Current Biology, Volume *23*

Supplemental Information

A Synthetic Gene Drive System for Local, Reversible Modification and Suppression of Insect Populations

Omar S. Akbari, Kelly D. Matzen, John M. Marshall, Haixia Huang, Catherine M. Ward, and Bruce A. Hay

Supplemental Inventory

Supplemental Experimental Procedures

Construction of the Base UD^{MEL} Plasmid

The Drosophila alpha-Tubulin 3'UTR was PCR amplified from genomic DNA using primers 5'-ATG CTA GCG GAT CCG GGA ATT GGG AAT TGG GCA ATA TTT AAA TAA AGA AAA ACA GTG GGG TTT-3' and 5'-CAT CAT CAT CAT CCG AAT TCC ACA GGC CGG CCA TTG GCG CGC CGC GTC ACG CCA CTT CAA CGC-3', producing a 363bp PCR product containing unique *Asc*I and *Fse*I5' restriction sites into which specific miRNA toxins could be cloned. These primers overlap the *Drosophila* attP plasmid [1] and the 3' end of the *Bicaudal C* promoter. The *Bicaudal C* promoter fragment was amplified from genomic DNA with primers 5'-GTT GAA GTG GCG TGA CGC GGC GCG CCA ATG GCC GGC CTG TGG AAT TCG GAT GAT GAT GAT GAT-3' and 5'-TTC AAC GCA CAC TTA TTA CGT GAG CGA TCG CAT CGC ATA ATT ATA TAA TAA TAA ACT GCA TGC CGC CA-3', producing a 2903bp PCR product. This product contains a unique *AsiS*I restriction site at the 5'end of the *Bicaudal C* promoter. The primers used to generate the fragment overlap the 5' end of alpha-Tubulin 3' UTR and the gypsy insulator. The gypsy insulator was PCR amplified from genomic DNA using primers 5'-ATG CAG TTT ATT ATT ATA TAA TTA TGC GAT GCG ATC GCT CAC GTA ATA AGT GTG CGT TGA-3' and 5'-GAG GCG TCC AGG ATC CCA TGG GGT TCA TCT AAT GTT TAA ACA ATT GAT CGG CTA AAT GGT ATG-3', producing a 356bp PCR product. This product contains a unique *Pme*I site between the gypsy insulator and the *bnk* promoter, and these primers overlap the 5'end of the *Bicaudal C* promoter and the 5' end of the *bnk* promoter. The *bnk* promoter was PCR amplified with primers 5'-TTT TCT TGC CAT ACC ATT TAG CCG ATC AAT TGT TTA AAC ATT AGA TGA ACC CCA TGG GAT C-3' and 5'-CGT GAC CTA CAT CGT CGA CAC TAG TGG ATC GCT AGT TAA TTA AGC CGA ATT CGT TGA CGG TTG A-3', producing a 787bp PCR product which carries a unique *Pac*I site on the 3' end of the *bnk* promoter for cloning in the rescue fragments. These primers also overlap the gypsy insulator and the *Drosophila* attP plasmid backbone. These 4 PCR products were purified with Qiagen (Valencia, CA) PCR purification columns and ligated using 1-step recombination technology [2] into an *Asc*I digested Drosophila attP plasmid [1], producing the UD^{MEL}-BicC-Gyp-Bnk-attb plasmid

Construction of miRNA Toxins

The Drosophila miRNA mir6.1 stem-loop was modified to target *dah*, *o-fut1*or *myd88* as described previously [3]. To make a miRNA that targets *dah* site 1, primers *dah*-1-f (5'-CTT AAT CAC AGC CTT TAA TGT AGG GAA ATA TAT AAC AAT ACA CTA AGT TAA TAT ACC ATA TCT-3') and dah-1-r (5'- ATG TTA GGC ACT TTA GGT ACA GGG AAA TAT ATA ACA ATA AAC TAG ATA TGG TAT ATT AAC TTA G -3') were generated. To target *dah* site 2, primers *dah* -2-f (5'- TTA AAC TTA ATC ACA GCC TTT AAT GTA ACC AGG ATG CGA ACT ATA CAC TAA GTT AAT ATA CCA TAT CTA G -3') and *dah* -2-r (5'- AAT GAT GTT AGG CAC TTT AGG TAC AAC CAG GAT GCG AAC TAT AAA CTA GAT ATG GTA TAT TAA CTT AG-3') were generated. To target *o-fut1* target site 1, primers *o-fut1*-1-f (5'-AAA CTT AAT CAC AGC CTT TAA TGT AGT TTT ATT ACA TTG ATT ACG CTA AGT TAA TAT ACC ATA TCT AG -3') and *o-fut1*-1-r (5'- AAT GAT GTT AGG CAC TTT AGG TAC AGT TTT ATT ACA TTG ATT AAG CTA GAT ATG GTA TAT TAA CTT AGC G -3') were generated. To target *o-fut1* target site 2, primers *o-fut1*-2-f (5'- ATC ACA GCC TTT AAT GTC AGG ATT ATC TAC TTA AAT CCT TAA GTT AAT ATA CCA TAT CTA AGT-3') and *o-fut1*-2-r (5'- ATG ATG TTA GGC ACT TTA GGT ACC AGG ATT ATC TAC TTA AAT ACT TAG ATA TGG TAT ATT AAC TTA AGG A-3') were generated. For *myd88*miRNA

