect Biochem Mol Biol,* 87**,** 81-89.

MAGNUSSON, K., MENDES, A. M., WINDBICHLER, N., PAPATHANOS, P. A., NOLAN, T., DOTTORINI, T., RIZZI, E., CHRISTOPHIDES, G. K. & CRISANTI, A. 2011. Transcription regulation of sex-biased genes during ontogeny in the malaria vector Anopheles gambiae. *PLoS One,* 6**,** e21572.

MAROIS, E., SCALI, C., SOICHOT, J., KAPPLER, C., LEVASHINA, E. A. & CATTERUCCIA, F. 2012. High-throughput sorting of mosquito larvae for laboratory studies and for future vector control interventions. *Malar J,* 11**,** 302.

MEREDITH, J. M., UNDERHILL, A., MCARTHUR, C. C. & EGGLESTON, P. 2013. Next-generation site-directed transgenesis in the malaria vector mosquito Anopheles gambiae: self-docking strains expressing germline-specific phiC31 integrase. *PLoS One,* 8**,** e59264.

PINELLO, L., CANVER, M. C., HOBAN, M. D., ORKIN, S. H., KOHN, D. B., BAUER, D. E. & YUAN, G. C. 2016. Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat Biotechnol,* 34**,** 695-7.

RANGAN, P., DEGENNARO, M., JAIME-BUSTAMANTE, K., COUX, R. X., MARTINHO, R. G. & LEHMANN, R. 2009. Temporal and spatial control of germ-plasm RNAs. *Curr Biol,* 19**,** 72-7.

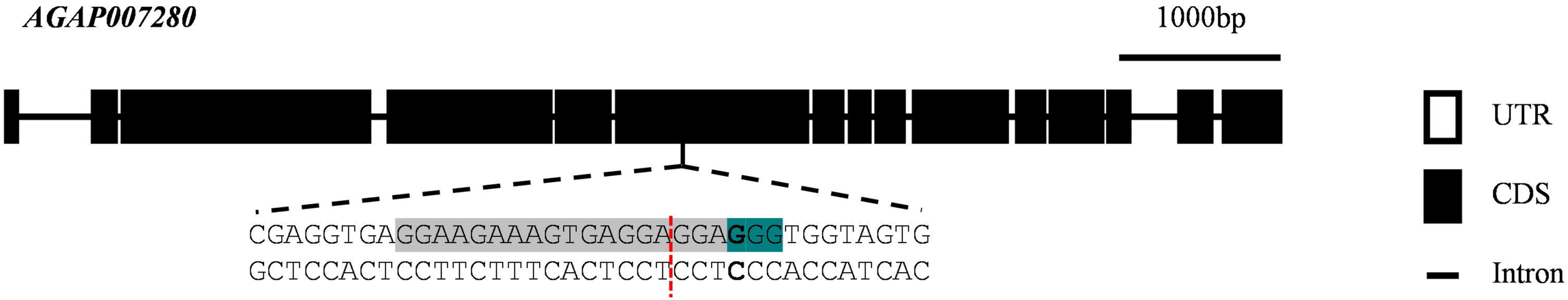
TAZUKE, S. I., SCHULZ, C., GILBOA, L., FOGARTY, M., MAHOWALD, A. P., GUICHET, A., EPHRUSSI, A., WOOD, C. G., LEHMANN, R. & FULLER, M. T. 2002. A germline-specific gap junction protein required for survival of differentiating early germ cells. *Development,* 129**,** 2529-39.

VOLOHONSKY, G., TERENZI, O., SOICHOT, J., NAUJOKS, D. A., NOLAN, T., WINDBICHLER, N., KAPPS, D., SMIDLER, A. L., VITTU, A., COSTA, G., STEINERT, S., LEVASHINA, E. A., BLANDIN, S. A. & MAROIS, E. 2015. Tools for Anopheles gambiae Transgenesis. *G3 (Bethesda),* 5**,** 1151-63.

