A male-biased sex-distorter gene drive for the human malaria vector *Anopheles gambiae*

Alekos Simoni[1,2†], Andrew M Hammond[1,3†], Andrea K Beaghton[1], Roberto Galizi[1,4], Chrysanthi Taxiarchi[1], Kyros Kyrou[1], Dario Meacci[1], Matthew Gribble[1], Giulia Morselli[1], Austin Burt[5], Tony Nolan[1,6] and Andrea Crisanti[1,7*]

Author affiliations:
[1] Imperial College London, London SW7 2AZ, UK
[2] Polo d'Innovazione Genomica, Genetica e Biologia, Terni, 05100, Italy
[3] W.Harry Feinstone Department of Molecular Microbiology and Immunology, Johns Hopkins University, Baltimore, MD 21205, USA
[4] Centre for Applied Entomology and Parasitology, School of Life Sciences, Keele University, ST5 5BG, UK
[5] Imperial College London, Silwood Park, Ascot, SL5 7PY, UK
[6] Liverpool School of Tropical Medicine, Liverpool, L3 5QA, UK
[7] Department of Molecular Medicine, University of Padova, Padova, Italy

[†] Contributed equally to this work

*Corresponding author acrs@imperial.ac.uk

Abstract

Only female insects transmit diseases such as malaria, dengue and Zika, therefore control methods that bias the sex ratio of insect offspring have long been sought. Genetic elements such as sex chromosome drives can distort sex ratios to produce unisexual populations that eventually collapse, but the underlying molecular mechanisms are unknown. We report a male-biased sex-distorter gene drive (SDGD) in the human malaria vector *Anopheles gambiae*. We induced super-Mendelian inheritance of the X-chromosome shredding I-PpoI nuclease by coupling a CRISPR-based gene drive into a conserved sequence of the *doublesex* gene (*dsx*). Invasion dynamics of SDGD are predicted to have a quicker impact on female mosquito

population than previously developed gene drives targeting female fertility. The SDGD at the *dsx* locus led to a male-only population from a 2.5% starting allelic frequency in 10-14 generations, with population collapse and no selection for resistance. Our results support the case for the use of SDGD for malaria vector control.

Introduction

Sex-chromosome drivers are genetic elements that interfere with chromosome segregation during meiosis and are over-represented in progeny (reviewed in [1]). In heterogametic sex they cause an unbalanced male to female ratio among offspring that can potentially lead to population suppression or extinction. Relatively few sex chromosome drives have been characterized, most likely because they produce an evolutionary conflict with the rest of the genome that selects for autosomal suppressors or resistant sex chromosomes[2,3].

Mathematical modelling predicts that a driving sex distorter will spread in a population, and, in the absence of resistance, cause eventual collapse[4,5]. Population collapse using natural sex chromosome drives has been reported in laboratory colonies of *Drosophila*[6,7]. In the field, a population crash of the species *Drosophila neotestacea* was detected in Washington State, USA, due to a natural X chromosome distorter which produced a female only population[8]. Therefore sex distorter drives could conceivably be harnessed for invasive pest or vector control[9,10].

Although Y drives are less common than X drives, they are been described in *Aedes aegypti* and *Culex pipiens* mosquitoes [11,12]. Y drives are particularly attractive for mosquito vector control because they would progressively reduce the number of females and hence disease transmission as they spread. Plus, Y drives are likely to be more effective than X drives because they will increase at a greater rate the fraction of heterogametic driving individuals[3–5]. Synthetic sex distorters have been generated in *Anopheles gambiae* mosquitoes by using site-specific nucleases such as I-PpoI or CRISPR/Cas9, which cleave conserved repeated sequences in the mosquito ribosomal DNA gene cluster exclusively located on the X chromosome[13,14]. These nucleases, when expressed during spermatozoa development, selectively cleave the X chromosome, thereby favoring the production of Y-bearing gametes and causing a 95% male-bias in the progeny[13,14]. However, attempts to convert synthetic sex-ratio distorters into Y chromosome drives have been unsuccessful so far. In most insect species, including *A. gambiae,* the sex chromosomes are

transcriptionally shut down during gametogenesis, a process known as meiotic sex chromosome inactivation (MSCI)[15,16] which prevents the transcription of X shredding nucleases if they are inserted into the Y chromosome (Crisanti and Galizi personal observation).

Recently, a gene drive that targeted the *doublesex* gene (*dsx*) reached 100% frequency in 7-11 generations and crashed a caged population of 600 mosquitoes without inducing resistance[17]. We hypothesized that it might be possible to circumvent MSCI by developing an autosomal male-biased sex distorter and coupling sex-ratio distortion with drive. This could result in a quicker impact on disease transmission and a synergistic effect (robustness) between the sex distorter and gene drive components. We report here the design and validation of a sex distorter gene drive to spread the X chromosome shredding I-PpoI endonuclease and produce a male-only insect population.

Designing a sex-distorter gene drive (SDGD)

We designed an SDGD system by combining (on the same construct) a CRISPR-based gene drive that targets a haplosufficient female fertility gene with the I-PpoI endonuclease, which in turn cleaves a conserved sequence in the X-linked ribosomal gene cluster (Fig. 1A and B). We used mathematical modelling to test the likely spread of this SDGD design. Our results indicated that our SDGD could spread rapidly from a low starting frequency to produce a largely unisexual male population, and would also impose a fitness load by impairing female fertility, which together would eliminate the population (Fig. 1C). This SDGD design is different from the previously reported CRISPR-based gene drives which target recessive female fertility genes and impose a fitness load by generation of homozygous sterile mutants[17,18]. Modelling predicted that this SDGD would quickly bias the population towards males, and gradually reduces the abundance of biting females, which both reduces pathogen transmission (by females only) and suppresses the population (Table 1 and Supplementary Fig. 1).

We generated distinct *Anopheles gambiae* SDGD strains targeting three haplosufficient genes (*AGAP011377*, *AGAP007280* and *AGAP005958*) with

established roles in female fertility[18]. We assessed the activity of these SDGD constructs (SDGD^{011377}, SDGD^{007280} and SDGD^{005958}) in the progeny of crosses between SDGD heterozygous (SDGD$^{-/+}$) and wild-type by scoring the fraction of offspring containing the drive element and the sex-ratio of the progeny. SDGD^{007280} had severely compromised fertility and we did not recover enough progeny to assess drive activity. We found average inheritance rates of 79% (± 0.17 s.d.) for SDGD^{011377} and 98% (± 0.08 s.d.) for SDGD^{005958} (Supplementary Fig. 2 and Supplementary Table 1). Furthermore, we observed a male bias ranging from 92% to 94% in the progeny of SDGD^{011377} and SDGD^{005958} heterozygous males. Monitoring life history traits revealed a dramatic reduction of female fertility in SDGD^{011377} and SDGD^{005958} heterozygous females (Supplementary Fig. 3), similarly to previous findings when targeting the same genes with a *vasa*-Cas9 gene drive construct[18]. We attributed the reduction in fertility of SDGD^{011377} and SDGD^{005958} heterozygous female to ectopic expression of the *vasa* promoter and subsequent conversion to a null genotype for the target gene in somatic tissues, where the gene product is required[18–20]. In addition, the *vasa* promoter is known to induce maternal deposition of Cas9 into the developing embryo, resulting in deleterious mutations of the paternally-inherited gene copy, in addition to the null allele inherited from the mother, imposing additional fitness cost to heterozygous female offspring. We also observed a strong reduction in fertility of heterozygous males, particularly in SDGD^{007280} and SDGD^{005958} (Supplementary Fig. 3). We hypothesised that male sterility in SDGD^{007280} and (partial) in SDGD^{005958} was due to locus-dependent high expression of the I-PpoI nuclease, which, if persisting in spermatozoa, shreds the maternally-inherited X chromosome in the fertilized embryo resulting in embryo lethality[13,21]. Despite high levels of drive transmission and male bias, unintended and severe fertility costs prevented the spread of SDGD^{011377} and SDGD^{005958} into caged mosquito populations when seeded at 12.5% allelic frequency (Supplementary Fig. 4). SDGD^{005958} failed to persist in the populations, and disappeared after two generations. SDGD^{011377} was stable for 8 generations, owing to a better balance of drive and fitness costs. This, in turn, generated a low-level population suppression by maintaining a sex ratio of approximately 65% males (Supplementary Fig. 4).

Optimization of temporal, spatial and level of expression of Cas9 and I-PpoI

Our initial findings revealed that SDGD constructs targeting female fertility genes could bias both their own inheritance and the sex ratio of progeny. However, fitness costs, most likely associated with non-optimal spatial and temporal activity of both the Cas9 and I-PpoI genes, impaired SDGD spread into mosquito populations. To minimise the ectopic activity of Cas9 we replaced the *vasa* promoter with the regulatory regions of the *zero population growth* (*zpg*) gene (*AGAP006241*). The *zpg* promoter has previously been applied to regulate Cas9 expression in gene drive constructs and increases fertility of heterozygous individuals compared to *vasa*[17,19]. Previous studies have also shown that the expression levels of I-PpoI during spermatogenesis are crucial in determining whether the outcome is sex bias or sterility; high levels of activity correlate with male sterility[13,21]. The destabilised version of I-PpoI (W124L[13]) used in this study was previously found to confer the highest levels of fertility whilst maintaining strong male bias from at least three independent genomic loci[13], but this I-PpoI variant impaired male fertility when expressed under the transgenic *beta2* promoter inserted into the *AGAP011377*, *AGAP007280* and *AGAP005958* loci. To reduce the transcriptional activity of the *beta2* promoter we generated three variants by inserting a GC-rich sequence of 100 bp in proximity to conserved sequences at position -244, -271 or -355 upstream of the ATG start codon (Supplementary Fig. 5). Each variant was tested for expression using a dual-fluorescence reporter system *in vivo* (Supplementary Fig. 5). For subsequent experiments we selected the *beta2* promoter variant 244 (*beta2*^{244}) showing a transcriptional activity of about 8.1% compared to wild type promoter sequence (Supplementary Fig. 6). The initial SDGD plasmid was then modified to replace the *vasa* promoter with the *zpg* regulatory sequences (as described in [17]) while the *beta2* promoter was replaced with the *beta2*^{244} variant.

An SDGD targeting the *doublesex* gene

To maximise the performance of SDGD we developed the construct SDGD^{dsx} containing the *zpg*-Cas9 transcription unit, the *beta2*^{244}-I-PpoI and a gRNA designed to target the intron 4–exon 5 boundary of the *doublesex* gene (AGAP004050),

because we previously reported that this site minimizes the development of resistance to a gene drive[17]. In addition, females that are homozygous for *dsxF* exhibit an "intersex" phenotype and are viable but unable to bite[17], therefore impacting on the vector competence of the population earlier than an SDGD targeting a standard female fertility locus (in which homozygous females are sterile but able to bite and transmit). Unlike SDGD^{007280}, SDGD^{011377} and SDGD^{005958}, SDGD^{dsx} had no measurable impact on heterozygotes fertility: larval output of SDGD^{dsx} males was comparable to control (126.7 ± 50.7 s.d. and 140.8 ± 40.8, respectively, p=0.39, Fig. 2A and Supplementary Table 1). The fertility of SDGD$^{dsx\ -/+}$ females measured as viable offspring is reduced compared to control individual (98.8 ± 63 S.D. and 140.8 ± 40.8, respectively, p=0.012) though it is still high enough to produce a large number of fertile individuals (Fig. 2A). High levels of maternal nuclease deposition can affect the fertility of females progeny[17,18,20]-, however we did not observe a significant difference in fertility when comparing females inheriting the transgene from a transgenic female or a transgenic male parent (Supplementary Fig. 7). As expected, we observed a marked male bias (93.1% ± 0.08 s.s.) in the offspring of heterozygous SDGD^{dsx} male (Fig. 2B). The sex distortion phenotype was stably transmitted from male mosquitoes to their transgenic male offspring and no differences were observed when comparing males that inherited the construct from a female or a male (Supplementary Table 3). A strong super-Mendelian inheritance of the construct of 96.0% ± 0.08 s.d. and 99.9% ± 0.01 s.d. was observed from both males and females, respectively, based on the frequency of RFP positive progeny from heterozygous parents (Fig. 2B), making the SDGD^{dsx} suitable for population suppression experiments.

SDGD^{dsx} invades caged mosquito populations

We used fertility, inheritance bias, sex distortion data and mutant phenotype information to develop both deterministic and stochastic discrete-generation model (see Online Methods and Supplementary Table 2, 4 and 5) to predict the spread of SDGD^{dsx} into mosquito populations, simulating the release of 10% and 50% heterozygous SDGD^{dsx} mosquitoes into caged populations of 600 individuals. The

stochastic model predicted that the transgene would quickly invade the population, reaching 100% frequency and leading to collapse of the population in 93% and 98% of 10,000 simulations after 30 generations from a 10% and 50% SDGD release, respectively (Fig. 3). The deterministic model, however, showed differences in outcome depending on the values for the fertility of heterozygous females and males, ranging from population elimination to suppression and to the disappearance of the SDGD^{dsx} if male fertility is below 0.5 compared to wild type (Supplementary Fig. 8 and 9). To test the model prediction, we released SDGD^{dsx} heterozygotes at 2.5% or 25% allelic frequency into two populations of 600 caged mosquitoes, each in two replicates. At each generation larvae were screened for the presence of the fluorescence marker linked to the transgene and subsequently the fraction of males and females in the population was assessed. We observed a rapid spread of SDGD^{dsx} in all 4 populations reaching 100% frequency between 4 to 12 generations. The spread of SDGD^{dsx} induced a strong bias of the population sex-ratio towards males, accompanied by a progressive reduction of egg output that led to population elimination at generation 5 and 6 for the replica cages started with 25% of SDGD^{dsx} allelic frequency and at generation 9 and 13 for the 2.5% SDGD^{dsx} release (Fig. 3).

Fitness of female progeny in SDGD^{dsx} males

SDGD$^{+/-}$ males generate <6% female progeny and the female offspring inherit an X chromosome from male gamtes that have been exposed to the I-PpoI nuclease during spermatogenesis. We investigated whether the inheritance of a potentially damaged X chromosome affected female fertility and SDGD HDR rate. We crossed females that carry one 'I-PpoI exposed' X chromosome from the father, to wild-type males and compared their fertility parameters to daughters of SDGD^{dsx} females that carry two unaffected copies of the X chromosome. We observed that females inheriting one 'I-PpoI exposed' X chromosome did not significantly differ in fertility (measured as the number of hatched larvae) nor in drive inheritance, suggesting that if there is a contribution to fitness of a damaged X chromosome in the females, this is not detectable in our assay. To further investigate the potential impact of 'I-PpoI exposed' X chromosomes we modelled additional fitness reductions in

individuals with a damaged X using deterministic discrete-generation cage simulations of a theoretical scenario of SDGD^{dsx} release (10% males and 50% males/50% females) into a caged population (Supplementary Fig. 10). The model predicts little or no effect during the initial spread of the transgene, but a reduction in the suppression load that correlates with the cost of the damaged X chromosome (Supplementary Fig. 10).