target sites 1 and 2, the primers are as described previously [3]. The above pairs of primary stem loop containing PCR products were amplified using primers mir6.15' *Not*I/*Fse*I/*Bgl*II (5'-TCG GGC GGC CGC ATT TGG CCG GCC AAA GAT CTT TTA AAG TCC ACA ACT CAT CAA GGA AAA TGA AAG TCA AAG TTG GCA GCT TAC TTA AAC TTA ATC ACA GCC TTT AAT GT- 3') and mir6.1 3' *EcoR*I/*Asc*I/*BamH*I (5' - TGA AGA ATT CAT TGG CGC GCC TTT GGA TCC AAA ACG GCA TGG TTA TTC GTG TGC CAA AAA AAAAAAAAA TTA AAT AAT GAT GTT AGG CAC TTT AGG TAC-3'). These primers add mir6.1 flanking sequences that are thought to promote miRNA processing, and several restriction sites. PCR products were purified with Qiagen (Valencia, CA) PCR purification columns, and then digested with appropriate restriction enzymes. For *dah*-1,*o-fut1*-1 and myd88-1 these were *Not*I and *BamH*I, and *Bgl*II and *Asc*I for *dah*-2, *o-fut1*-2 and myd88-2. Two digestion products, each containing a single miRNA targeting the same gene, were ligated together and cloned into the pAc5.1/V5-HisB shuttle vector (Invitrogen), producing a pair of miRNAs designed to silence *dah*, *o-fut1* or *myd88*. The copy number of miRNAs in pAc5.1 was doubled by digesting the above plasmids with either *Not*I and *BamH* or *Bgl*II and *Asc*I. The products generated each contain 2 miRNAs. These were ligated together into the pAc5.1 shuttle vector cut with Not1 and Asc1, producing a final vector containing two copies each of 2 miRNAs designed to silence each gene. This procedure was repeated once more to produce polycistronic 8-mer miRNA toxins specific to *dah*, *o-fut1* or *myd88*. These miRNAs were then digested out of the pAc5.1 shuttle vector using *Fse*I/*Asc*I and ligated into UD^{MEL}-BicC-Gyp-Bnk-attb, previously digested with*Fse*I and *Asc*I, generating UD^{MEL}-BicC-o-fut1-Gyp-Bnk-attb, UD^{MEL}-BicC-dah-Gyp-Bnk-attb and UD^{MEL}-BicC-myd88-Gyp-Bnk-attb.

Construction of the Antidotes and Final UD^{MEL} Constructs

The O-fut1 coding region was amplified from a cDNA library using primers o-fut1-anti-f (5'-ACA TTC GTA CTT CAA CCG TCA ACG AAT TCG GCT TAA TTA AAT GCA GTG GCT CAA AAT GAA GC-3') and o-fut1-anti-r (5'-AGA AGT AAG GTT CCT TCA CAA AGA TCC TGC GGC CGC TTA CAG CTC CTC GTG CAC GTT TGT-3'), producing a 1284bp PCR product bearing a unique *Pac*I restriction site at the 5' end of the o-fut CDS and a unique *Not* I site at the 3' end. This PCR product overlaps the 3' end of the *bnk* promoter and the 5' end of the SV40-3'UTR. The SV40-3'UTR was PCR amplified using primers Sv40-f (5'-TAC AAA CGT GCA CGA GGA GCT GTA AGC GGC CGC AGG ATC TTT GTG AAG GAA CCT TAC TTC-3') and Sv40-R (5'-TAC AAA CGT GCA CGA GGA GCT GTA AGC GGC CGC AGG ATC TTT GTG AAG GAA CCT TAC TTC-3'), from p(UWG) [4], producing a 764bp product that overlaps the 3'end of the *o-fut1* CDS and the attB backbone. These 2 PCR products were ligated together using one-step recombination, as above, into a *Pac*I-digested UD^{MEL}-BicC-dah-Gyp-Bnk-attb, producing final UD^{MEL}-dah^{T}-o-$fut1^{A}$. Note that because the *o-fut1* transcript lacks the 3' UTR present in the endogenous *o-fut1* transcript, it is not silenced by mir6.1-*o-fut-1* or mir6.1-*o-fut1*-2, which target the native *o-fut1* 3' UTR. The *dah* coding region was amplified from a cDNA library using primers *dah*-anti-f-1 (5'-TCA ACA GCA CAT TCG TAC TTC AAC CGT CAA CGA ATT CGG CAT GCT GAG ATC GTC GGT GCC CGT-3') and *dah* -anti-r-1 (5'- GTT GCC CTG TCC AAC TTG TAA TTG GCG TCT TGA TTG AAA TGG CCT AGT TTC TCG CAG GC-3'), producing a 455bp PCR product, and *dah* -anti-f-2 (5'- GCC TGC GAG AAA CTA GGC CAT TTC AAT CAA GAC GCC AAT TAC AAG TTG GAC AGG GCA AC-3') and *dah* -anti-r-2 (5'- AGA CCG TGA CCT ACA TCG TCG ACA CTA GTG GAT CTC TAG CGG CCG CTC ACG TGC TGA TGC GC -3'), producing a 1638bp PCR product bearing a unique *Not*I site 3' end of the *dah* rescue. The dah 3' UTR was amplified from genomic