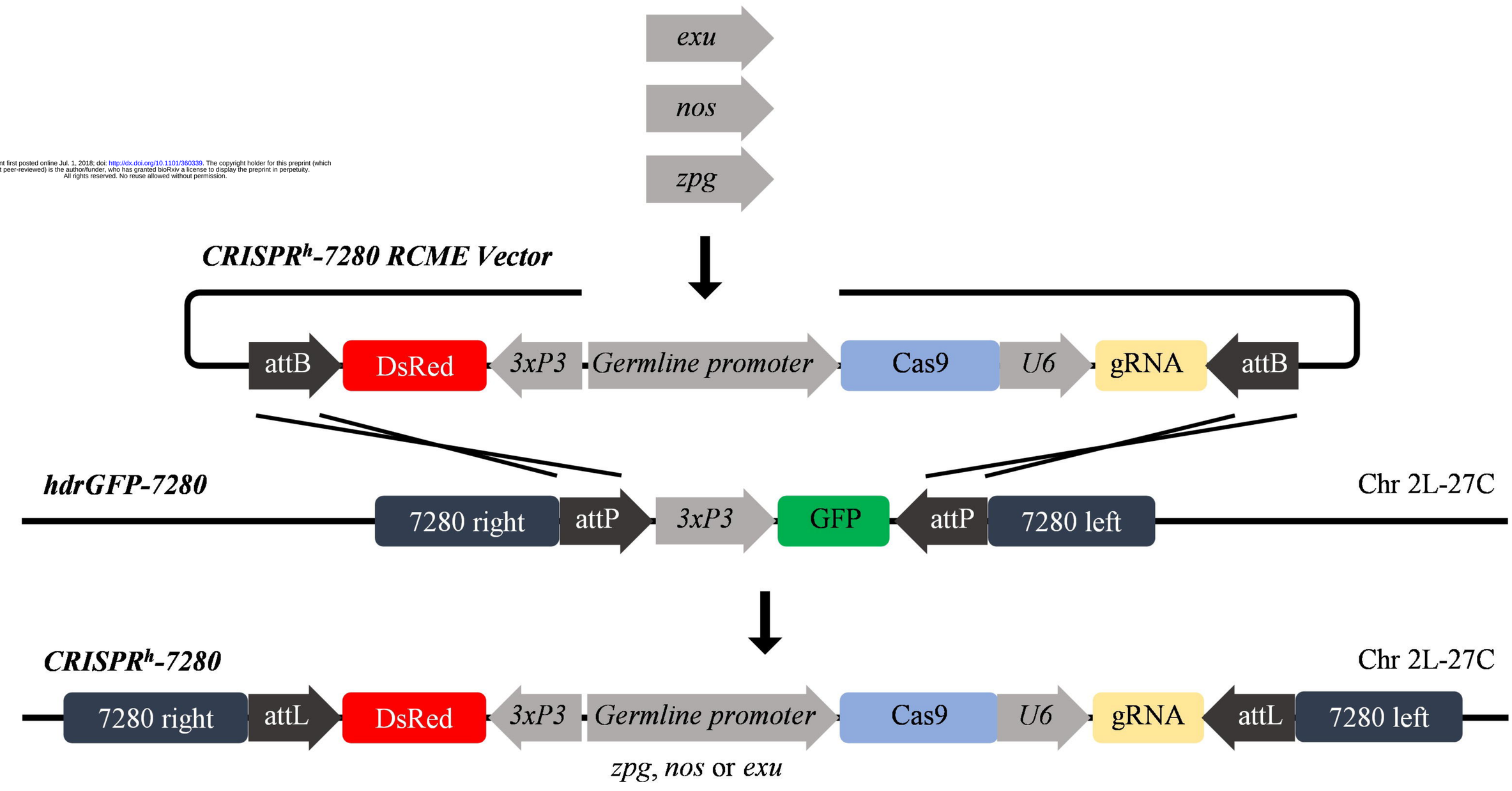
WANG, C., DICKINSON, L. K. & LEHMANN, R. 1994. Genetics of nanos localization in Drosophila. *Dev Dyn,* 199**,** 103-15.