Dynamics of sex distorter drive

Driving a sex distorter into a female fertility locus could impose a sufficiently high load on the population to the point that the population is suppressed and eliminated. However, the dynamics of a sex-distorter gene drive are complex and depend not only on the fertility of heterozygous SDGD^{dsx} individuals (Supplementary Fig. 8 and 9) but also on the rate of male bias (Supplementary Fig. 11) and in certain scenarios these dynamics are not intuitive. For example, when female (W/D) fertility is reduced (e.g. below 0.5), the load on the population increases with increasing sex distortion, whereas for higher female fertility (e.g. above 0.5), the load is greater when there is no sex distortion (equivalent to a gene drive without sex distorter, m=0.5, Supplementary Fig. 11). The sex distorter allows the SDGD construct to spread at low (or even zero) female fertility imposing a substantial load (see also Supplementary Fig. 12). This is because the male sex bias mitigates the effect of low female fitness. Overall, increasing sex distortion makes the construct less sensitive to variation in female heterozygous fertility (Supplementary Fig. 12). At the limit of complete male bias (male progeny = 100%), the load is independent of female fertility, since no SDGD females are created and only SDGD heterozygous males can pass on the construct. Based on our experimental parameter estimates for the SDGD^{dsx}, the SDGD allele is predicted to be present in an intermediate equilibrium with wildtype and non-functional resistance alleles at sufficient frequency to still induce a dramatic population reduction, and possibly preventing reinvasion events (Supplementary Fig. 13).

Discussion

Our results show that SDGD^{dsx} functions as a sex-distorter autosomal gene drive. In four cage experiments SDGD^{dsx} progressively biased the sex-ratio towards males, with eventual population collapse. Notably we did not observe the development of functional mutations at the target *dsx* site that blocked the spread of the distorter. This observation further supports the notion that the *doublesex* sequence at the boundary of intron4-exon 5 is highly functionally constrained and validates its use as a target sequence for gene drive solutions in anopheline mosquitoes. It should also be noted that a sex-distorter that simultaneously destroys the female isoform of the *dsx* gene while reducing the female population also decreases the opportunity of resistant mutations to arise (because they are not selected in males). In addition, targeting a sequence present in hundreds of copies on the X chromosome reduces the likelihood that nuclease-induced resistance will evolve to block the sex-distorter component.

Our SDGD solution also combines a number of features in terms of efficacy, robustness and predicted time to impact (on disease transmission), which differ from previously described gene drives or autosomal sex-distorter systems, making it particularly attractive for field implementation (Table 1). In two replicate caged experiments SDGD^{dsx} consistently induced population collapse starting from an allelic frequency of 2.5%. For field experiments this translates into mosquito numbers to be released that are within the range of production capability: Recent studies modelling the impact of hypothetical X-shredder Y-drive mosquitoes on a national scale predicts that the release of as little as 10 males in 1% of human settlements achieve more than 90% population suppression after 4 years[22].

SDGD^{dsx} is predicted to show a higher level of robustness than a gene drive alone, even if one of the critical components breaks down or mutates, due to the synergy the components. The loss or inactivation of the I-PpoI sequence will result in the generation of functional *dsx* gene drive that will also contribute to population suppression (Supplementary Fig. 12) and the loss of function of either one of the two drive components (Cas9 and gRNA) will produce non-functional *dsx* alleles (R) that in heterozygous individuals will still contribute to the production of male biased progeny due to the presence of a functional I-PpoI. Mutations and recombination events of the constructs involving both the drive and the distorter will generate R

non-functional *dsx* mutations. These R mutations are constantly generated at the target locus by the action of the nuclease[17], but they are not selected because they do not restore function of the *dsx* gene and homozygous R females are sterile therefore they are continuously lost as they arise.

Modelling based on our experimental data shows that SDGD^{dsx} offers some important advantages in short-term drive dynamics and long-term outcomes. Importantly, the number of transmission competent (i.e. biting) females is reduced faster by SDGD than by a standard gene drive targeting the same locus (Supplementary Fig. 1) and could lead to a potential strong effect on disease transmission (time to impact after release). When comparing a distorting and a non-distorting gene drive, the equilibrium load imposed by SDGD is much less sensitive to female fitness costs, which is particularly relevant given the uncertain extrapolation of fitness effect measurements from the lab to the field.

Previous modelling of gene drive without the sex distorter showed that under certain conditions (e.g. leaky expression of the drive construct) there can be an accumulation of non-functional cleavage-resistant sequences which prevents the transgene from going to fixation[17,23]. Deterministic modelling of SDGD^{dsx} indicates that there is also the potential for the transgene to go to an intermediate equilibrium frequency and population suppression rather than complete fixation and elimination (Supplementary Figs. 8, 9, 12). The lower the SDGD fertility in heterozygous individuals, the more likely an intermediate equilibrium is reached. For the observed fertility values of SDGD^{dsx} heterozygous females, stochastic models predict population elimination for finite cage populations in 93-98% of the simulations with kinetics of spread in line with observed data. Under field or semi-field conditions the fertility estimates of heterozygous individuals could differ and tilt the balance in one way or the other towards population reduction rather than population elimination. Achieving a strong population reduction may be regarded as less effective than elimination in a field scenario, however it could help achieve long-term stable vector control via a higher tolerance to repopulation through migration compared to a system which quickly eliminates an entire target population.

Males carrying a non-driving I-PpoI construct designed to cause dominant male sterility[21] were recently released in a field location of Burkina Faso[24] as part of

a phased, step-by-step assessment of novel genetic approaches to malaria control, following independent guidance and recommendations[25,26]. This opened the way to the utilization of I-PpoI-based distorter for the implementation of genetic vector control measures.

We believe that SDGD^{dsx} outperforms other anopheline gene drives combining efficacy, resistance management and robustness and is well suited as an anti-malaria intervention.

Acknowledgments

This work was funded by grants from the Bill & Melinda Gates Foundation and additionally from the European Commission (731060 - Infravec2) to A.C.

Author contribution

A.S., A.M.H., T.N. and A.C. designed the research; A.S., A.M.H., R.G., K.K. and C.T. performed the research. A.K.B. and A.B. performed modelling; D.M., M.G. and G.M. performed insect husbandry; A.S., A.M.H. and A.C. analyzed data; A.S. and A.C. wrote the paper, with input from all authors.

Data availability statement

Full sequence of vectors is provided through the NCBI database. GeneBank accession code for vector p172: MT270142 and for p182-244: MT270141. Sanger sequencing of vector p182-244 is available as supplementary information. Additional vectors sequences are provided as Supplementary data.

Competing financial Interests Statement

The authors declare no competing financial interests.

Ethics statement.

All animal work was conducted according to UK Home Office Regulations and approved under Home Office License PPL 70/8914.

References

1. Helleu, Q., Gérard, P. R. & Montchamp-Moreau, C. Sex Chromosome Drive. *Cold Spring Harb. Perspect. Biol.* **7**, (2015).
2. Taylor, J. E. & Jaenike, J. Sperm competition and the dynamics of X chromosome drive: stability and extinction. *Genetics* **160**, 1721–1731 (2002).
3. BURT, A., Trivers, R. & Burt, A. *Genes in Conflict: The Biology of Selfish Genetic Elements*. (Harvard University Press, 2009).
4. Hamilton, W. D. Extraordinary Sex Ratios. *Science* **156**, 477–488 (1967).
5. Deredec, A., Godfray, H. C. J. & Burt, A. Requirements for effective malaria control with homing endonuclease genes. *Proc. Natl. Acad. Sci.* **108**, E874–E880 (2011).
6. Lyttle, T. W. EXPERIMENTAL POPULATION GENETICS OF MEIOTIC DRIVE SYSTEMS1s2I. PSEUDO-Y CHROMOSOMAL DRIVE AS A MEANS OF ELIMINATING CAGE POPULATIONS OF. *Genetics* **76**, 413–445 (1977).
7. Price, T. A. R., Hodgson, D. J., Lewis, Z., Hurst, G. D. D. & Wedell, N. Selfish Genetic Elements Promote Polyandry in a Fly. *Science* **322**, 1241–1243 (2008).
8. Pinzone, C. A. & Dyer, K. A. Association of polyandry and sex-ratio drive prevalence in natural populations of Drosophila neotestacea. *Proc. Biol. Sci.* **280**, 20131397 (2013).
9. Hastings, I. M. Selfish DNA as a method of pest control. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **344**, 313–324 (1994).
10. Sinkins, S. P. & Gould, F. Gene drive systems for insect disease vectors. *Nat. Rev. Genet.* **7**, 427–435 (2006).
11. Hickey, W. A. & Craig, G. B. GENETIC DISTORTION OF SEX RATIO IN A MOSQUITO, AEDES AEGYPTP. *Genetics* **53**, 1177–1196 (1966).
12. Sweeny, T. L. & Barr, R. SEX RATIO DISTORTION CAUSED BY MEIOTIC DRIVE. *Genetics* **88**, 427–446 (1978).
13. Galizi, R. *et al.* A synthetic sex ratio distortion system for the control of the human malaria mosquito. *Nat Commun* **5**, 3977 (2014).
14. Galizi, R. *et al.* A CRISPR-Cas9 sex-ratio distortion system for genetic control. *Sci Rep* **6**, 31139 (2016).
15. Turner, J. M. A. Meiotic sex chromosome inactivation. *Dev. Camb. Engl.* **134**, 1823–1831 (2007).
16. Taxiarchi, C. *et al.* High-resolution transcriptional profiling of Anopheles gambiae spermatogenesis reveals mechanisms of sex chromosome regulation. *Sci. Rep.* **9**, 1–12 (2019).
17. Kyrou, K. *et al.* A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. *Nat. Biotechnol.* **36**, 1062–1066 (2018).
18. 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 (2016).
19. Hammond, A. M. *et al.* Improved CRISPR-based suppression gene drives mitigate resistance and impose a large reproductive load on laboratory-contained mosquito populations. *bioRxiv* (2018) doi:10.1101/360339.
20. 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.* **112**, E6736–E6743 (2015).

21. Windbichler, N., Papathanos, P. A. & Crisanti, A. Targeting the X chromosome during spermatogenesis induces Y chromosome transmission ratio distortion and early dominant embryo lethality in Anopheles gambiae. *PLoS Genet* **4**, e1000291 (2008).
22. North, A. R., Burt, A. & Godfray, H. C. J. Modelling the potential of genetic control of malaria mosquitoes at national scale. *BMC Biol.* **17**, 26 (2019).
23. Beaghton, A. K., Hammond, A., Nolan, T., Crisanti, A. & Burt, A. Gene drive for population genetic control: non-functional resistance and parental effects. *Proc. R. Soc. B Biol. Sci.* **286**, 20191586 (2019).
24. Scudellari, M. Self-destructing mosquitoes and sterilized rodents: the promise of gene drives. *Nature* **571**, 160 (2019).
25. James, S. *et al.* Pathway to Deployment of Gene Drive Mosquitoes as a Potential Biocontrol Tool for Elimination of Malaria in Sub-Saharan Africa: Recommendations of a Scientific Working Group †. *Am. J. Trop. Med. Hyg.* **98**, 1–49 (2018).
26. National Academies of Sciences, E. *Gene Drives on the Horizon: Advancing Science, Navigating Uncertainty, and Aligning Research with Public Values*. (2016). doi:10.17226/23405.

FIGURES LEGENDS

F**igure 1**. **Driving a sex distorter system in the autosome**.
a) Schematic overview of the construct utilised to build a sex distorter gene drive (SDGD), which contains four transcription units: the I-Ppo*I* nuclease (variant W124L) expressed a fusion protein with the visual marker eGFP, under the male-specific *beta2-tubulin* germline promoter; the hSpCas9 nuclease regulated by a promoter that is active in the germline of both males and (from the vasa or zpg gene); a gRNA under the control of the ubiquitous U6 PolIII promoter, designed for homing at previously characterized haplosufficient fertility genes; and a 3xP3::DsRed gene as a fluorescent integration marker. b) Mode of action of the autosomal SDGD. The sex distorter component (I-Ppo*I* blue square) and the gene drive component (CRISPR^{h} red square) are linked head-to-tail into the same construct that is integrated in the autosome within a fertility gene. In the germline of a female transgenic mosquito (highlighted in red) the CRISPR^{h} component is active (red arrowhead) leading to super-Mendelian inheritance of the transgene by homology directed repair (HDR). In the germline of a male transgenic, both gene drive (red arrowhead) and the sex distorter (blue arrowhead) transcription units are active leading to homing of the construct (by action of the CRISPR^{h}) and shredding of the X chromosome (by action of I-PpoI targeting ribosomal rDNA repeats, indicated by vertical lines). This results in a bias of the sex ratio towards males in the progeny and super-Mendelian inheritance of the transgene. c) Idealised predictions (discrete-generation deterministic model) of transgenic frequency for spread in a population (solid line) alongside with the load on the target population (dotted line) for a sex distorter gene drive construct (left-hand panel, with fraction of male progeny m = 0.95) and a gene drive (right-hand panel, m=0.5) targeting a female fertility gene. Blue shading represents the fraction of males in the population. Pink shading indicates the fraction of female in the population, with the fraction of fertile females indicated in a darker colour. This idealised model makes several assumptions that are likely to vary by strain, including but not limited to: full fitness in males and heterozygous females (fully recessive female fertility gene); complete sterility in homozygous females; 95% SDGD transmission in male and female heterozygotes; no generation of drive-resistant mutations; no loss of function of sex distorter; single release of male drive heterozygotes is 1% of the male population.

Figure 2. Fertility (a), sex and inheritance bias (b) of a sex distorter gene drive targeting the female isoform of the sex determination gene AGAP004050 (*doublesex*).
a) Counts of eggs and hatched larvae determined in individual crosses ($n \geq 33$) of heterozygous SDGD^{dsx} females and males to wild-type mosquitoes. While male fertility is comparable with wild-type (male fertility 0.86, ns), females showed 37% reduction in larval output compared to wild-type (female fertility 0.627. * P=0.0124 Kruskal-Wallis test). Values on the right indicate mean count ± s.e.m. and larval hatching rate in parenthesis. **b)** Scattered plots showing the fraction of SDGD^{dsx} transgene inheritance (y-axis) against sex bias (x-axis) in the progeny of individual SDGD^{dsx} trans-heterozygous males (left-hand panel, n = 63) and females (right-hand panel, n = 39) crossed to wild-type individuals. Individual blue and pink dots represent the progeny derived from a single female and the red dot indicates the average of the population. Error bars indicate standard deviation. Both males and females SDGD^{dsx}/+ showed a super-Mendelian inheritance of the transgene determined

by scoring the presence of RFP marker in the progeny. Males SDGD^{dsx}/+ showed a strong bias in sex ratio towards males (0.93 ± 0.09). Dotted lines indicate fraction of males (x-axis) and the fraction of SDGD (y-axis) as expected by Mendelian inheritance.