DNA with primers *dah*-UTR-f-1 (5'-CAT CAG CAC GTG AGC GGC CGC AAC GGT ACC GGA TC-3') and *dah*-UTR-R-1 (5'-GAG ACC GTG ACC TAC ATC GTC GAC ACT AGT GGA TCT CTA GAG CAT TGG AAA TCT ACA AAG TTG AT-3'), producing a 374bp PCR product that overlaps the *dah* CDS and attP backbone. These three PCR fragments were ligated together into *Pac*I-digested UD^{MEL}-BicC-o-fut1-Gyp-Bnk-attb and UD^{MEL} -BicC-myd88-Gyp-Bnk-attb, producing the final UD^{MEL}-o-fut1^{T}-DAH^{A} and UD^{MEL}-myd88^{T}-DAH^{A} constructs Note that because this *dah* transcript lacks a 5' UTR present in the endogenous *dah* transcript, it is not silenced by mir6.1-*dah*-1 which targets the *dah* 5' UTR. mir6.1-*dah*-2 targets exon 2 of the CDS in *dah.* To render the antidote insensitive to this miRNA, we recoded the nucleotide sequence in *dah* in such that it codes for the same amino acid sequence, but is not targeted by the miRNA. The *myd88* coding region was PCR amplified from genomic DNA using primers myd88-f-1 (5'-ATT CGT ACT TCA ACC GTC AAC GAA TTC GGC TTA ATT AAA TGC GCC CTC GAT TT GTA TGC C-3') and myd88-R-1 (5'-GTA AGG TTC CTT CAC AAA GAT CCT CTA GAC CGC GGC CGC TCA GCC CGG CGT CTG CAG CTT GC-3'), producing a 1690bp PCR product bearing a unique *Pac*I site on the 5' end of the *myd88* CDS and a unique *Not*I site on the 3' end of the *myd88* CDS. The *myd88* CDS PCR product overlaps the 3' end of the *bnk* promoter and the 5' end of the SV40 3' UTR. The SV40 3' was PCR amplified using Sv40-M-F (5'-CAG CAG CAA GCT GCA GAC GCC GGG CTG AGC GGC CGC GGT CTA GAG GAT CTT TGT GAA GGA ACC-3') and SV40-M-R (5'-CCT ACA TCG TCG ACA CTA GTG GAT CTC TAG AGG ATC CAG ACA TGA TAA GAT ACA TTG ATG-3'), producing a 775bp PCR product that overlaps the 3' end of the *myd88* rescue and the attP plasmid. These 2 PCR products were ligated together into *Pac*I digested UD^{MEL}-BicC-dah-Gyp-Bnk-attb, using one step recombination, producing the final UD^{MEL}-dah^{T}-myd88^{A} construct. Note that because this *myd88* transcript lacks the 5' UTR present in the endogenous myd88 transcript, it is not silenced by myd88-1 and myd88-2 miRNA toxins, which are designed to target the *myd88* 5'UTR.

Transgenesis and UD^{MEL} System Generation

Germline transformation of *D. melanogaster* was performed by Rainbow Transgenic Flies, Inc (www.rainbowgene.com, Camarillo, CA). The transgenic lines were generated by site specific PhiC31 integration into attp docking sites on either the 2nd chromosome at cytological location 31B1 (Bloomington fly stock # 9724, PBac(y[+]-attP-3B}VK00003a) (constructs UD^{MEL}-o-fut1^{T}-DAH^{A} and UD^{MEL}-myd88^{T}-DAH^{A}), or the 3rd chromosome at cytological location 86E18 (Bloomington fly stock # 24486, M(vas-int.Dm)ZH-2A, M(3xP3-RFP.attP')ZH-86Fa 2A3) (constructs UD^{MEL}-dah^{T}-myd88^{A}, UD^{MEL}-o-fut1^{T}-DAH^{A}, UD^{MEL}-myd88^{T}-DAH^{A} and UD^{MEL}-dah^{T}-o-fut1^{A}). Each of these lines, UD^{MEL}-dah^{T}-myd88^{A} at attb site 86fa, UD^{MEL}-o-fut1^{T}-DAH^{A} at attb site 9724, UD^{MEL}-o-fut-1^{T}-DAH^{A} at attb site 86FA, UD^{MEL}-myd88^{T}-DAH^{A} at attb site 9724, UD^{MEL}-myd88^{T}-DAH^{A} at attb site 86FA and UD^{MEL}-dah^{T}-o-fut1^{A} at attb site 86fa, were maintained (25+ generations) by out crossing the transgenic males to WT w^{1118} (+/+) virgin females.

To generate the single and two locus UD^{MEL} configurations, virgin females carrying one toxin-antidote combination were crossed with males carrying the complementary toxin-antidote combination to produce stable lines. In the case of two-locus UD^{MEL}, homozygosity for each chromosome was achieved by carrying out single pair crosses between animals that must be at least transheterozygous for both constructs, for multiple generations. Homozygosity was confirmed using single fly PCR involving primers designed to amplify sequences to either side of the attp insertion site. For the 2nd chromosome 9724 insertion site, primers 9724-F (5'- ACA

TTT ATA TTT TCG TTT GCG ACC GA-3') and 9724-R (5'-CCC AAA AGA CTT GGC TCG GAT GCA CTG A-3') were used. For the third chromosome 86FA insertion site we out-crossed individual males to wild-type w^{1118} (+/+) virgin females and determined whether all offspring carried the 3x3p RFP transgene associated with this insertion site. We also used PCR primers 86fa-F (5'-ATC TGT AGG CTA GCG TAT TTA G-3') and 86fa-R (5'-GAT CCA AAA GAA TAC ATA GCA ATG CGA-3') to carry out PCR from single flies. These primers produce an 88bp PCR product from wild-type DNA.