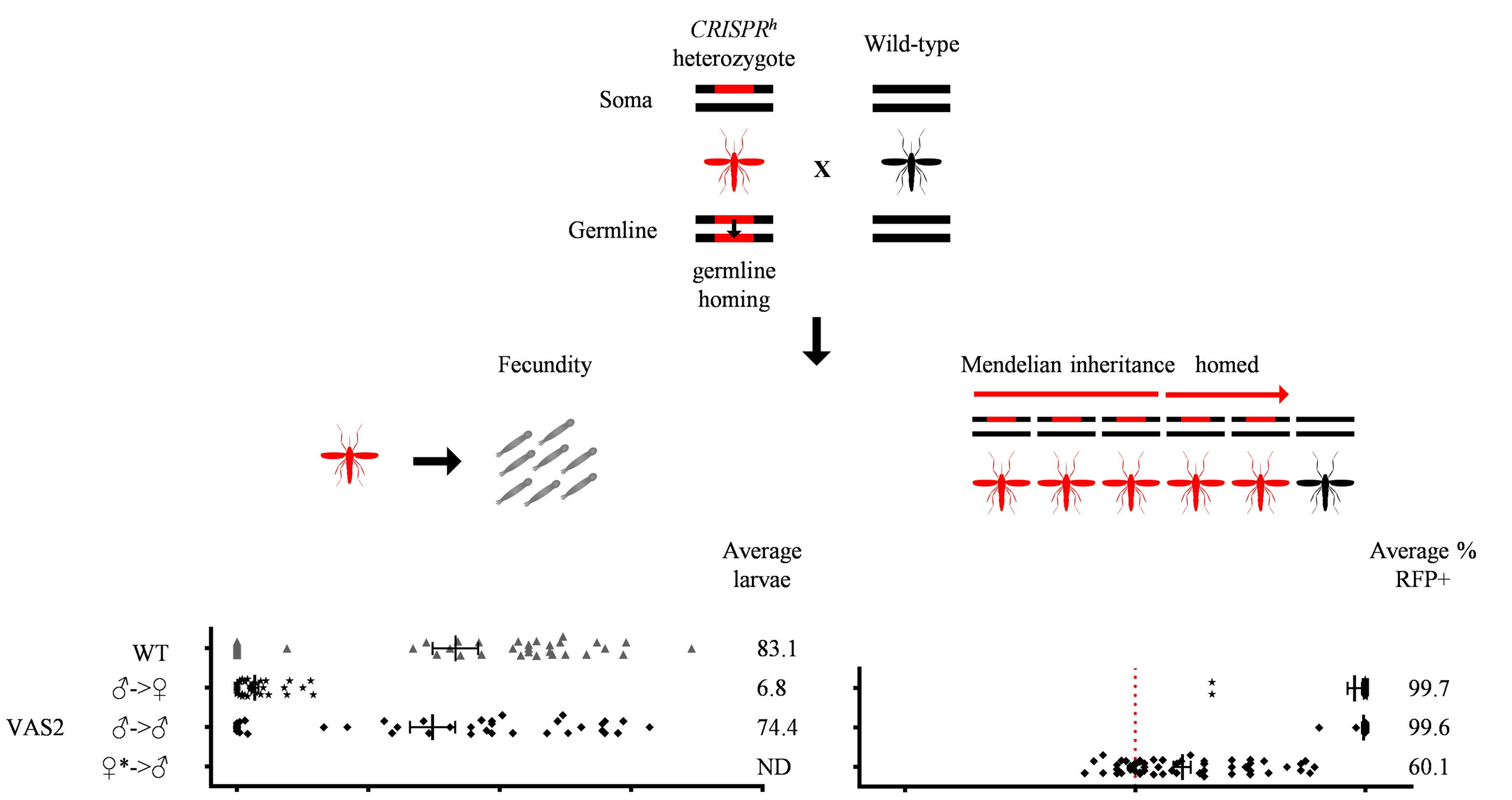
WANG, C. & LEHMANN, R. 1991. Nanos is the localized posterior determinant in Drosophila. *Cell,* 66**,** 637-47.