Figure 3. **Kinetics of SDGD^{dsx} spread in target mosquito populations.**
The spread of the sex distorter gene drive was investigated in two different experiments starting with an allelic frequency of 2.5% (10% male release) and 25% (50% male and female release) respectively, in replica (cage A and cage B). The 10% release cages were set up with a starting population of 300 wild-type females, 270 wt males and 30 SDGD^{dsx} -/+ males. The 50% release were started with 150 wild-type females, 150 wt males, 150 SDGD^{dsx} /+ males and 150 SDGD^{dsx} /+ females (SDGD allele frequency of 25%). Each consecutive generation was established by selecting 600 larvae. The frequency of the transgene (as fraction of RFP+ individual), the sex ratio (female/male) and the relative egg output (fraction of egg produced relative to first generation) was recorded at each generation. **a)** The bar plots represent the fraction of males and females (blue and light red shading, respectively) for each population and the fraction of transgenic individuals is shown in striped pattern. Black lines indicate the total fraction of individual containing the SDGD (as fraction of RFP+). **b)** The frequency of the transgene, the sex ratio and the relative egg output are superimposed on both a deterministic model (dotted black lines) and 20 representative stochastic simulations (grey solid lines) of dynamics of invasion of SDGD based on release scenarios of 25% and 2.5 % SDGD^{dsx} allelic frequency. In 93% and 98% of the stochastic simulations (out of 10,000 runs) the release of SDGD/+ individuals at a starting frequency of 2.5% and 25%, respectively, is predicted to collapse the population within 30 generations. Dotted lines indicate the expected mendelian distribution of gender. Fitness and life history parameter estimates used in the model are provided in Supplementary Table 2.

TABLES AU NUMBER REFERENCES IN THE TABLE PLEASE

Construct type	**Construct name**	**Homing rate**	**Male-biased sex-ratio distortion**	**Spread in caged population**	**Population suppression**	**Development of resistance in cages**	**Impact of heterozygotes on population size†**	**Component redundancy**	**Reference**
Gene drive	dsxF^{CRISPRh}	92% males 99% females	50%	Yes	Yes	No	No	No	17
Sex distorter gene drive	SDGD^{dsx}	92% males 99% females	93%	Yes	Yes	No	Yes	Yes	This study
Autosomal sex-distorter	gfp111A-2	0%	95%	No	Yes (overflooding ratio of 3X)	No	Yes	No	13
Y-drive	n.a	100% males 0% females*	95%*	Yes*	n.a.	n.a.	Yes*	No	n.a.

Table 1. Comparison of performance of gene drive and sex distorter genetic control approaches in terms of efficacy, spread and robustness. Homing rate is defined as the fraction of transgenic progeny above Mendelian inheritance. * values based on hypothetical X-shredder construct inserted on the Y chromosome generating 95% male offspring, all of which inherit the transgene. † Indicates the ability of the construct to have an impact on the population size (i.e. number of females) in heterozygosity (or hemizygosity for Y-drive) compared to constructs targeting recessive female fertility loci which impact population size when homozygote transgenic females are generated.

ONLINE METHODS

Generation of Sex Distorter Gene Drive constructs

To create the SDGD vectors p172 (*vas2*) (GenBank accession MT270142) and p182 (*zpg*), the β2-eGFP(F2A)I-PpoI transcription unit from pBac[3xP3-DsRed]β2-eGFP::I-PpoI-124L[13] was excised by AscI digestion and cloned into AscI-digested p165 (*vas2*-CRISPR^{h}) [18] (GenBank accession code KU189142) and p174 (*zpg*-CRISPR^{h}) [19] (GenBank accession code MH541847) respectively. SDGD vectors were further modified by BsaI-mediated Golden Gate assembly to contain gRNA spacers targeting AGAP011377 (GCAGACGTAGAAATTTTC), AGAP007280 (GGAAGAAAGTGAGGAGGA), AGAP005958 (GAGATACTGGAGCCGCGAGC) [18] and AGAP004050 (GTTTAACACAGGTCAAGCGG) [17]. To include the beta2^{244} promoter modification, the plasmid p182 was further modified to generate p182-244 (GenBank accession code MT270141) according the beta2^{244} variant described below. Additional sequences of all vectors are available as supplementary information.

Microinjection of embryos and selection of transformed mosquitoes

All mosquitoes were reared under standard conditions of 80% relative humidity and 28°C. The mosquitoes were blood-fed on anesthetized mice or by Hemotek, and freshly laid embryos were aligned and used for microinjections as described before[27]. To generate SDGD mosquitoes, we injected respective docking lines [17,18] embryos with solution containing p174 or p182-244 and a plasmid-based source of PhiC31 integrase[28] (at 200 ng/μl and 400 ng/μl, respectively). All the surviving G_0 larvae were crossed to wild-type mosquitoes and G_1 positive transformants were identified using a fluorescence microscope (Nikon, Eclipse TE200) as RFP^{+} larvae for the RMCE events.

Containment of gene drive mosquitoes

All mosquitoes were housed at Imperial College London in an insectary that is compliant with Arthropod Containment Guidelines Level 2 (ref.[29]). All GM work was performed under institutionally approved biosafety and GM protocols. In particular, GM mosquitoes containing constructs with the potential to show gene drive were housed in dedicated cubicles, separated by at least six doors from the external environment and requiring two levels of security card access. Moreover, because of its location in a city with a northern temperate climate, *A. gambiae* mosquitoes housed in the insectary are also ecologically contained. The physical and ecological containment of the insectary are compliant with guidelines set out in a recent commentary calling for safeguards in the study of synthetic gene drive technologies[30].

Mutagenesis of beta2 promoter

Bioinformatic analysis the regulatory region of the beta2-tubulin gene (AGAP008622) was performed using the promoter2.0 prediction server [31] and the Neural Network Promoter Prediction tool [32] to identify conserved region. A synthetic 100 bp DNA sequence with a GC content of 65% (sequence reported in Supplementary Fig. 3) was designed using Geneious R11 (https://www.geneious.com), and cloned into the beta2promoter at position -244, -271 and -355 relative to the ATG start codon using site-specific mutagenesis of the plasmid pBac[3xP3-DsRed]β2-eGFP::I-PpoI-124L [13] by nested PCR using primers pairs B2-355_r and B2-355_f, B2-271_r and B2-3271_f, B2-244_r and B2-244_f followed by Spac-fwd and Spac-rev, for Beat2^{355}, Beta2^{271} and Beta2^{244} variants, respectively. A second unmodified copy of the beta2 promoter was cloned to express the mCherry gene.

Name	Sequence
B2-355_r	GGCCAACTCGGGTCCGAGTCGTCTTCTTGGATGGGATGATG

B2-355_f	CGCCAGCACTCTCAGACTCAATACGAATTTATTTGTGGCATCG
B2-271_r	GGCCAACTCGGGTCCGAGTCATATGACTACTATGATCATCTTTTGC
B2-271_f	CGCCAGCACTCTCAGACTCAGAG CCG TAC GTG CCG G
B2-244_r	GGCCAACTCGGGTCCGAGTCCACGAAATGATCCGGCAC
B2-244_f	CGCCAGCACTCTCAGACTCACAGAACCTTCAGAGACGTTG
Spac-fwd	GTGAGAAGTGCGCGTCTCGTTCCCGCAGCTCGCCAGCACTCTCAGACTCA
Spac-rev	CATCCGCCCTAACTCCGCCCGTGGGTCGTTGGCCAACTCGGGTCCGAGTC

Dual-fluorescence assay experiment

3 to 5 days old adult male heterozygous mosquitoes were collected in Falcon tubes and anesthetized on ice 5 min before dissection. Testes were micro-dissected using an Olympus SZX7 optical microscopes and pictures of gonads were taken using the EVOS imaging system (Thermo-Fisher) with magnification 20X and the following exposure settings: Bright field: gain 50%; GFP channel: gain 30% 120 ms; RFP channel: gain 80%, 120 ms. Unmodified pictures were then analysed using ImageJ software [33] as follow. Testes areas were selected using free-form selection tool and integrated density and mean grey values were measured for GFP and RFP channel independently using the same selection area. A reading for the background (with same selection area) was then subtracted to the integrated density value for each testis to remove background noise. Value for fluorescent intensity was measured as ratio between GFP reading and mCherry readings and normalized to the value of unmodified beta2 control.

Phenotypic assays

Phenotypic assays designed to examine SDGD inheritance and relative fecundity in mosquitoes carrying out essentially as described before[17,18]. Briefly, the offspring of heterozygous individuals to wild-type counterpart were screened for RFP expression. Nonfluorescent progeny were kept as controls. Groups of 50 male and 50 female mosquitoes were mated to an equal number of wild-type mosquitoes for 5 d, blood-fed, and a minimum of 40 females allowed to lay individually. The entire egg and larval progeny were counted for each lay. Females that failed to give progeny and had no evidence of sperm in their spermathecae were excluded from the analysis. To determine inheritance and sex-ratio bias of SDGD, the entire larval progeny was screened for presence of DsRed, which is linked to the SDGD allele, and all the progeny was sexed at the pupal stage to determine the sex ratio. Statistical differences between genotypes were assessed using the Kruskal–Wallis test.

Cage trial assays

To performed cage trials of SDGD^{011377}, SDGD^{005958} we introduced 100 heterozygous transgenic males into a population of 100 wild-type males and 200 wild-type females (transgenic allele frequency of 12.5%) in triplicate. As control, 100 heterozygous transgenic males from the autosomal self-limiting sex-distorter ^{gfp}124L-2 line [13] was released at the same frequency in a separate population, in triplicate. In addition, a population of 200 wild-type males and 200 wild-type females served as negative control.

For the starting generation only, age-matched male and female pupae were allowed to emerge in separate cages and were mixed only when all the pupae had emerged. Mosquitoes were left to mate for 5 days before they were blood fed on anesthetized mice. Two days after, the mosquitoes were set to lay in a 300-ml egg bowl filled with water and lined with filter paper. The eggs produced from the cage were photographed and counted using JMicroVision V1.27. Prior to counting, eggs were dispersed using gentle water spraying in the egg bowl to homogenize the population, and 450 eggs were randomly selected to seed the next generation. Larvae emerging from the 450 eggs were counted and screened for the presence of the RFP marker to score the transgenic rate of the progeny. All the pupae were sexed to determine the sex-ratio of the population.

To perform cage trials of SDGD^{dsx}, we set up two different experiments, in replicate. The 10% release cages were set up with a starting population of 300 wild-type females, 270 wt males and 30

heterozygous SDGD^{dsx} -/+ males (starting allelic frequency 2.5%). The 50% release were started with 150 wild-type females, 150 wt males, 150 heterozygous SDGD^{dsx} -/+ males and 150 heterozygous SDGD^{dsx} -/+ females (SDGD allele frequency of 25%). For the starting generation only, age-matched male and female pupae were allowed to emerge in separate cages and were mixed only when all the pupae had emerged. Mosquitoes were left to mate for 5 days before they were blood fed on anesthetized mice. Two days after, the mosquitoes were set to lay in a 300-ml egg bowl filled with water and lined with filter paper. All larvae were allowed to hatch, and each consecutive generation was established by randomly selecting 600 larvae, split in 3 trays of 200 larvae each. All 600 larvae were screened for the presence of the RFP marker and the pupae from 1 tray were sexed to determine the sex-ratio. On day 8 mosquitoes were offered a second blood-meal and all the eggs produced were photographed and counted using the Egg counter software[34].

Statistical analysis

Statistical analysis was performed as indicated using GraphPad Prism version 7.0 (La Jolla California USA).

Population genetics model

Population genetics model (discrete time). To model the results of the cage experiments, we use discrete-generation recursion equations for the genotype frequencies, with males and females treated separately similarly to [17]. We extend [17] to model the sex distorter gene drive by including a sex bias and possible X-chromosome damage in progeny of SDGD males, although here we do not include parental effects on fitness (as these effects were not strongly observed). We consider three alleles at the female fertility target site, W (wildtype), D (driving sex-distorter), and R (non-functional nuclease-resistant). We also differentiate between two possible types of X-chromosome: x (wildtype) and X, which denotes an X-chromosome that has passed through a SDGD male and survived X-shredding but may be damaged, resulting in an additional fitness cost to the individual carrying it. $F_{ij,pq}(t)$ and $M_{ij,qY}(t)$ denote the genotype frequency of females (or males) in the total population, where the first set of indices denotes alleles at the target site $\{\mathrm{WW}, \mathrm{WD}, \mathrm{WR}, \mathrm{DD}, \mathrm{DR}, \mathrm{RR}\}$, and the second set denotes the sex chromosomes, $pq = \{\mathrm{xx}, \mathrm{xX}, \mathrm{XX}\}$ for females and $q = \{\mathrm{x}, \mathrm{X}\}$ for males. There are 18 female genotypes and 12 male genotypes; six types of eggs in proportions $E_{\mathrm{W,x}}, E_{\mathrm{D,x}}, E_{\mathrm{R,x}}, E_{\mathrm{W,X}}, E_{\mathrm{D,X}}, E_{\mathrm{R,X}}$, where the first index refers to the target site allele and the second to the sex chromosome; and eight types of sperm, $S_{\mathrm{W,x}}$, $S_{\mathrm{R,x}}$ (no $S_{\mathrm{D,x}}$, since we assume that SDGD males only contribute X chromosomes), $S_{\mathrm{W,X}}, S_{\mathrm{D,X}}, S_{\mathrm{R,X}}, S_{\mathrm{W,Y}}, S_{\mathrm{D,Y}}, S_{\mathrm{R,Y}}$.

Homing. Adults of genotype $\mathrm{W/D}$ at the target site produce gametes at meiosis in the ratio $\mathrm{W}:\mathrm{D}:\mathrm{R}$ as follows:

$(1-d_f)(1-u_f):d_f:(1-d_f)u_f$ in females

$(1-d_m)(1-u_m):d_m:(1-d_m)u_m$ in males

Here, d_f and d_m are the rates of transmission of the driver allele in the two sexes and u_f and u_m are the fractions of non-drive gametes at the target site that are repaired by meiotic end-joining and are non-functional and resistant to the drive (R). In all other genotypes, inheritance at the target site is Mendelian.

Sex distortion. The SDGD X-shredder only affects the sex ratio of the progeny if it is in males. It destroys the X chromosome while males are making their sperm, resulting in mostly Y-bearing sperm. From male SDGD heterozygotes, progeny will therefore consist of m_1 ($1/2 < m_1 \leq 1$) males and $(1-m_1)$ females; from male SDGD homozygotes (D/D), progeny will be m_2 $(1/2 < m_2 \leq 1)$ males and $(1-m_2)$ females. For simplicity when comparing to experiment, we assume $m_1 = m_2 = m$. We assume no mutations that cause loss of function of the sex distorter from the construct or resistance to X-shredding.

All X chromosomes contributed by SDGD males that have survived X-shredding are assumed to be 'damaged' X (vs wildtype x), which is reflected in reduced reproductive fitness of the individual carrying it (see next section). We assume a damaged X chromosome is susceptible to further shredding

if it is inherited by a SDGD male, and for simplicity, that the fitness cost of carrying a damaged X is the same no matter how many times the chromosome has passed through a SDGD male and survived X-shredding.