Embryo Viability Determination

For embryo viability counts, 2-4 day old adult virgin females were allowed to mate with males of the relevant genotypes for 2-3 days in egg collection chambers, supplemented with yeast paste. On the following day, a 3 hr egg collection was carried out, after first having cleared old eggs from the females through a pre-collection period on a separate plate for three hrs. Embryos were isolated into groups of 100 and kept on an agar surface at 25^{o}C for 48-72 hrs. The % survival was then determined by counting the number of unhatched embryos. One group of 100 embryos per cross were scored in each experiment, and each experiment was carried out three times. The results presented are averages from these three experiments. Embryo survival was normalized with respect to the % survival observed in parallel experiments carried out with the w^{1118} strain used for line maintenance. For the embryo counts for crosses between heterozygous females and either heterozygous males or WT males we observed 100% maternal-effect killing with no surviving embryos with a sample size of greater then 10,000 embryos for each construct (Figure 5A). For Adult fly counts (figure 5A-bottom), individual heterozygous males (M/+) were mated to multiple WT (+/+) virgin females (N=3-5) for each of the six constructs tested and this experiment was executed three times. 100% of the progeny from these crosses (between 417-742 progeny) were counted, and the results of the three experiments were averaged together.

Population Cage Experiments

All fly experiments were carried out at 25^{o}C, ambient humidity in 250 ml bottles containing Lewis medium supplemented with live, dry yeast. Drive experiments were tested against a w^{1118} white eyed wild-type strain, and fly rearing was carried out in a light-tight, dark chamber, to eliminate any fitness benefit associated with expression of the w+ transgene. G_0 flies were collected as virgins (males and females were collected concurrently) and then aged for 3 days before crosses were set. Flies were anesthetized concurrently and introduced to the bottles in a single group, so that no bias was introduced from some flies waking before others. The largest practical populations (N=150) were used to ensure good mixing, as this was found to be important in the first generation (data not shown). Flies were then allowed to mate and lay eggs for 5 days, at which point adults were cleared from the bottles. 14 days post-mixing, flies were anesthetized, and divided into two groups. One group was immediately introduced into the new bottle while the other was counted to determine genotype/phenotype frequencies (wild-type and transgene-bearing were the only classes scored). For experiments in which males were released in both the first and second generations, males were again collected young and aged for three days before introduction into bottles. Progeny males and females from the first generation cross were collected as virgins and young males, and kept isolated for three days. Young transheterozygous males from the UD^{MEL} stock were collected and similarly aged at the

same time. Subsequently, transheterozygous males and a cohort of flies from the first generation were introduced together into a new population cage, maintaining the 60% introduction frequency present in the first generation. Subsequent generations were scored as above.

Population Genetic Modelling and Fitness Cost Estimation

We use a simple difference equation framework to model the spread of single and two-locus UD^{MEL} through a randomly mating population. In doing so, we assume discrete generations and infinite population size. To investigate the confinement properties of single- and two-locus UD^{MEL} we follow the framework of Marshall and Hay [5]. We consider two scenarios, one in which migration occurs before mating (mi-ma) and a second in which mating occurs before migration (ma-mi). Details are provided below.

Fitness Cost Estimation (Single-Locus)

In order to estimate the fitness costs of the UD^{MEL} constructs for both single and two-locus systems, we assume discrete generations, random mating and infinite population size. There are two transgenic constructs, each consisting of a maternal toxin and a zygotic antidote to the toxin on the complementary construct. For the single-locus case, both of these are present at the same locus. We denote the transgenic alleles by "*A*" and "*B*" and the null allele at this locus by "*a*." There are six possible genotypes – *AB*, *AA*, *BB*, *Aa*, *Ba* and *aa* – and we denote the proportion of the k th generation that are fruit flies having these genotypes by p_k^{AB}, p_k^{AA}, p_k^{BB}, p_k^{Aa}, p_k^{Ba} and p_k^{aa}, respectively.

By considering all possible mating pairs, taking into account that offspring are unviable if they do not have the antidote for any maternal toxin possessed by their mother, the genotypes of embryos in the next generation are described by the ratio $\hat{p}_{k+1}^{AB} : \hat{p}_{k+1}^{AA} : \hat{p}_{k+1}^{BB} : \hat{p}_{k+1}^{Aa} : \hat{p}_{k+1}^{Ba} : \hat{p}_{k+1}^{aa}$. The equations for these ratios follow from the schematic in supplementary Figure 1A; however, given the large number of possible mating pairs, it is not feasible to show them here. The simulation code that utilizes them is available from the authors upon request. The genotype frequencies in the next generation can then be calculated by accounting for the genotype-specific fitness costs, i.e.,

$$(p_{k+1}^{AB}, p_{k+1}^{AA}, p_{k+1}^{BB}, p_{k+1}^{Aa}, p_{k+1}^{Ba}, p_{k+1}^{aa}) = (\hat{p}_{k+1}^{AB}, \hat{p}_{k+1}^{AA}, \hat{p}_{k+1}^{BB}, \hat{p}_{k+1}^{Aa}, \hat{p}_{k+1}^{Ba}, \hat{p}_{k+1}^{aa}) \times (1-s_{AB}, 1-s_{AA}, 1-s_{BB}, 1-s_{Aa}, 1-s_{Ba}, 1)/W_{k+1}$$