A

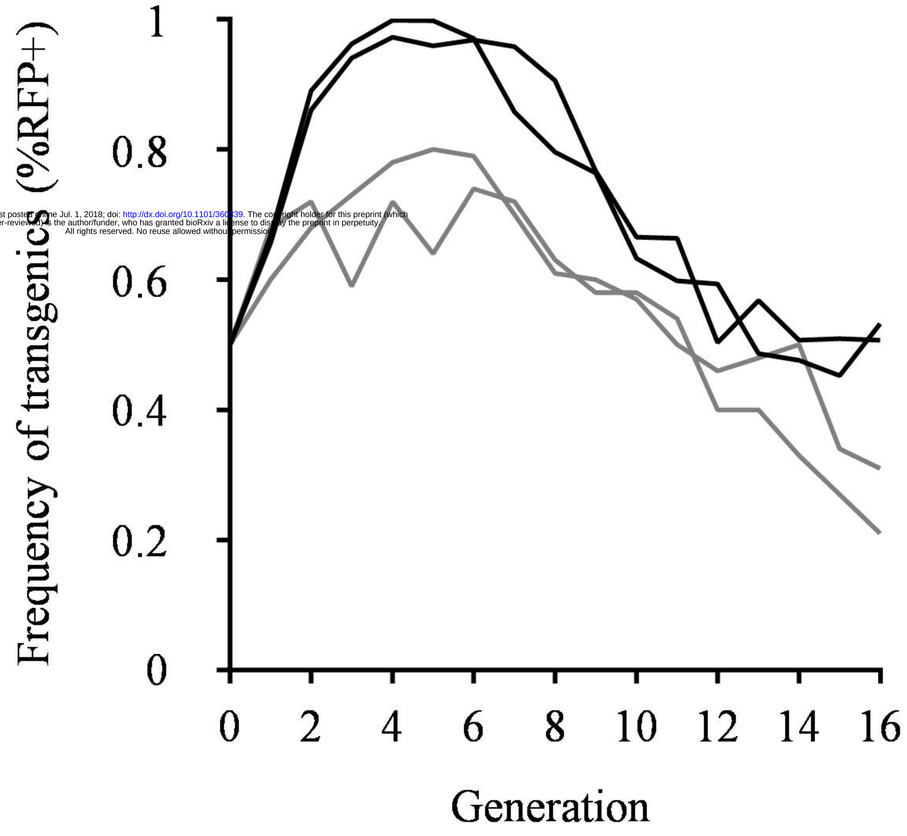


B

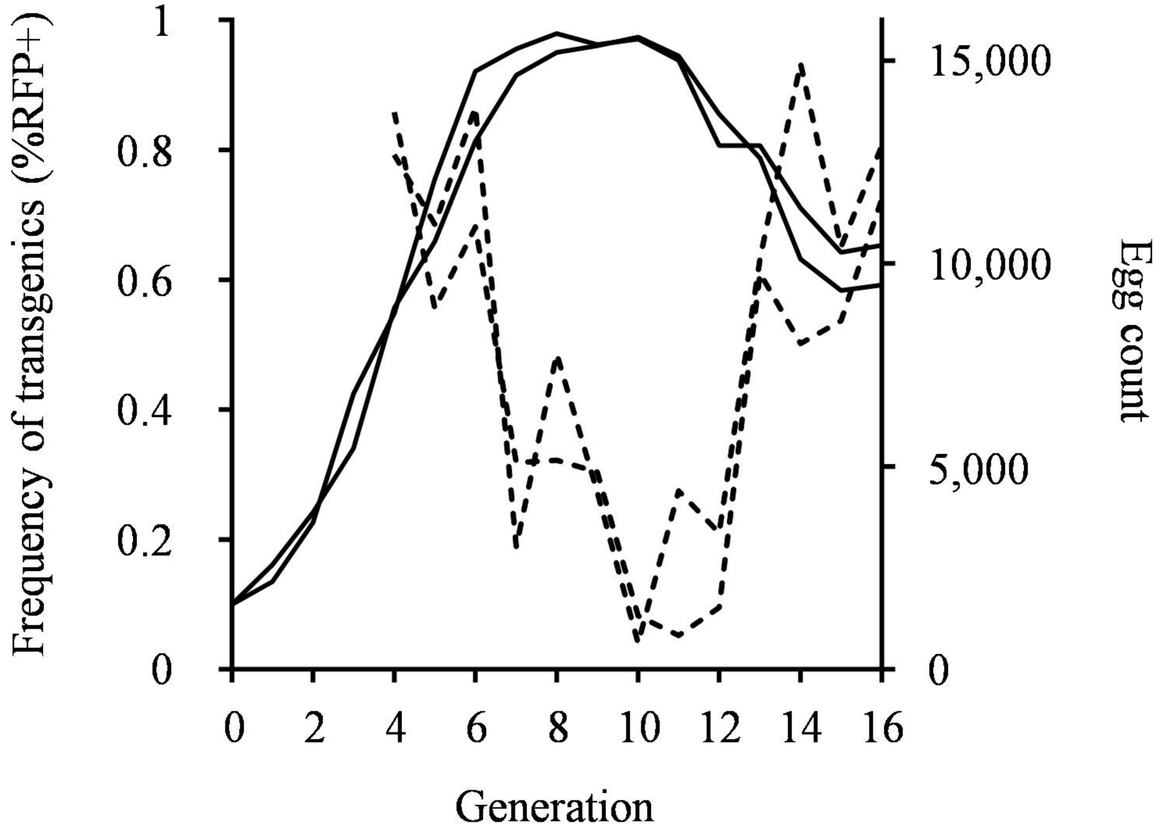




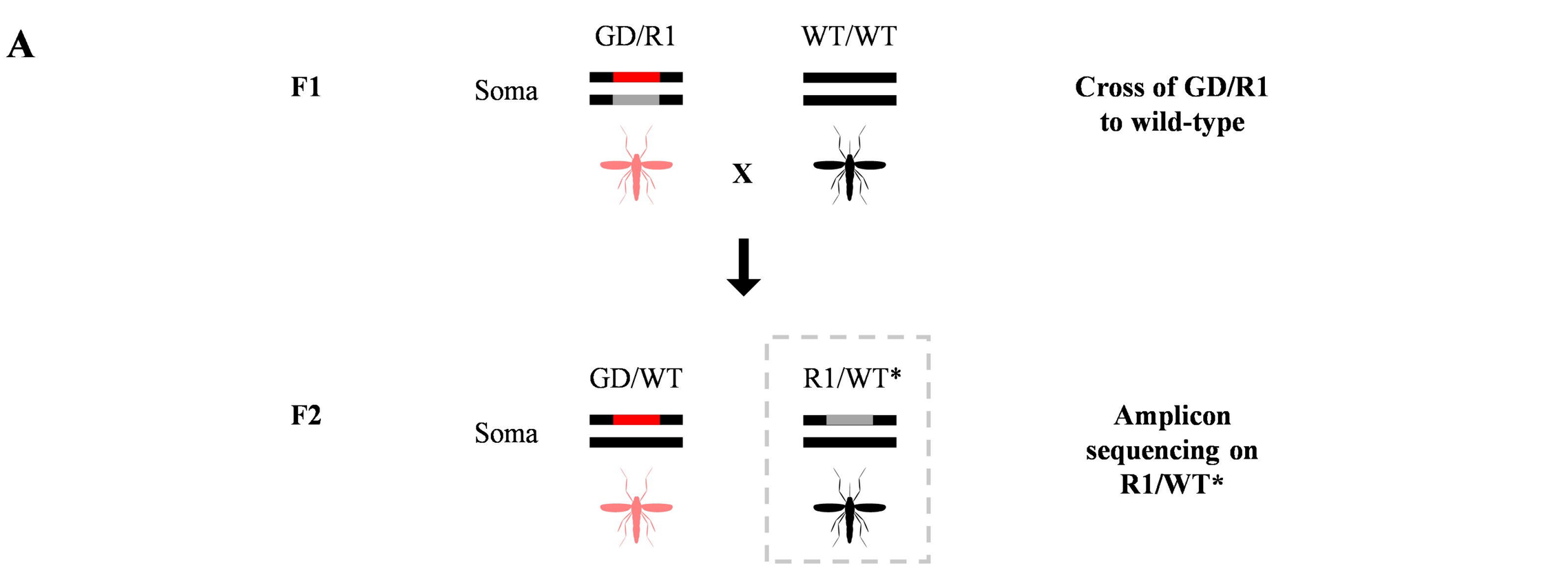
A



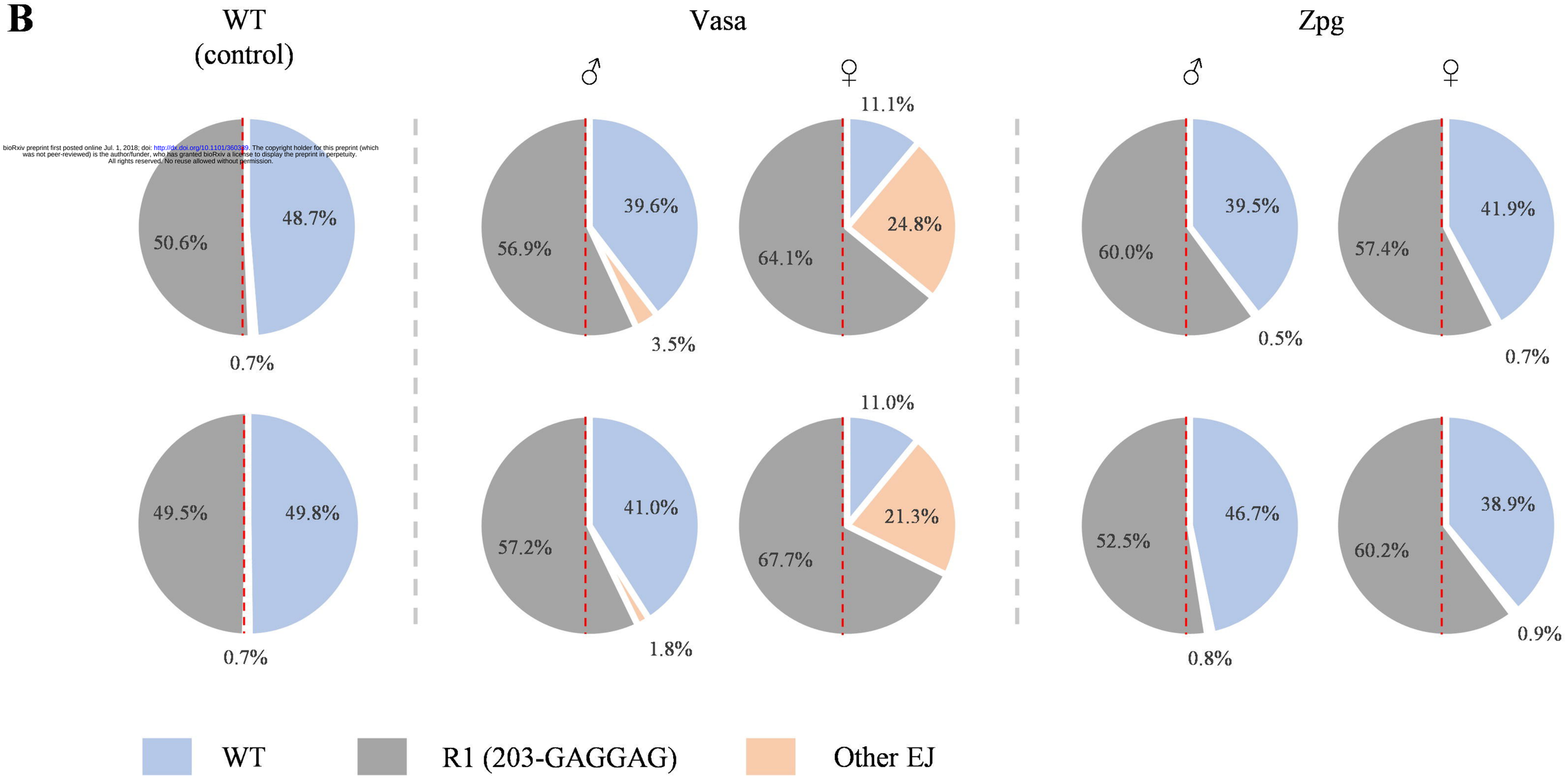
B

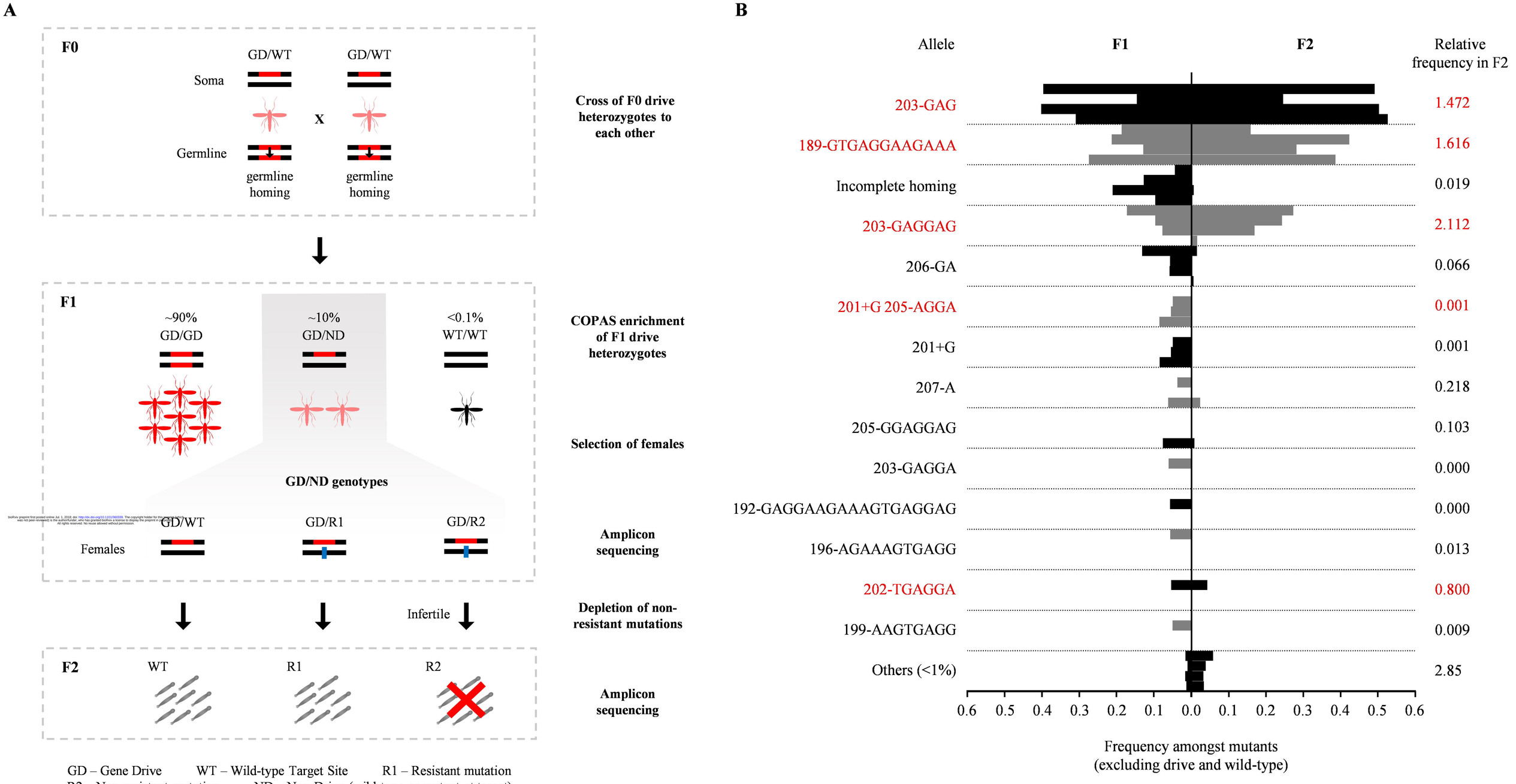


A



B




A
F0
GD/WT
GD/WT
Soma
X
Germline
germline homing
germline homing
Cross of F0 drive heterozygotes to each other
F1
~90% GD/GD
~10% GD/ND
<0.1% WT/WT
COPAS enrichment of F1 drive heterozygotes
GD/ND genotypes
Selection of females
GD/WT
GD/R1
GD/R2
Females
Amplicon sequencing
Infertile
Depletion of non-resistant mutations
F2
WT
R1
R2
Amplicon sequencing
GD – Gene Drive
WT – Wild-type Target Site
R1 – Resistant mutation
R2 – Non-resistant mutation
ND – Non-Drive (wild-type or mutant at target)
B
Allele
F1
F2
Relative frequency in F2
203-GAG
1.472
189-GTGAGGAAGAAA
1.616
Incomplete homing
0.019
203-GAGGAG
2.112
206-GA
0.066
201+G 205-AGGA
0.001
201+G
0.001
207-A
0.218
205-GGAGGAG
0.103
203-GAGGA
0.000
192-GAGGAAGAAAGTGAGGAG
0.000
196-AGAAAGTGAGG
0.013
202-TGAGGA
0.800
199-AAGTGAGG
0.009
Others (<1%)
2.85
0.6 0.5 0.4 0.3 0.2 0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6
Frequency amongst mutants
(excluding drive and wild-type)