Fitness. Let $w_{ij,pq}, w_{ij,qY} \leq 1$ represent the reproductive fitnesses of female and male genotypes relative to fitness one for the wild-type homozygotes, where {*ij*} denotes alleles at the target site of the construct $\{\mathrm{WW}, \mathrm{WD}, \mathrm{WR}, \mathrm{DD}, \mathrm{DR}, \mathrm{RR}\}$, and the second set of indices $pq = \{\mathrm{xx}, \mathrm{xX}, \mathrm{XX}\}$ for females and $q = \{\mathrm{x}, \mathrm{X}\}$ for males.While all fitness parameters are retained in the recursion equations for generality, for comparison with experiment, we assume that the target gene is needed for female fertility, thus females with D/D, D/R and R/R at the target site are sterile. There is no reduction in fitness in W/R females from carrying only one copy of the target gene (W/R), but W/D females have reduced fitness due to the presence of the SDGD construct, as observed experimentally (Supplementary Table 2). We assume no costs to males that have no copies of the driving sex-distorter (W/R and R/R), but that males with one or two copies of the SDGD (W/D, D/D, D/R) have a fitness reduction consistent with experimental observation (Supplementary Table 2).
If the individual also carries a damaged X chromosome, we assume that this imposes an additional cost that affects reproductive success. To calculate the overall fitness of the genotype, the fitness value associated with carrying the damaged X chromosome is multiplied by the fitness value associated with D (or R) alleles at the target site (Supplementary Tables 4 and 5). Reduced fitness in males with a copy of the damaged X is $(1 - s_{\mathrm{X,m}})$, and in females with two copies of the damaged X is $(1 - s_{\mathrm{X,f}})$, with $s_{\mathrm{X,f}}, s_{\mathrm{X,m}} = 0$ if no cost and 1 if the damaged X causes sterility. For females with one damaged X and one wildtype x chromosome, the reduction is $(1 - h_{\mathrm{X,f}}\, s_{\mathrm{X,f}})$, where $h_{\mathrm{X,f}}$ is the dominance coefficient (0 for fully recessive, 1 for fully dominant). For baseline parameters, we assume these costs are zero.

Recursion equations. We firstly consider the gamete contributions from each genotype. The proportions $E_{k,l}(t)$ with allele $k = \{\mathrm{W}, \mathrm{D}, \mathrm{R}\}$ at the target site and sex chromosome $l = \{\mathrm{x}, \mathrm{X}\}$ in eggs produced by females participating in reproduction are given in terms of the female genotype frequencies $F_{ij,pq}(t)$:

$$E_{k,l}(t) = \frac{\sum_{i=1}^{3}\sum_{j=i}^{3}\sum_{pq=xx;xX;XX}\left(c_{ij,pq}^{k,l} w_{ij,pq} F_{ij,pq}(t)\right)}{\sum_{i=1}^{3}\sum_{j=i}^{3}\sum_{pq=xx;xX;XX}\left(w_{ij,pq} F_{ij,pq}(t)\right)} \tag{1a}$$

where i and j are each summed such that {1,2,3} corresponds to {W, D, R}. The coefficients $c_{ij,pq}^{k,l}$ in (1a) correspond to the proportion of the gametes from female individuals of type (ij, pq) that carry alleles (k, l) and are shown in Supp. Table 4 with rows corresponding to genotypes (ij, pq) and columns to alleles (k, l).
The proportions $S_{k,l}(t)$ with allele $k = \{\mathrm{W}, \mathrm{D}, \mathrm{R}\}$ at the target site and sex chromosome $l = \{\mathrm{x}, \mathrm{X}, \mathrm{Y}\}$ in sperm are given in terms of the male genotype frequencies $M_{ij,qY}(t)$:

$$S_{k,l}(t) = \frac{\sum_{i=1}^{3}\sum_{j=i}^{3}\sum_{q=x,X}\left(c_{ij,qY}^{k,l}\, w_{ij,qY} M_{ij,qY}(t)\right)}{\sum_{i=1}^{3}\sum_{j=i}^{3}\sum_{q=x,X}\left(w_{ij,qY} M_{ij,qY}(t)\right)} \tag{1b}$$

where again, i and j are each summed such that {1,2,3} corresponds to {W, D, R}. The coefficients $c_{ij,qY}^{k,l}$in (1b) correspond to the proportion of the gametes from male individuals of type (ij, qY) that carry alleles (k, l) and are shown in Supp. Table 5 with rows corresponding to genotypes (ij, qY) and columns to alleles (k, l). Note that $S_{D,x}(t) = 0$ since SDGD males only contribute damaged X chromosomes, so there is no entry for this included in Supp. Table 5.

We define the proportion of females in the population as:

$$F(t) = \sum_{i=1}^{3}\sum_{j=i}^{3}\sum_{pq=xx;xX;XX} F_{ij,pq}(t)$$

and the average female reproductive fitness as
$\overline{w}_f(t) = \sum_{i=1}^{3}\sum_{j=i}^{3}\sum_{pq=xx;xX;XX}\left(\frac{w_{ij,pq}F_{ij,pq}(t)}{F(t)}\right)$, and analogously for the male proportion,
$M(t) = \sum_{i=1}^{3}\sum_{j=i}^{3}\sum_{q=x,X} M_{ij,qY}(t)$ and average male fitness:

$$\overline{w}_m(t) = \sum_{i=1}^{3}\sum_{j=i}^{3}\sum_{q=x,X}\left(\frac{w_{ij,qY}M_{ij,qY}(t)}{M(t)}\right).$$

Note that in (1a) and (1b), the normalization factor in the denominator is therefore $\overline{w}_f(t)F(t)$ and $\overline{w}_m(t)M(t)$.
The load on the population incorporates reductions in female and male fertility and decreased frequency of females due to the SDGD spreading in the population, and at time t is defined as:

$$L(t) = 1 - 2\,F(t)\,\overline{w}_f(t)\,\overline{w}_m(t)$$

It is zero when only wildtypes are present, and one if the SDGD has established and the average female fitness, or fraction of females present, is zero. We note that increases in load predicted by the cage model do not predict absolute changes in population density in the field but can be an indication of comparative potential reductions[35].

To model cage experiments, we start with an equal number of males and females. For 50% release, the initial frequency for wildtype female and males is $F_{\mathrm{WW,xx}} = M_{\mathrm{WW,xY}} = 1/4$ and for heterozygote drive females and males is $F_{\mathrm{WD,xx}} = M_{\mathrm{WD,xY}} = 1/4$. For 10% release of males only, $M_{\mathrm{WW,xY}} = 9/20$ and $M_{\mathrm{WD,xY}} = 1/20$ and all females are wildtype, $F_{\mathrm{WW,xx}} = 1/2$. Assuming random mating, we obtain the following recursion equations for the female genotype frequencies in generation $(t+1)$:

$$F_{ij,pq}(t+1) = \left(1-\frac{\delta_{ij}}{2}\right)\left(1-\frac{\delta_{pq}}{2}\right)\Big(E_{i,p}(t)S_{j,q}(t) + E_{j,p}(t)S_{i,q}(t) + E_{i,q}(t)S_{j,p}(t) + E_{j,q}(t)S_{i,p}(t)\Big)$$

Where $pq = \{xx, xX, XX\}$, and δ_{ij} is the Kronecker delta. The factors $\left(1-\frac{\delta_{ij}}{2}\right)$, $\left(1-\frac{\delta_{pq}}{2}\right)$ account for the factor of 1/2 for homozygosity at the target site (for ij = {W/W, D/D, R/R}) and at the sex chromosomes (for pq ={xx, XX}). For the male genotype frequencies:

$$M_{ij,qY}(t+1) = \left(1-\frac{\delta_{ij}}{2}\right)\Big(E_{i,q}(t)S_{j,Y}(t) + E_{j,q}(t)S_{i,Y}(t)\Big)$$

where $q = \{x, X\}$, and $\left(1-\frac{\delta_{ij}}{2}\right)$ accounts for the factor of 1/2 for homozygosity at the target site (for ij = {W/W, D/D, R/R})

Stochastic version. In the stochastic version of the model described above, random values for probabilistic events are taken from the appropriate multinomial distributions, with probabilities estimated from experiment where applicable (Supplementary Table 2). To model the cage experiments, 150 female and 150 male wildtype adults (or 300 females and 270 males for 10% release of males only) along with 150 female and 150 male heterozygotes (or no females and 30 males for 10% release) are initially present. Females may fail to mate, or mate once in their life, with a male of a given genotype according to its frequency in the male population, chosen randomly with replacement such that males may mate multiple times. The number of eggs from each mated female is multiplied by the egg production of the male relative to the wildtype male, to account for experimental observations of reduced egg production from SDGD fathers. The eggs hatch or not with a probability that depends on the product of larval hatching values from the mother and father,

relative to wildtype. To start the next generation, 600 larvae are randomly selected, unless less than 600 larvae have hatched, in which case the smaller amount initiates the next generation, following experiment. The probability of subsequent survival to adulthood is assumed to be equal across genotypes. Assuming very large population sizes gives results for the genotype frequencies that are indistinguishable from the deterministic model. For the deterministic egg count, we use the large population limit of the stochastic model.

Population dynamics model (continuous time). To model changing population sizes in the field (for Supplementary Fig. 1), we use a continuous-time population dynamics model with one life stage and logistic density-dependence in the recruitment rate based on models developed previously[36,37]. Here, $n(t)$ represents the abundance of adult individuals, $f(t)$ and $m(t)$ as the total abundances of adult females and males, and $f_{ij,pq}(t)$ and $m_{ij,qY}(t)$ are the genotype abundances where, as above, the first set of indices denotes alleles at the target site and the second set denotes the sex chromosomes. Populations are normalized with the pre-release wildtype population size such that $n(t=0)=1$ and time is continuous and measured in generations. The dynamics of the total population size are given by the following differential equation:

$$\frac{dn(t)}{dt} = 2\left(\frac{R_m}{1+2(R_m-1)\,\overline{w}_m(t)\overline{w}_f(t)f(t)}\right)\overline{w}_m(t)\overline{w}_f(t)f(t) - n(t)$$

The total recruitment rate of adults incorporates a density-dependent factor (bracketed term) based on Deredec et al. [5], and depends upon the total numbers of females, $f(t)$, times the average female fitness, $\overline{w}_f(t) = \sum_{i=1}^{3}\sum_{j=i}^{3}\sum_{pq=xx;xX;XX}\left(w_{ij,pq}\, f_{ij,pq}(t)/f(t)\right)$. Since SDGD males may have reduced fertility, the recruitment rate is also dependent on the average male fitness, $\overline{w}_m(t) = \sum_{i=1}^{3}\sum_{j=i}^{3}\sum_{q=x,X}\left(w_{ij,qY}\, m_{ij,qY}(t)/m(t)\right)$ (we assume the number of males is not limiting and that all males participate in mating). R_m is the intrinsic growth rate of the population per generation at low density.

The equations for the individual genotype populations for females and males are:

$$\frac{df_{ij,pq}(t)}{dt} = 2\left(\frac{R_m\overline{w}_m(t)\overline{w}_f(t)f(t)}{1+2(R_m-1)\overline{w}_m(t)\overline{w}_f(t)f(t)}\right)\left(1-\frac{\delta_{ij}}{2}\right)\left(1-\frac{\delta_{pq}}{2}\right)\Big(e_{i,p}(t)s_{j,q}(t) + e_{j,p}(t)s_{i,q}(t) + e_{i,q}(t)s_{j,p}(t) + e_{j,q}(t)s_{i,p}(t)\Big) - f_{ij,pq}(t)$$

$$\frac{dm_{ij,qY}(t)}{dt} = 2\left(\frac{R_m\overline{w}_m(t)\overline{w}_f(t)f(t)}{1+2(R_m-1)\,\overline{w}_m(t)\overline{w}_f(t)f(t)}\right)\left(1-\frac{\delta_{ij}}{2}\right)\Big(e_{i,q}(t)s_{j,Y}(t) + e_{j,q}(t)s_{i,Y}(t)\Big) - m_{ij,qY}(t)$$

Egg and sperm proportions $e_{k,l}(t)$ and $s_{k,l}(t)$ are as defined in (1a) and (1b) in the discrete-generation model above, with $f_{ij,pq}(t)$ and $m_{ij,pq}(t)$ instead of frequencies $F_{ij,pq}(t)$ and $M_{ij,pq}(t)$.

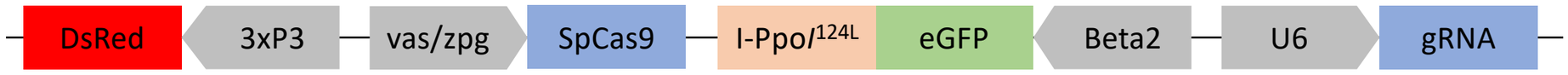
All calculations are carried out using Wolfram Mathematica[38].

Methods-only References

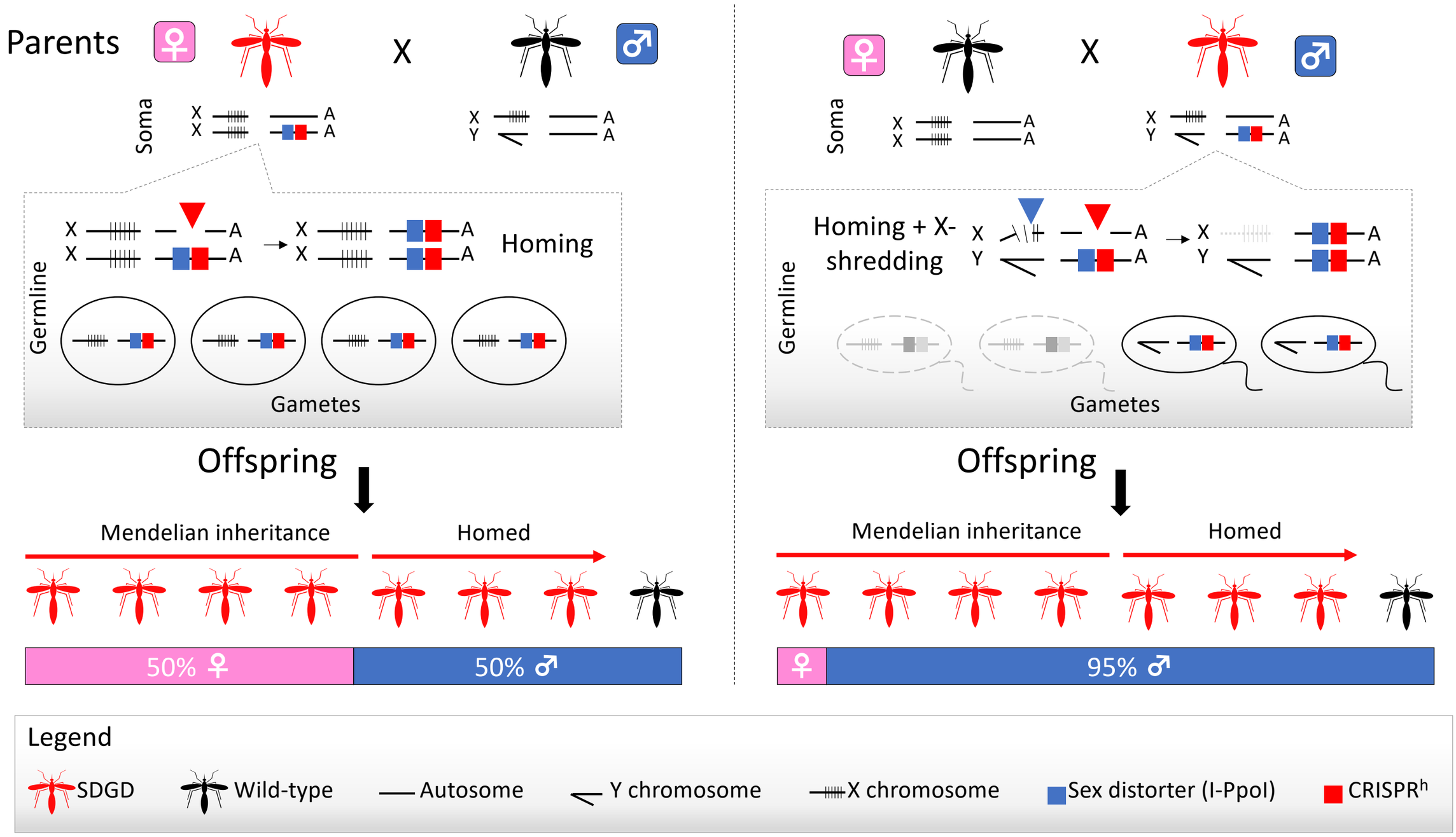
27. Fuchs, S., Nolan, T. & Crisanti, A. Mosquito Transgenic Technologies to Reduce Plasmodium Transmission. in *Malaria: Methods and Protocols* (ed. Ménard, R.) 601–622 (Humana Press, 2013). doi:10.1007/978-1-62703-026-7_41.