Here, s represents the fitness cost associated with each genotype, with the genotypes being listed as subscripts. W_{k+1} is a normalizing term given by,

$$W_{k+1} = \hat{p}_{k+1}^{AB}(1-s_{AB}) + \hat{p}_{k+1}^{AA}(1-s_{AA}) + \hat{p}_{k+1}^{BB}(1-s_{BB}) + \hat{p}_{k+1}^{Aa}(1-s_{Aa}) + \hat{p}_{k+1}^{Ba}(1-s_{Ba}) + \hat{p}_{k+1}^{aa}$$

We investigated a number of different fitness cost models and tested them by seeing which provides the best fit to the data. The simplest model is one in which all transgenics have the same fitness cost (i.e. $s_{AB} = s_{AA} = s_{BB} = s_{Aa} = s_{Ba} = s$). An alternative model assumes that

fitness costs are additive (i.e. $s_{AB} = s_{AA} = s_{BB} = s$ and $s_{Aa} = s_{Ba} = s/2$). In both cases, we also looked into models where fitness cost varies linearly with transgenic allele frequency,

$$s = (s_{tr} - s_{wt})(p_k^{AB} + p_k^{AA} + p_k^{BB} + 0.5(p_k^{Aa} + p_k^{Ba})) + s_{wt} ,$$

and where it varies linearly with wild-type frequency,

$$s = (s_{wt} - s_{tr})\, p_k^{aa} + s_{tr} .$$

Here, s_{wt} represents the fitness cost of a transgenic homozygote in a (nearly) fully wild-type population, and s_{tr} represents the fitness cost of a transgenic homozygote in a (nearly) fully transgenic population.

These models were compared according to their Akaike Information Criterion (AIC) values. The likelihood of the data was calculated by assuming a binomial distribution of WT and red-eyed individuals, and by using the model predictions to generate expected proportions for each fitness cost and model, i.e. by calculating the log-likelihood,

$$\log L(\theta) = \sum_i \sum_k \log \binom{R_{i,k} + WT_{i,k}}{R_{i,k}} + WT_{i,k} \log(p_k^{aa}(\theta)) + R_{i,k} \log(p_k^{AB}(\theta) + p_k^{AA}(\theta) + p_k^{BB}(\theta) + p_k^{Aa}(\theta) + p_k^{Ba}(\theta)) .$$

Here, $R_{i,k}$ and $WT_{i,k}$ are the number of red-eyed and WT individuals at generation k in experiment i, and the expected genotype frequencies depend on the fitness cost model and parameters, θ. The best estimate of the model parameters are those having the highest log-likelihood, and the best fitting model is that having the lowest AIC value,

$$AIC = 2n - 2\log L(\theta) ,$$

where n represents the number of model parameters. The AIC values and parameter estimates for each of the fitness cost models are shown in the table below. This table shows that the best-fitting model is one in which fitness costs are dominant (i.e. transgenic homozygotes and heterozygotes have the same fitness cost) and the magnitude of the fitness cost depends on the frequency of wild-type (i.e. *aa*) individuals in the population.

Fitness Cost Model	**AIC**	s_{wt}	s_{tr}
Constant, dominant fitness costs	1575	0.167	0.167
Constant, additive fitness costs	1448	0.201	0.201
Transgenic allele frequency-dependent, dominant fitness costs	1029	-0.213	0.389
Transgenic allele frequency-dependent, additive fitness costs	1102	-0.301	0.470
Wild-type frequency-dependent, dominant fitness costs	1027	-0.242	0.276
Wild-type frequency-dependent, additive fitness costs	1092	-0.389	0.300

Fitness Cost Estimation (Two-Locus)

For the two-locus case, the two transgenic constructs are present at different loci. At the first locus, we denote the transgenic allele by "*A*" and the null allele by "*a*," and at the second locus, we denote the transgenic allele by "*B*" and the null allele by "*b*." This time, there are nine possible genotypes – *AABB*, *AABb*, *AAbb*, *AaBB*,*AaBb*, *Aabb*,*aaBB*, *aaBb* and *aabb* – and we denote the proportion of the k th generation that are fruit flies having these genotypes by p_k^{AABB}, p_k^{AABb}, p_k^{AAbb}, p_k^{AaBB}, p_k^{AaBb}, p_k^{Aabb}, p_k^{aaBB}, p_k^{aaBb} and p_k^{aabb}, respectively.

By considering all possible mating pairs, taking into account that offspring are unviable if they do not have the antidote for any maternal toxin possessed by their mother, the genotypes of embryos in the next generation are described by the ratio $\hat{p}_{k+1}^{AABB} : \hat{p}_{k+1}^{AABb} : \hat{p}_{k+1}^{AAbb} : \hat{p}_{k+1}^{AaBB} : \hat{p}_{k+1}^{AaBb} : \hat{p}_{k+1}^{Aabb} : \hat{p}_{k+1}^{aaBB} : \hat{p}_{k+1}^{aaBb} : \hat{p}_{k+1}^{aabb}$. The equations for these ratios follow from the schematic in supplementary Figure 1B; however, given the large number of possible mating pairs, it is not feasible to show them here, and the simulation code that utilizes them is available from the authors upon request. The genotype frequencies in the next generation can then be calculated by accounting for the genotype-specific fitness costs, i.e.,