28. Volohonsky, G. *et al.* Tools for Anopheles gambiae Transgenesis. *G3amp58 GenesGenomesGenetics* **5**, 1151–1163 (2015).
29. Scott, T. W. Containment of Arthropod Disease Vectors. *ILAR J.* **46**, 53–61 (2005).
30. Akbari, O. S. *et al.* Safeguarding gene drive experiments in the laboratory. *Science* **349**, 927–929 (2015).
31. Knudsen, S. Promoter2.0: for the recognition of PolII promoter sequences. *Bioinforma. Oxf. Engl.* **15**, 356–361 (1999).
32. Reese, M. G., Harris, N. L. & Eeckman, F. H. Large scale sequencing specific neural networks for promoter and splice site recognition. in *Biocomputing: Proceedings of the 1996 pacific symposium* 737–738 (World Scientific, 1996).
33. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
34. Mollahosseini, A. *et al.* A user-friendly software to easily count Anopheles egg batches. *Parasit. Vectors* **5**, 122 (2012).
35. 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 (2008).
36. Beaghton, A., Beaghton, P. J. & Burt, A. Vector control with driving Y chromosomes: modelling the evolution of resistance. *Malar. J.* **16**, 286 (2017).
37. Beaghton, A. *et al.* Requirements for Driving Antipathogen Effector Genes into Populations of Disease Vectors by Homing. *Genetics* **205**, 1587–1596 (2017).
38. Wolfram Research, Inc. *Mathematica*. (Wolfram Research, Inc., 2019).

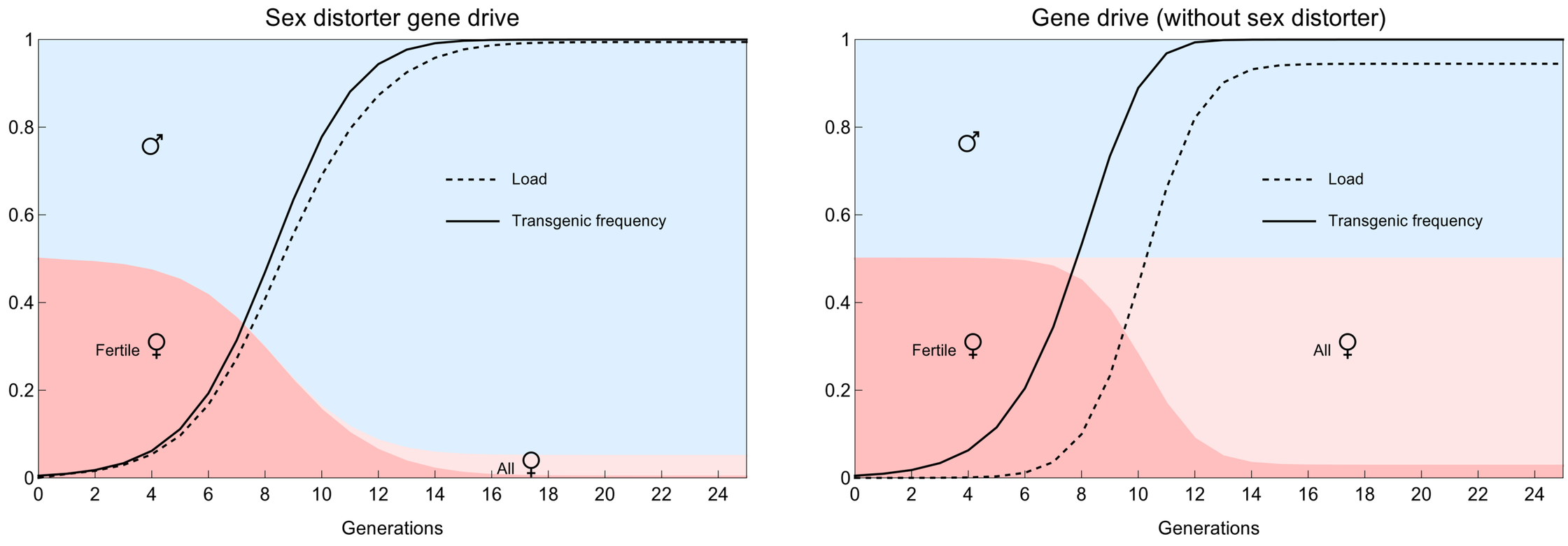
a)

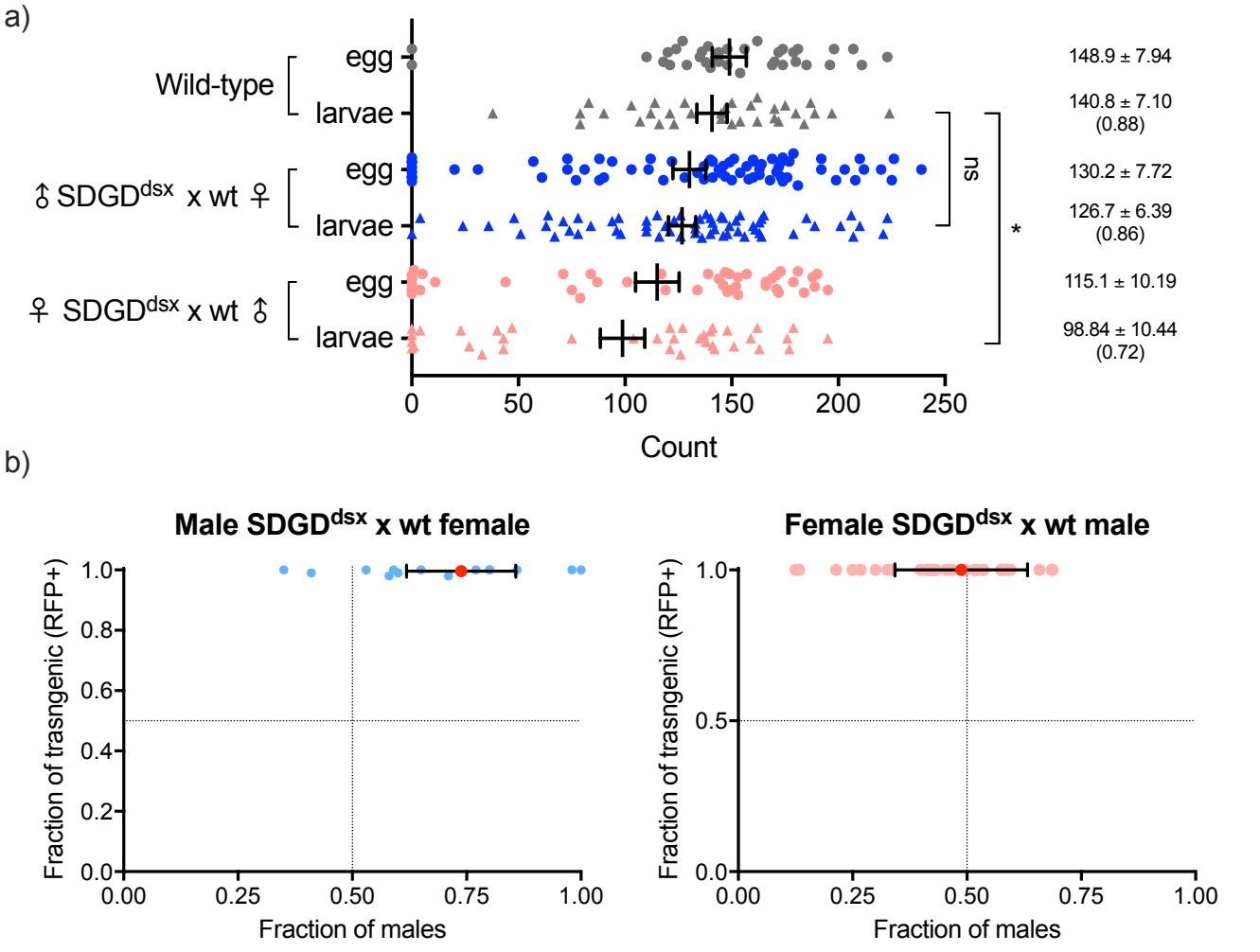

DsRed
3xP3
vas/zpg
SpCas9
I-Ppo/124L
eGFP
Beta2
U6
gRNA

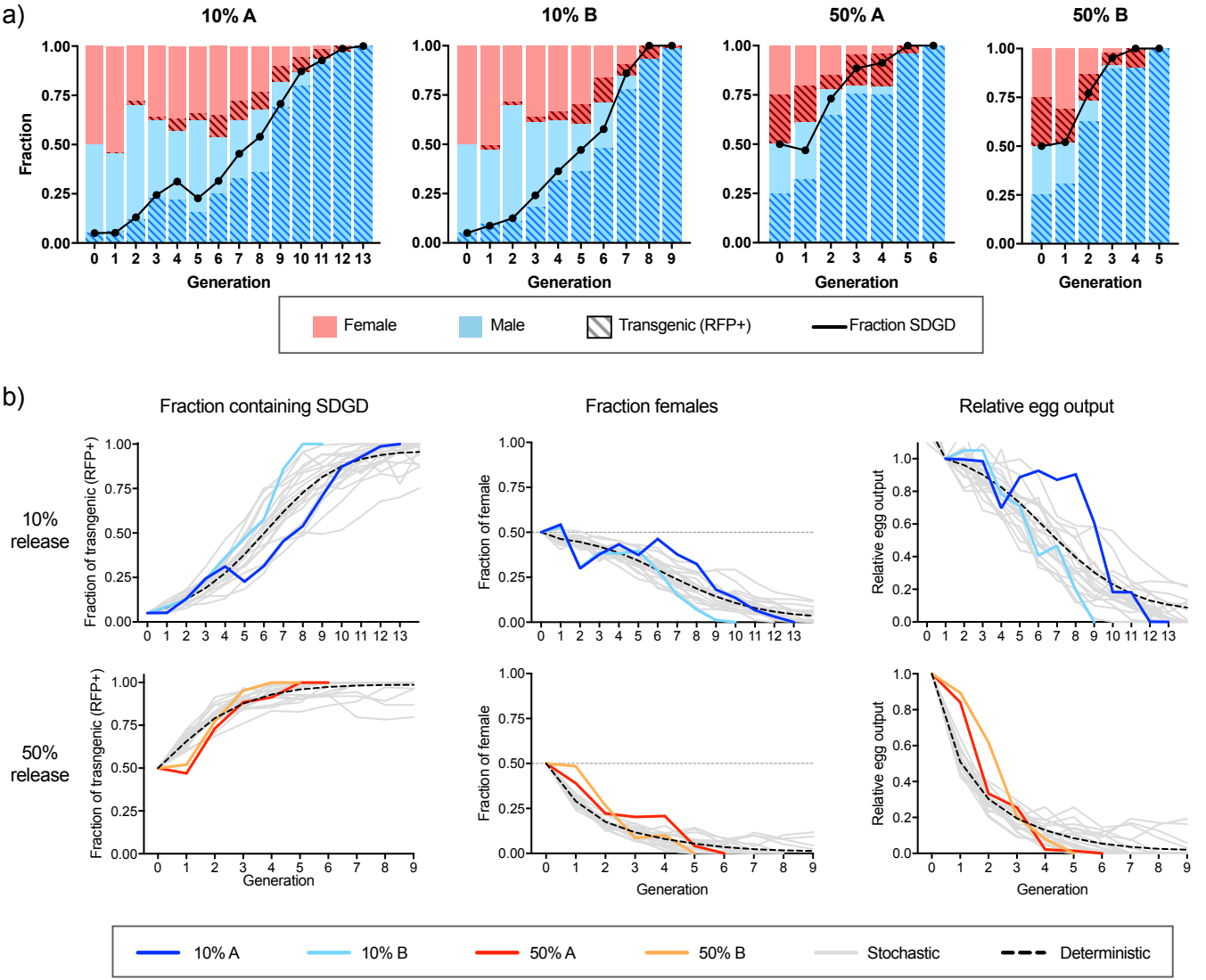
b)

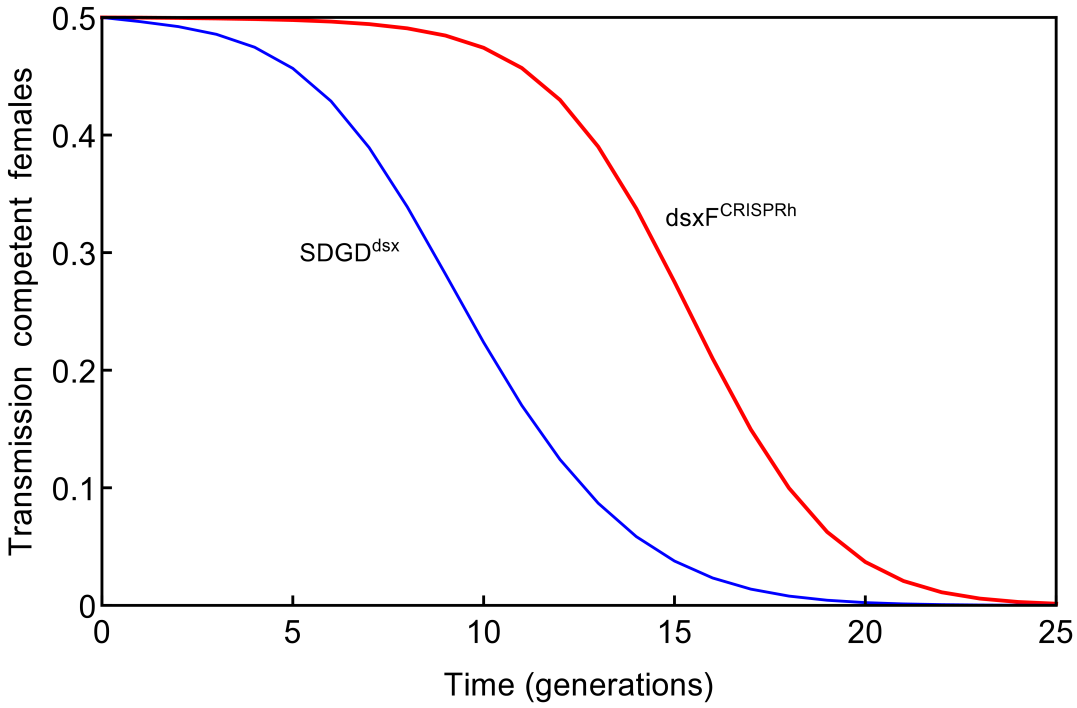

Parents
X
Soma
Germline
Homing
Gametes
Offspring
Mendelian inheritance
Homed
50% ♀
50% ♂
Homing + X-shredding
95% ♂
Legend
SDGD
Wild-type
Autosome
Y chromosome
X chromosome
Sex distorter (I-PpoI)
CRISPRh

c)