$$(p_{k+1}^{AABB}, p_{k+1}^{AABb}, p_{k+1}^{AAbb}, p_{k+1}^{AaBB}, p_{k+1}^{AaBb}, p_{k+1}^{Aabb}, p_{k+1}^{aaBB}, p_{k+1}^{aaBb}, p_{k+1}^{aabb}) =$$
$$(\hat{p}_{k+1}^{AABB}, \hat{p}_{k+1}^{AABb}, \hat{p}_{k+1}^{AAbb}, \hat{p}_{k+1}^{AaBB}, \hat{p}_{k+1}^{AaBb}, \hat{p}_{k+1}^{Aabb}, \hat{p}_{k+1}^{aaBB}, \hat{p}_{k+1}^{aaBb}, \hat{p}_{k+1}^{aabb})$$
$$\times (1-s_{AABB}, 1-s_{AABb}, 1-s_{AAbb}, 1-s_{AaBB}, 1-s_{AaBb}, 1-s_{Aabb}, 1-s_{aaBB}, 1-s_{aaBb}, 1)/W_{k+1}$$

Here, s represents the fitness cost associated with each genotype, with the genotypes being listed as subscripts. W_{k+1} is a normalizing term given by,

$$W_{k+1} = \hat{p}_{k+1}^{AABB}(1-s_{AABB}) + \hat{p}_{k+1}^{AABb}(1-s_{AABb}) + \hat{p}_{k+1}^{AAbb}(1-s_{AAbb}) + \hat{p}_{k+1}^{AaBB}(1-s_{AaBB})$$
$$+ \hat{p}_{k+1}^{AaBb}(1-s_{AaBb}) + \hat{p}_{k+1}^{Aabb}(1-s_{Aabb}) + \hat{p}_{k+1}^{aaBB}(1-s_{aaBB}) + \hat{p}_{k+1}^{aaBb}(1-s_{aaBb}) + \hat{p}_{k+1}^{aabb}.$$

As for the single-locus case, we investigated a number of different fitness cost models and tested them by seeing which provides the best fit to the data. The simplest model is one in which all transgenic individuals have the same fitness cost (i.e. $s_{AABB} = s_{AABb} = s_{AAbb} = s_{AaBB} = s_{AaBb} = s_{Aabb} = s_{aaBB} = s_{aaBb} = s$). A second model assumes that each transgenic allele has the same fitness cost (i.e. $s_{AABB} = s_{AABb} = s_{AaBB} = s_{AaBb} = s + s = 2s$ and $s_{AAbb} = s_{Aabb} = s_{aaBB} = s_{aaBb} = s$); and a third model assumes that the fitness costs of each allele are additive (i.e. $s_{AABB} = 2s$, $s_{AABb} = s_{AaBB} = 3s/2$, $s_{AaBb} = s_{AAbb} = s_{aaBB} = s$ and $s_{Aabb} = s_{aaBb} = s/2$). In each case, we also looked into models where fitness cost varies linearly with wild-type frequency,

$$s = (s_{wt} - s_{tr})\, p_k^{aabb} + s_{tr},$$

and where it varies with transgenic allele frequency,

$$s = (s_{tr} - s_{wt}) \frac{(p_k^{AA} + 0.5 p_k^{Aa}) + (p_k^{BB} + 0.5 p_k^{Bb})}{2} + s_{wt},$$

where

$$p_k^{AA} = p_k^{AABB} + p_k^{AABb} + p_k^{AAbb},$$
$$p_k^{Aa} = p_k^{AaBB} + p_k^{AaBb} + p_k^{Aabb},$$
$$p_k^{BB} = p_k^{AABB} + p_k^{AaBB} + p_k^{aaBB},$$
$$p_k^{Bb} = p_k^{AABb} + p_k^{AaBb} + p_k^{aaBb}.$$

Here, s_{wt} represents the fitness cost of a transgenic homozygote at both loci in a (nearly) fully wild-type population, and s_{tr} represents the fitness cost of a transgenic homozygote at both loci in an (almost) fully transgenic population.

These models were compared according to their Akaike Information Criterion (AIC) values. As for the single-locus case, the likelihood of the data was calculated by assuming a binomial distribution of WT and red-eyed individuals, and by using the model predictions to generate expected proportions for each fitness cost and model, i.e. by calculating the log-likelihood,

$$\log L(\theta) = \sum_i \sum_k \log\binom{R_{i,k} + WT_{i,k}}{R_{i,k}} + WT_{i,k} \log(p_k^{aabb}(\theta)) + R_{i,k} \log(p_k^{AABB}(\theta) + p_k^{AABb}(\theta) + p_k^{AAbb}(\theta) + p_k^{AaBB}(\theta) + p_k^{AaBb}(\theta) + p_k^{Aabb}(\theta) + p_k^{aaBB}(\theta) + p_k^{aaBb}(\theta))$$

Here, $R_{i,k}$ and $WT_{i,k}$ are the number of red-eyed and WT individuals at generation k in experiment i, and the expected genotype frequencies depend on the fitness cost model and parameters, θ. The best estimate of the model parameters are those having the highest log-likelihood, and the best fitting model is that having the lowest AIC value. The AIC values and parameter estimates for each of the fitness cost models are shown in the table below. Interestingly, the best-fitting model for the two-locus system is the same as for the single-locus system. In this model, fitness costs are dominant, such that all transgenic individuals suffer the same fitness cost, and the magnitude of the fitness cost depends on the frequency of wild-type (i.e. *aabb*) individuals in the population.

Fitness Cost Model	**AIC**	s_{wt}	s_{tr}
Constant, dominant fitness costs	4114	0.202	0.202
Constant, dominant allelic fitness costs	3847	0.104	0.104
Constant, additive allelic fitness costs	3673	0.168	0.168
Transgenic allele frequency-dependent, dominant fitness costs	2663	-0.088	0.677
Transgenic allele frequency-dependent, dominant allelic fitness costs	2680	-0.047	0.345
Transgenic allele frequency-dependent, additive allelic fitness costs	2723	-0.083	0.582

To summarize, the best-fit model (Table S1) for single-locus UD^{MEL} is one in which fitness costs are dominant (i.e. transgenic homozygotes and heterozygotes have the same fitness cost) and the magnitude of the fitness cost depends on the frequency of wild-type (i.e. *aa*) individuals in the population. We calculated 95% confidence intervals for the parameters used in this model using a Markov Chain Monte Carlo sampling procedure. Interestingly, results suggest that transgenics experience a fitness benefit in a (nearly) fully wild-type population of 24.2% (95% confidence interval: 20.5%-27.6%) and a transgenic fitness cost in a (nearly) fully transgenic population of 27.6% (95% confidence interval: 26.5%-28.6%). The best-fit model for the two-locus system has similar overall characteristics (Table S2), though the calculated fitness costs are different. The basis for these effects remains to be explored.

Migration Thresholds and Confinement

We first investigate the confinement properties of single-locus UD^{MEL} following the framework of Marshall and Hay [5], in which migration occurs before mating. We consider a two-population model in which the mating pool in both populations is made up of individuals from populations 1 and 2. The proportion of the k th generation that are individuals having the genotypes *AB*, *AA*, *BB*, *Aa*, *Ba* and *aa* in population 1 are denoted by $p_{1,k}^{AB}$, $p_{1,k}^{AA}$, $p_{1,k}^{BB}$, $p_{1,k}^{Aa}$, $p_{1,k}^{Ba}$ and $p_{1,k}^{aa}$, respectively, with corresponding symbols applying to population 2. For a migration rate of μ in both directions, we make the following substitutions into the equations described earlier for UD^{MEL} dynamics in population 1,

$$
\begin{aligned}
&(p_{1,k}^{AB}, p_{1,k}^{AB}, p_{1,k}^{AB}, p_{1,k}^{AB}, p_{1,k}^{AB}, p_{1,k}^{AB}, p_{1,k}^{AB}) \\
&\leftarrow (p_{1,k}^{AB}, p_{1,k}^{AB}, p_{1,k}^{AB}, p_{1,k}^{AB}, p_{1,k}^{AB}, p_{1,k}^{AB}, p_{1,k}^{AB})(1-\mu) \\
&+ (p_{2,k}^{AB}, p_{2,k}^{AB}, p_{2,k}^{AB}, p_{2,k}^{AB}, p_{2,k}^{AB}, p_{2,k}^{AB}, p_{2,k}^{AB})\mu
\end{aligned}
$$

.

Iterating these equations confirms that single-locus UD^{MEL} is confineable to a partially-isolated population and that, rather than spreading into neighboring populations at high migrations, the system is actually eliminated from both populations, similar to the case for single-locus engineered UD. We applied the same modeling framework to determine the migration and confinement properties of the two-locus UD^{MEL} systems.

In the case where mating occurs before migration the analysis proceeds similarly. We consider a two-population model in which mating occurs in populations 1 and 2 separately, following which each population exchanges a fraction, μ, of its population with the other. The proportion of the k th generation that are individuals having the genotypes *AB*, *AA*, *BB*, *Aa*, *Ba* and *aa* in population 1 are denoted by $p_{1,k}^{AB}$, $p_{1,k}^{AA}$, $p_{1,k}^{BB}$, $p_{1,k}^{Aa}$, $p_{1,k}^{Ba}$ and $p_{1,k}^{aa}$, respectively, with corresponding symbols applying to population 2. Iterating these equations confirms that single-locus UD^{MEL} is confineable to a partially-isolated population and that, rather than spreading into neighboring populations at high migration rates, the system is actually eliminated from both populations, similar to the case for single-locus engineered UD. As above, this same modeling framework was applied to determine the migration and confinement properties of the two-locus UD^{MEL} systems.

Population Suppression and Reversion

We investigate the possibility of suppressing and reverting UD^{MEL}-replaced populations using a stochastic population model analogous to the population frequency model previously described. A stochastic model was chosen because it can accommodate the small populations resulting from population suppression. Density-dependence is an important consideration because, at low population sizes, larval competition is reduced and a single female can produce more offspring than survive to adulthood. We adapt a general mosquito population biology model [6,7], modified to run in discrete time with units of one generation. In this model, the number of adult females in the population at generation k is given by,

$$N_{f,k} = \sigma_k f N_{f,k-1} \theta_0 \theta_L F(L_k) \theta_P .$$

Here, σ_k represents the proportion of embryos that are viable after the toxin-antidote effects of the UD^{MEL} constructs are taken into account, f represents the average number of female eggs laid by an adult female in her lifetime, θ_0 represents the proportion of eggs that survive the egg life stage, θ_L represents the proportion of larvae that survive the larval life stage in the absence of density-dependence, $F(L_k)$ represents the proportion of larvae that survive the effects of density-dependent mortality (density-dependence is assumed to act at the larval stage), and θ_P represents the proportion of pupae that survive the pupal life stage. The density-dependent function is given by,

$$F(L_k) = \frac{\alpha}{\alpha + L_k} .$$

Here, L_k represents the number of larvae of both sexes at generation k. This is given by,