Sex distorter gene drive
Gene drive (without sex distorter)
Load
Transgenic frequency
Fertile ♀
All ♀
Generations


a)
Wild-type
egg
larvae
♂ SDGD^dsx x wt ♀
egg
larvae
♀ SDGD^dsx x wt ♂
egg
larvae
ns
*
148.9 ± 7.94
140.8 ± 7.10
(0.88)
130.2 ± 7.72
126.7 ± 6.39
(0.86)
115.1 ± 10.19
98.84 ± 10.44
(0.72)
0
50
100
150
200
250
Count
b)
Male SDGD^dsx x wt female
Fraction of trasngenic (RFP+)
1.0
0.8
0.6
0.4
0.2
0.0
0.00
0.25
0.50
0.75
1.00
Fraction of males
Female SDGD^dsx x wt male
Fraction of trasngenic (RFP+)
1.0
0.5
0.0
0.00
0.25
0.50
0.75
1.00
Fraction of males

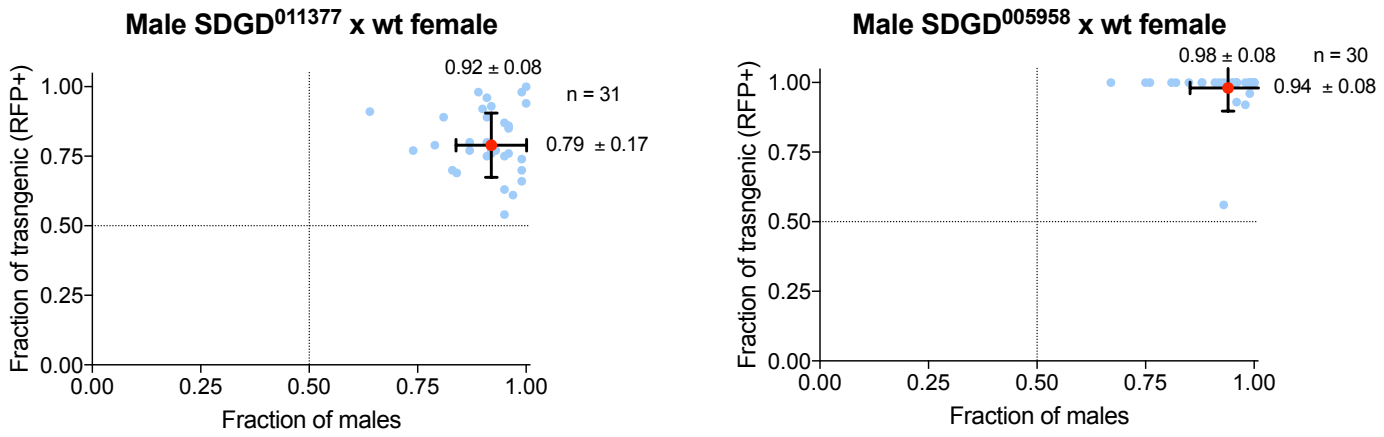

a)
10% A
10% B
50% A
50% B
Fraction
Generation
Female
Male
Transgenic (RFP+)
Fraction SDGD
b)
Fraction containing SDGD
Fraction females
Relative egg output
10% release
50% release
Fraction of trasngenic (RFP+)
Fraction of female
Relative egg output
Generation
10% A
10% B
50% A
50% B
Stochastic
Deterministic



Supplementary Figure 1

Model prediction of the reduction in abundance of transmission competent females.

Model prediction of the reduction over time in abundance of transmission competent (i.e. biting) females, normalised by the initial total population size, for the SDGD^{dsx} construct compared to dsxF^{CRISPRh} (Kyrou et al., 2018), using a continuous-time population dynamics model (see Supp. Modelling Methods) for 'field' release of 1% heterozygous transgenic males in male population. The SDGD^{dsx} construct is predicted to suppress the population of transmission competent females faster than the dsxF^{CRISPRh}, mainly due to the creation of a male bias in the population by the sex distorter). Parameters used for SDGD^{dsx} are in Table S2; dsxF^{CRISPRh} parameters were estimated from Kyrou et al. (2018) using an average W/D female fitness of 0.4335; for both, R_m (intrinsic growth rate per generation) = 6. At long times (not shown), the SDGD^{dsx} population rebounds to an intermediate equilibrium (suppressed) population.

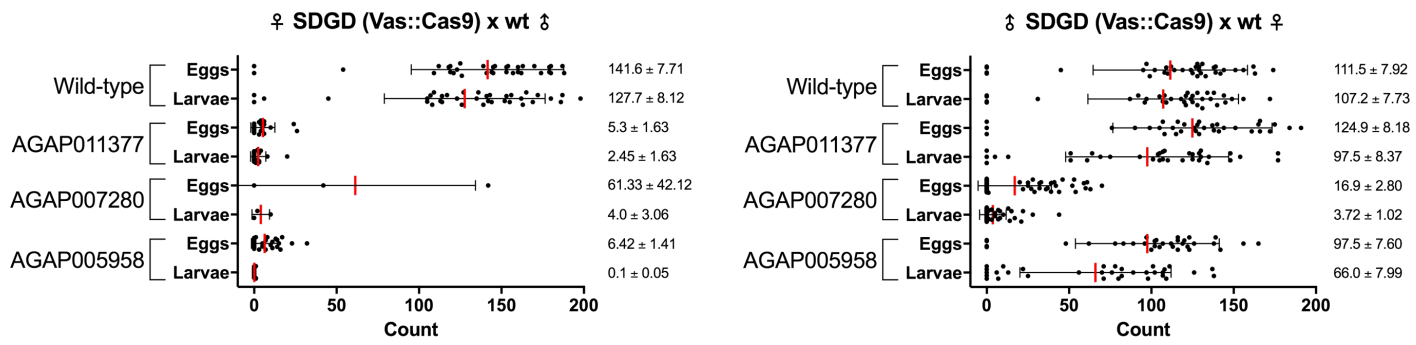


Supplementary Figure 2

Sex and inheritance bias caused by SDGD^{011377} and SDGD^{005958} males.

Scattered plots showing the fraction of transgene inheritance (y-axis) against male bias (x-axis) in the progeny of heterozygous male SDGD^{011377} (left-hand panel) SDGD^{005958} and (right-hand panel) crossed to wild type females. Individual coloured dots represent the progeny derived from a single female and the red dots indicate the average of the population (with respective values indicated next to the plot ± s.e.m.). Error bars indicate standard deviation. SDGD at both loci showed a high transmission rate of the transgene determined by scoring in the progeny the presence of RFP marker that is linked to the SDGD allele. The progeny of SDGD/+ at both loci showed a strong sex ratio distortion towards males. Dotted lines indicate expected Mendelian inheritance.

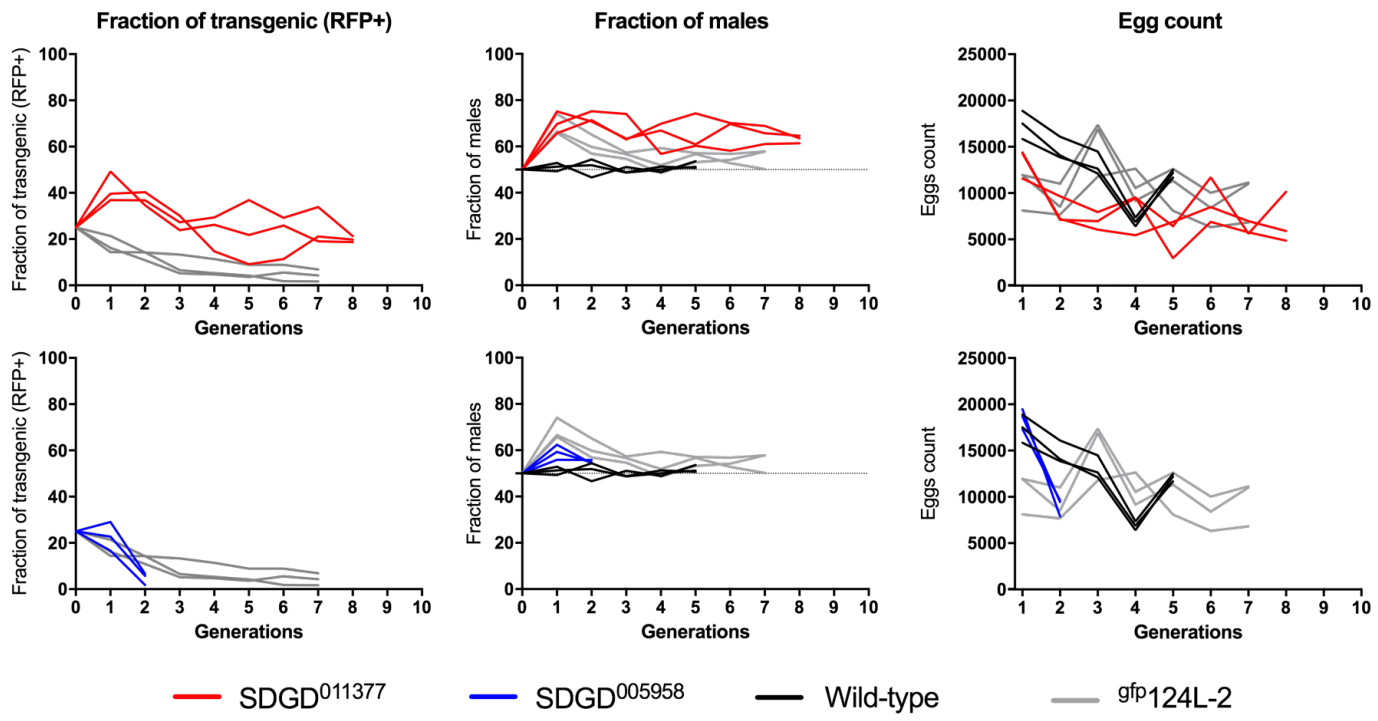
.



Supplementary Figure 3

Fecundity phenotype of SDGD targeting 3 different fertility loci in *An. gambiae*

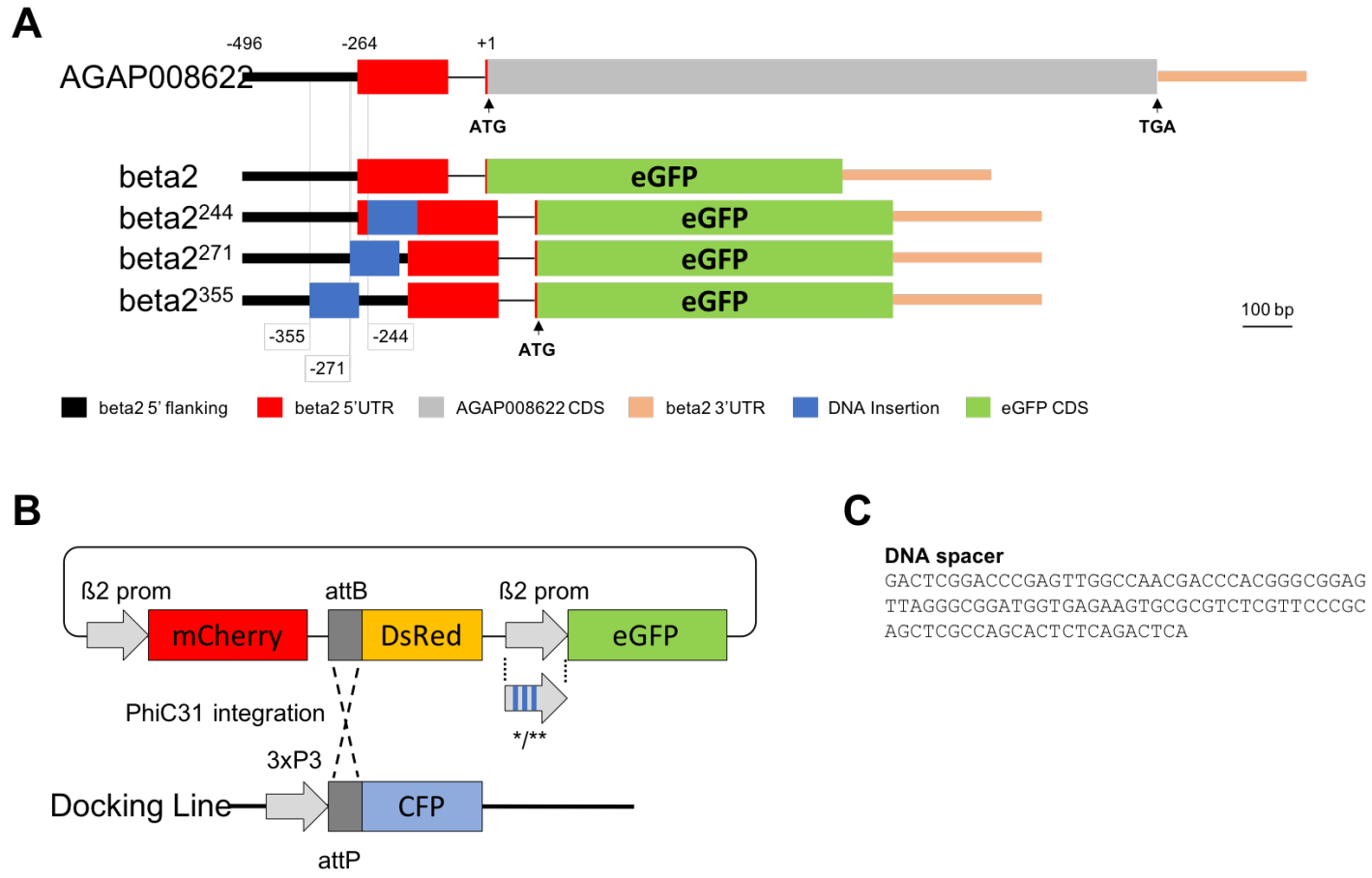
SDGD constructs expressing the Cas9 nuclease under the control of the *Vas2* promoter were generated targeting the fertility loci *AGAP011377*, *AGAP007280* and *AGAP005859* (as indicated). SDGD heterozygous male and female were crossed to wild-type counterparts. Each dot represents progeny of individual females. Fecundity was measured by counting the number of eggs per female and the hatched larvae. Values on the right represent average ± s.e.m. A strong fertility effect was observed in heterozygous SDGD females at 3 loci, while male fecundity was strongly impaired by targeting 7280 and 5859 loci. Vertical red bars indicate average count, and error bars indicate standard deviation. A minimum of 20 females were analysed for each cross.



Supplementary Figure 4

Kinetics of SDGD^{011377} and SDGD^{005958} spread in target mosquito populations.

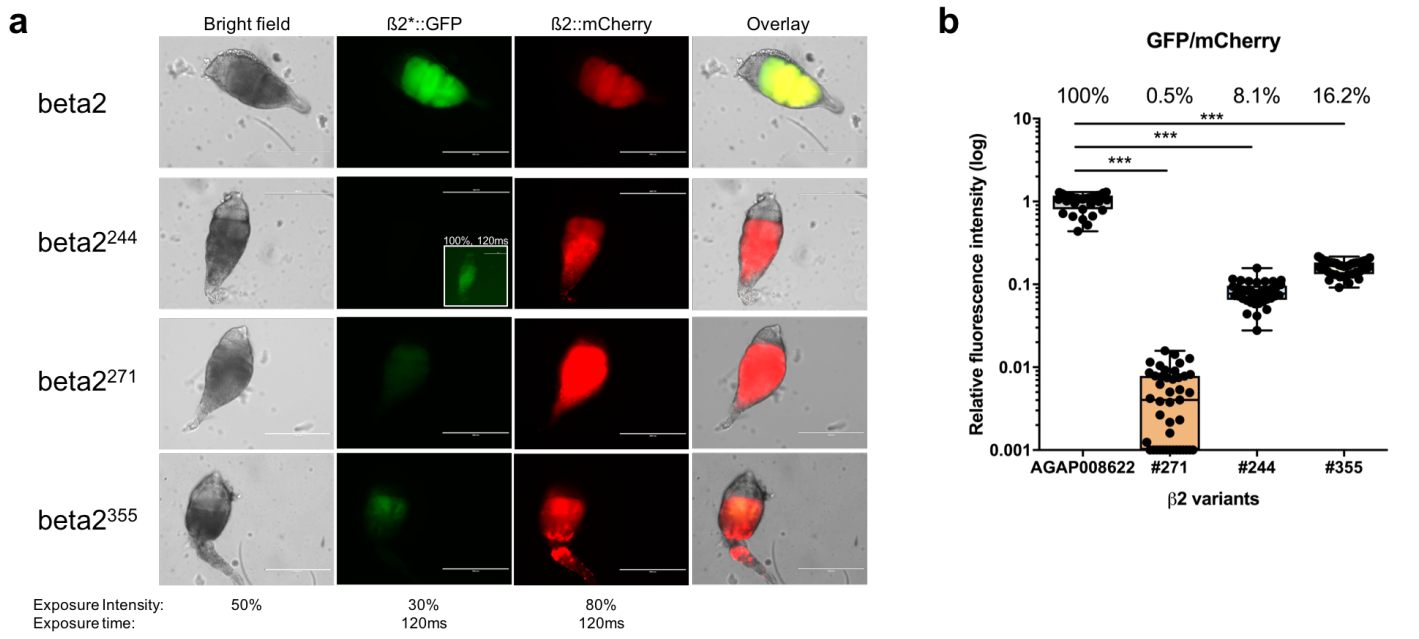
In these experiments 100 heterozygous transgenic males were introduced into a population of 100 wild-type males and 200 wild-type females (transgenic allele frequency of 12.5%). The frequency of the transgene was monitored every generations together with the fraction of males in the population and the total number of eggs laid. Each consecutive generation was established by collecting 450 eggs. The frequency of the SDGD^{011377} (red lines) and SDGD^{005958} (blue lines) was compared to that of the autosomal self-limiting sex-distorter ^{gfp}124L-2 (grey lines) (Galizi et al., 2014) as well as to that of wild-type populations (black lines) as control. Each genotype was tested in triplicate cages. The SDGD^{005958} allele disappeared from the populations at generation 2 due to the strong fertility effects. The SDGD^{011377} alleles persisted in the populations despite the fertility effects but failed to increase over the frequency of release on subsequent generations. The fraction of males in the population was stably biased to about 65%.



Supplementary Figure 5

Mutagenesis of the beta2 tubulin promoter.

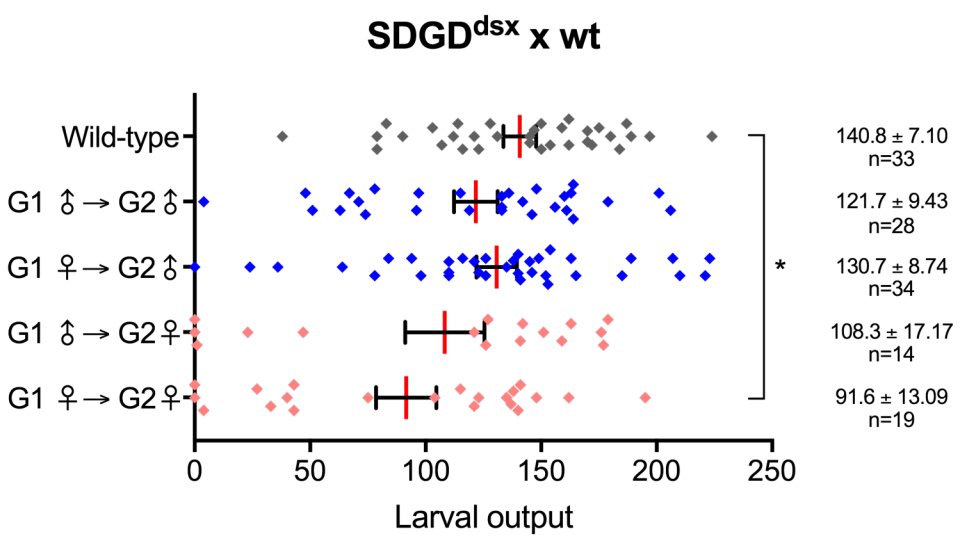
A) Schematic overview of the *AGAP008622* (*beta2-tubulin*) locus and the three *beta2* promoter variants generated by placing a 100bp GC-rich DNA sequence (blue rectangle) upstream of the start codon at position 244, 271 and 355. B) The double-fluorescence reporter assay developed to detect the effects of the *beta2* promoter modifications on its transcription level. A construct for each modification was generated harbouring the modified *beta2* promoter (stars key) driving an eGFP marker, while a second, unmodified, *beta2* promoter was driving a mCherry protein. The constructs were integrated within the same autosomal docking line by PhiC31-mediated integration by replacing a 3xP3::CFP cassette with a 3xP3::DsRed as integration marker. C) The sequence of 100bp GC-rich DNA region inserted at the 5' of the *beta2* promoter.



Supplementary Figure 6

GFP and mCherry signal quantification from mosquito testes transformed with modified *beta2* promoters.

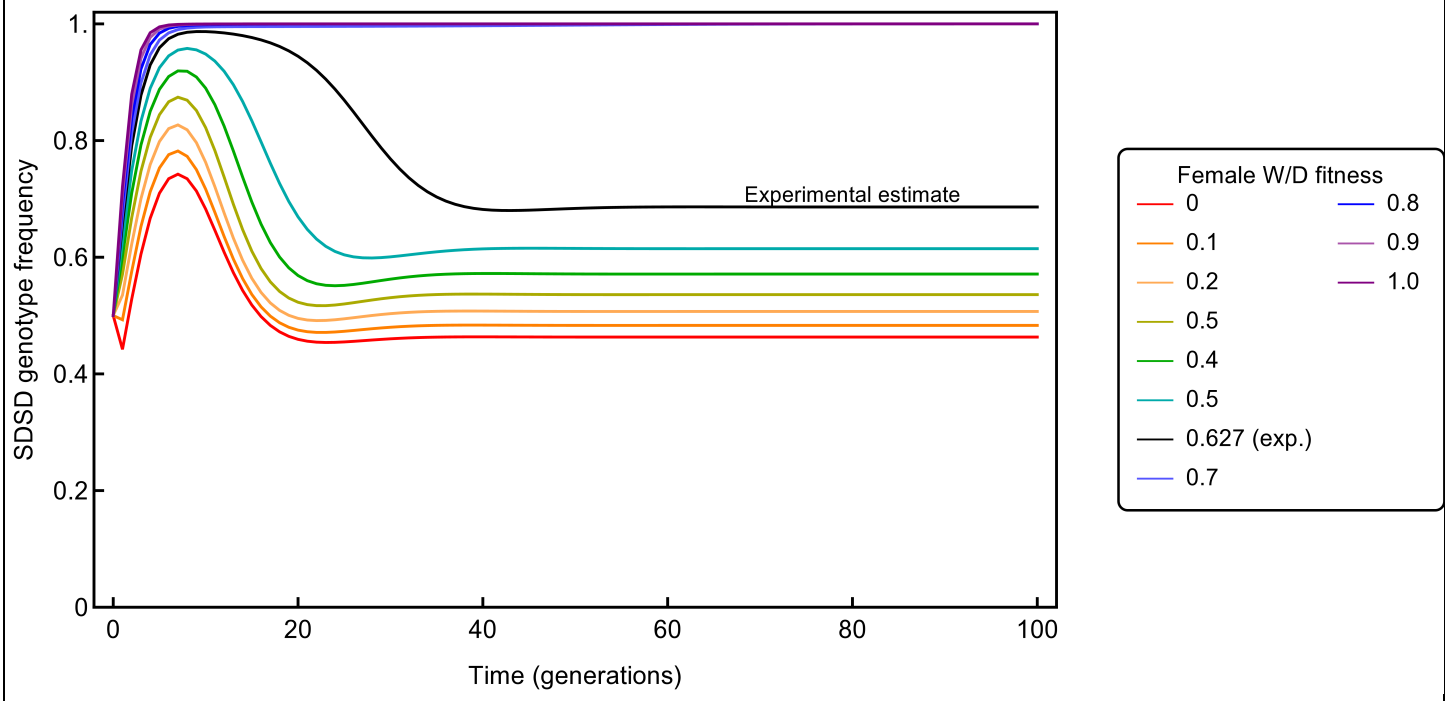
a) Dissected testis from beta2^{wt}, beta2^{271}, beta2^{244} and beta2^{355} transgenic lines containing a single heterozygous insertion were microphotographed under the same exposure settings (as indicated), using mCherry as internal control. Line beta2^{271} showed GFP fluorescence intensity comparable to background; GFP expression was detectable at increased exposure (100% gain, 120ms, inset). Scale bar, 200 µm. b) Quantification of fluorescence intensity as GFP/mCherry ratio, normalized to the beta2^{wt} control (100%). Average relative intensity is indicated above the bars. *** indicates P value < 0.001 (ordinary One-way ANOVA). A minimum of 31 testes were analysed from individual expressing each promoter variant.



Supplementary Figure 7

Maternal or paternal contribution to the fecundity of the SDGD^{dsx} allele

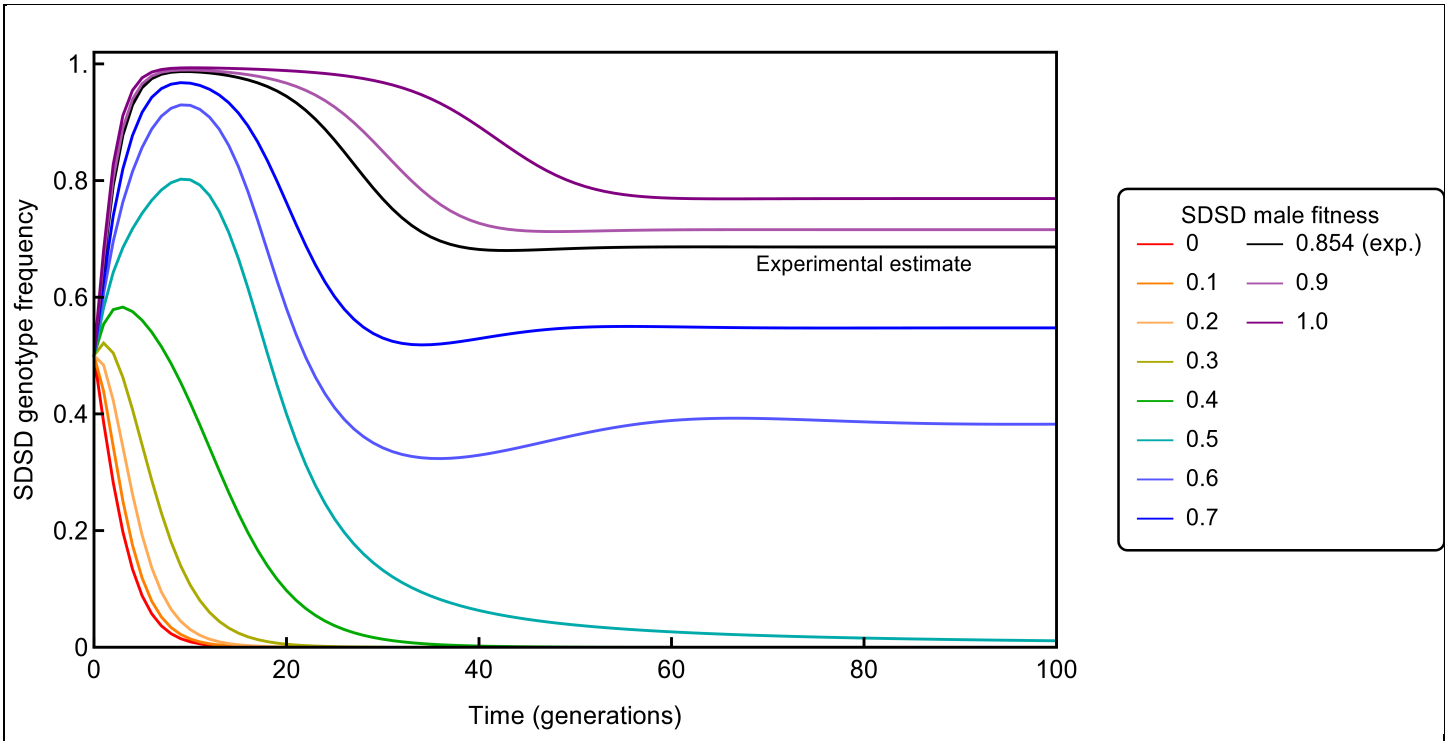
Male and female SDGD^{dsx} heterozygotes that had inherited a maternal or paternal copy of the SDGD^{dsx} allele were crossed to wild type counterparts and assessed for fecundity. The total larval output is plotted for individual females (dots). Red bars indicate the average and the mean count (± s.e.m.) is shown. Females inheriting the transgene from the mother (G1♀ → G2♀) have significantly lower larval progeny (* P = 0.0256, Kruskal-Wallis test) compared to wild-type control.



Supplementary Figure 8

Time dynamics of the frequency of SDGD^{dsx} as function of female fitness

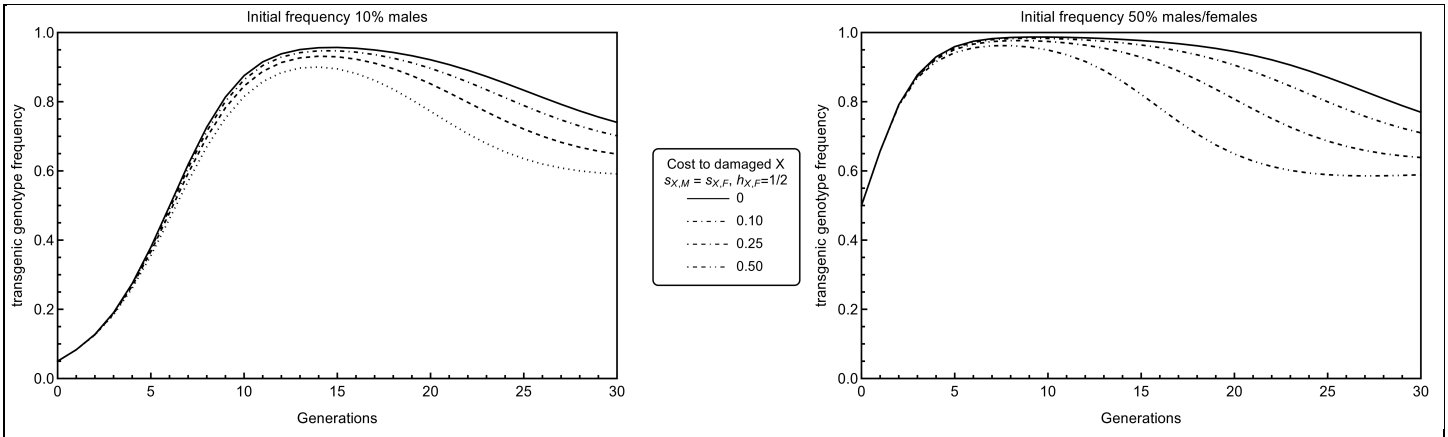
Time dynamics of the frequency of SDGD^{dsx} transgenic individuals in the population as a function of W/D (where D represent the SDGD allele and W the wild-type allele) female fitness ($w_{\mathrm{WD,xx}}$), as predicted by the deterministic discrete-generation model at 25% initial allelic frequency. The graph shows the frequency of SDGD heterozygote males and females as a proportion of the male (or female) population, with other parameter estimates and baseline values given in Supp Table 2 (SDGD male fitness = 0.854; m (sex distortion) = 0.93). The predicted outcome at high W/D female fitness is elimination of the population, and at lower fitness, an intermediate equilibrium with W, R and D alleles. The result (black line) for the experimental estimate for female SDGD heterozygote fitness, $w_{\mathrm{WD,xx}} = 0.627$, is in a parameter region where even a small (positive) change leads to a prediction of population elimination instead of suppression.



Supplementary Figure 9

Time dynamics of the frequency of SDGD^{dsx} as function of male fitness

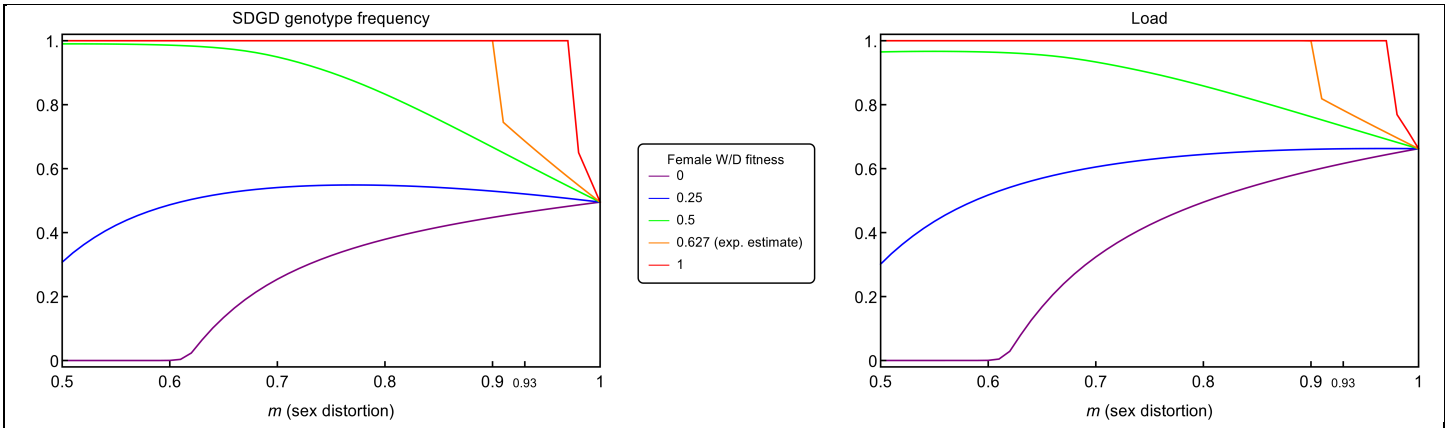
Time dynamics of the frequency of SDGD^{dsx} transgenic individuals in the a population as a function of the fitness of SDGD males (assume $w_{\mathrm{WD,xY}} = w_{\mathrm{DR,xY}} = w_{\mathrm{DD,xY}}$) as predicted by the deterministic discrete-generation model. Initial release is 50% of SDGD heterozygote males and females as a proportion of the male (or female) population, with other parameter estimates and baseline values given in Supp Table 2 (W/D female fitness $w_{\mathrm{WD,xx}}$ = 0.627; m = 0.93). For low SDGD male fitness ($<\approx 0.5$), the construct is eventually lost.



Supplementary Figure 10

Impact of fitness due to damaged X chromosome.

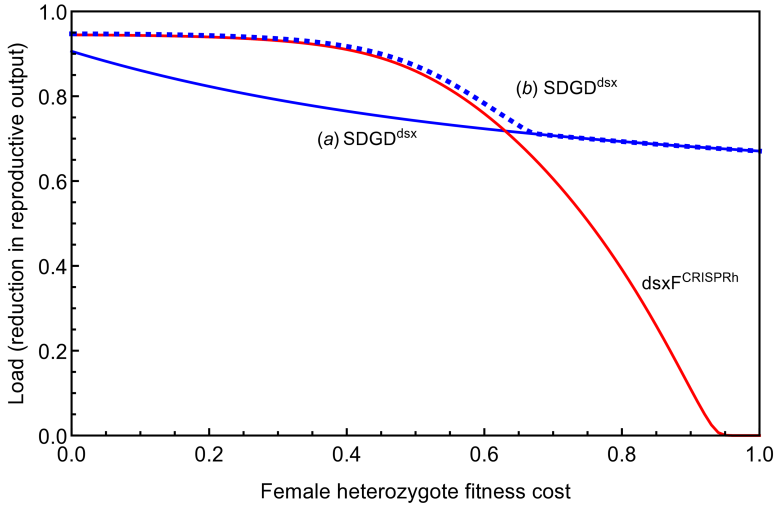
Modelling the impact of fitness reduction, due to the inheritance of damaged X chromosomes passed through X-shredding in a SDGD male, on the spread of the SDGD transgenics as predicted by the deterministic discrete-generation cage model. For simplicity, we assume that the additional cost to carrying one damaged X chromosome in males is the same as the cost in females that carry two copies of the damaged X ($s_{\mathrm{X,f}} = s_{\mathrm{X,m}}$), and females with one damaged X and one wildtype X chromosome have only half the fitness cost of females with two copies (dominance coefficient $h_{\mathrm{X,f}} = 1/2$). Estimates used for other parameters given in Supplementary Table 2.

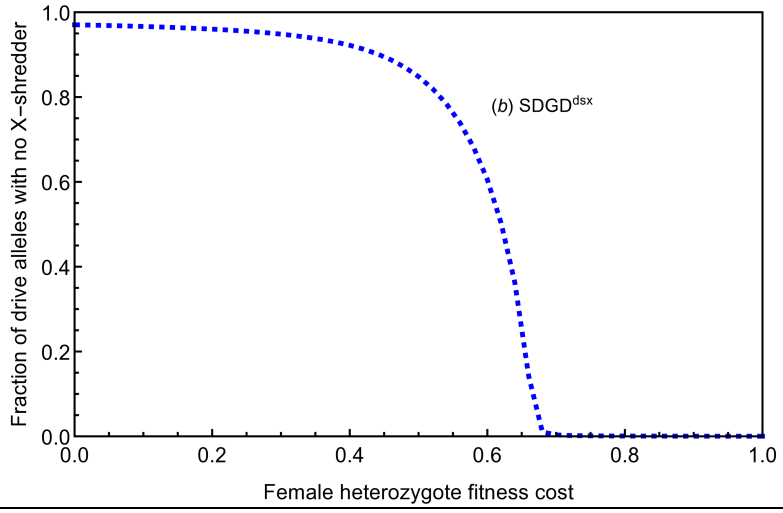


Supplementary Figure 11

Effect of female fitness on SDGD frequency and population load.

Effect of female W/D heterozygote fitness (D refers to the SDGD^{dsx} allele) on the SDGD genotype frequency (i.e., individuals with at least one copy of the SDGD) and load after 200 generations, as predicted by the deterministic discrete-generation model. Parameter estimates and baseline values given in Supp Table 2 (SDGD male fitness $= 0.854$). The possible outcomes (load $= 1$ and population elimination; intermediate equilibrium with W, D and R; or load $= 0$ and construct lost) depend upon the sex distortion *m* (0.5 [no sex bias] to 1 [only male progeny]) and the female W/D heterozygote fitness ($0 \leq w_{\mathrm{WD,xx}} \leq 1$). For low female fitness, adding an X-shredder (increasing sex distortion m) is predicted to beneficially increase the load on the population. For complete female heterozygous sterility (purple line), the drive construct disappears from the population if there is no sex distortion (*m* = 1/2), whereas sufficiently increasing the sex distortion allows the SDGD^{dsx} to spread and impose a load on the population. The presence of the X-shredder keeps the construct predominantly in males, therefore mitigating the heterozygous female sterility effect. By contrast if female fitness is sufficiently high (lines of fitness 0.5 or greater on plot), the load decreases at high sex distortion because due to male bias, SDGD^{dsx} males replace high-fitness W/D heterozygote females which decreases the ability of the construct to spread. For a complete male sex bias, *m* = 1, no W/D heterozygote females are created (and therefore no female/male SDGD^{dsx} homozygotes), since all X chromosomes are shredded and SDGD^{dsx} males have no female progeny; thus the load at *m* = 1 does not depend on the fitness of female W/D heterozygote individuals since none are present. Only SDGD^{dsx} heterozygous males can pass on the construct, with the SDGD^{dsx} allele present in reduced frequency in an intermediate equilibrium with wildtype and resistance alleles and a load less than one for these parameters. (In general, the amount of reduction in SDGD^{dsx} frequency and load at *m* = 1 will depend on parameters such as the homing rate, here less than 100%, the rate of resistant R mutation, and the relative fertility of the R and SDGD^{dsx} allele).

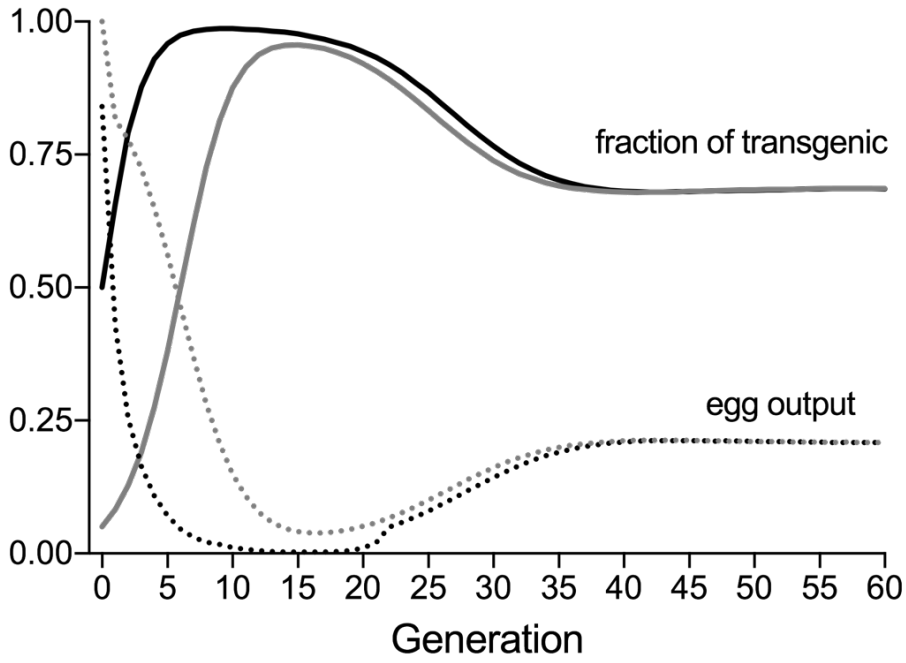




Supplementary Figure 12

Comparison of the predicted equilibrium for release of SDGD^{dsx} or dsxF^{CRISPRh} into a wild-type population

(Top) Comparison of the predicted equilibrium load (i.e., the reduction in reproductive output by the population after 400 generations) for release of SDGD^{dsx} or dsxF^{CRISPRh} into a wild-type population, varying the fitness cost of heterozygote W/D females. The discrete generation model predicts that the SDGD^{dsx} construct (blue solid line) is more robust to reductions in female heterozygote fitness compared to dsxF^{CRISPRh} (red line), still maintaining a substantial load even at 100% reduction in female heterozygote fitness (i.e. females heterozygous for the drive are completely non-viable). We also consider the possibility that the X-shredder component may be lost from the SDGD^{dsx} construct during homing (bottom panel, blue dashed line), such that out of the drive alleles transmitted from female or male W/D individuals, 0.01% will not have a functioning X-shredder component. For low to mid-fitness costs, the predicted load is similar to that of the dsxF^{CRISPRh} drive-only construct since the fraction of drive individuals without an intact X-shredder is high; for high fitness costs, the load merges with that of intact SDGDdsx (blue solid line) since almost all drive individuals have an intact X-shredder. We use representative parameters for both constructs for comparison (drive transmission $d_f = d_m = 0.95$ for both males and females; rate of resistance $u_f = u_m = 0.5$; X-shredding parameter for SDGD^{dsx} is m=0.95; no reduction in fitness for heterozygote females or males).



Supplementary Figure 13

Time dynamics of the frequency of SDGDdsx transgenic individuals and relative eggs output

Time dynamics of the frequency of SDGD^{dsx} transgenic individuals (solid lines) and relative eggs output (dotted lines) in the population as predicted by the deterministic discrete-generation model using experimental parameters given in Supp Table 2 and assuming two initial releases of 50% of SDGD heterozygote males and females (black line) or 10% SDGD heterozygous males only (grey line). Independently of the release scenarios, for these parameters, the frequency of transgenic individuals reaches an intermediate equilibrium while W, R and D alleles and the egg output is reduced (population suppressed).