$$L_k = 2\sigma_k f N_{f,k-1} \theta_0 .$$

Finally, α is a term that specifies the strength of density-dependence and, for a population with an adult carrying capacity of K, is given by,

$$\alpha = \frac{fK\theta_0}{f\theta_0\theta_L\theta_P - 1} .$$

The total population size is given by $N_k = 2N_{f,k}$. Stochastic effects are incorporated by sampling the number of individuals having each genotype from a Poisson distribution with a mean equal to the expected number of individuals having this genotype. For the single-locus system, suppression of the UD^{MEL}-replaced population can then be achieved by consecutive releases of 10,000 wild-type males, and population reversion can be achieved by releasing small numbers of wild-type males and females into the suppressed populations so that wild-type individuals exceed the required threshold for re-colonization. For the two-locus system, population suppression can't be achieved through the release of wild-type males; however, three consecutive releases of 5,000 wild-type males and 5,000 wild-type females are sufficient for wild-types to exceed the required threshold for reversion to a fully wild-type population.

Parameters are described in the following table and code is available from the authors upon request.

Symbol	Parameter	Value	Reference
f	Average number of female eggs a female adult lays in her lifetime	130	[8]
θ_0	Probability of surviving the egg life stage	0.831	[9]
θ_L	Probability of surviving the larval life stage (no density dependence)	0.076	[9]
θ_P	Probability of surviving the pupal life stage	0.831	[9]
K	Adult population carrying capacity	10,000	

Supplemental References

1. Fish, M.P., Groth, A.C., Calos, M.P., and Nusse, R. (2007). Creating transgenic Drosophila by microinjecting the site-specific phiC31 integrase mRNA and a transgene-containing donor plasmid. Nature protocols *2*, 2325-2331.
2. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature methods *6*, 343-345.
3. Chen, C.H., Huang, H., Ward, C.M., Su, J.T., Schaeffer, L.V., Guo, M., and Hay, B.A. (2007). A synthetic maternal-effect selfish genetic element drives population replacement in Drosophila. Science *316*, 597-600.
4. Akbari, O.S., Oliver, D., Eyer, K., and Pai, C.Y. (2009). An Entry/Gateway cloning system for general expression of genes with molecular tags in Drosophila melanogaster. BMC Cell Biol *10*, 8.
5. Marshall, J.M., and Hay, B.A. (2012). Confinement of gene drive systems to local populations: a comparative analysis. Journal of Theoretical Biology *294*, 153-171.
6. Hancock, P. A., and Godfray, H. C. J. (2007). Application of the lumped age-class technique to studying the dynamics of malaria-mosquito-human interactions. Malaria Journal *6*, 98.
7. Deredec, A., Godfray, H. C. J., and Burt, A. (2011). Requirements for effective malaria control with homing endonuclease genes. Proc. Natl. Acad. Sci. USA *108*, E874-E880.
8. Molineaux, L., and Gramiccia, G. (2007). The Garki Project: Research on the Epidemiology and COntrol of Malaria in the Sudan Savanna of West Africa. World Health Organization, Geneva.
9. Depinay, J. M. O., Mbogo, C. M., Killeen, G., Knols, B., Beier, J., Carlson, J., Dushoff, J., Billingsley, P., Mwambi, H., Githure, J., Toure, A. M., and McKenzie, F. E. (2004). A simulation model of African *Anopheles* ecology and population dynamics for the analysis of malaria transmission. Malaria Journal *3*, 29.

Supplemental Figure Legends

Figure S1. Parental and Progeny Genotypes and Survival for Two-Locus and Single-Locus UD^{MEL}
Punnet squares for single-locus (A), and two-locus (B) UD^{MEL}, indicating all possible parental and offspring genotypes. Progeny expected to perish are indicated in pink.

Figure S2. UD^{MEL} Single- and Two-Locus Systems Are Predicted to Show Threshold-Dependent Gene Drive in the Presence of a 10% Fitness Cost
The threshold frequency above which a UD^{MEL} drive system spreads into a population, and below which it is eliminated from the population when each element carries a 10% fitness cost, was calculated using a deterministic model and graphed as in Figure 2. Fitness costs are additive. Thus, for the two-locus system there is a 10% cost per locus when homozygous at that locus, and a 5% cost when heterozygous. Release thresholds are calculated for two single locus scenarios: a single, all-male release of transheterozygotes (A) and two all male releases of transheterozygotes in the first and second generation (B). For X-autosome two-locus UD^{MEL} (C) and autosome-autosome two-locus UD^{MEL} (D) single releases of doubly homozygous males are illustrated. Introduction frequencies/transgene frequencies represent the fraction of individuals in the total population carrying at least one UD^{MEL} construct. The symbol s refers to fitness cost.

Figure S3. Fate of the Wild-Type Allele during Population Replacement with Single- and Two-Locus UD^{MEL} Elements with a 10% Fitness Cost
Panels are as in Figure S2, with the Y axis indicating the wild-type allele frequency.

A

Female

Male	A/B	A/+	B/+	+/+
A/B	A/B A/A B/B	A/A A/+ A/B B/+	A/B A/+ B/B B/+	A/+ B/+
A/+	A/A A/B A/+ B/+	A/A +/+ A/+	A/B A/+ B/+ +/+	A/+ +/+
B/+	A/B B/B A/+ B/+	A/B B/+ A/+ +/+	B/B +/+ B/+	B/+ +/+
+/+	A/+ B/+	A/+ +/+	B/+ +/+	+/+

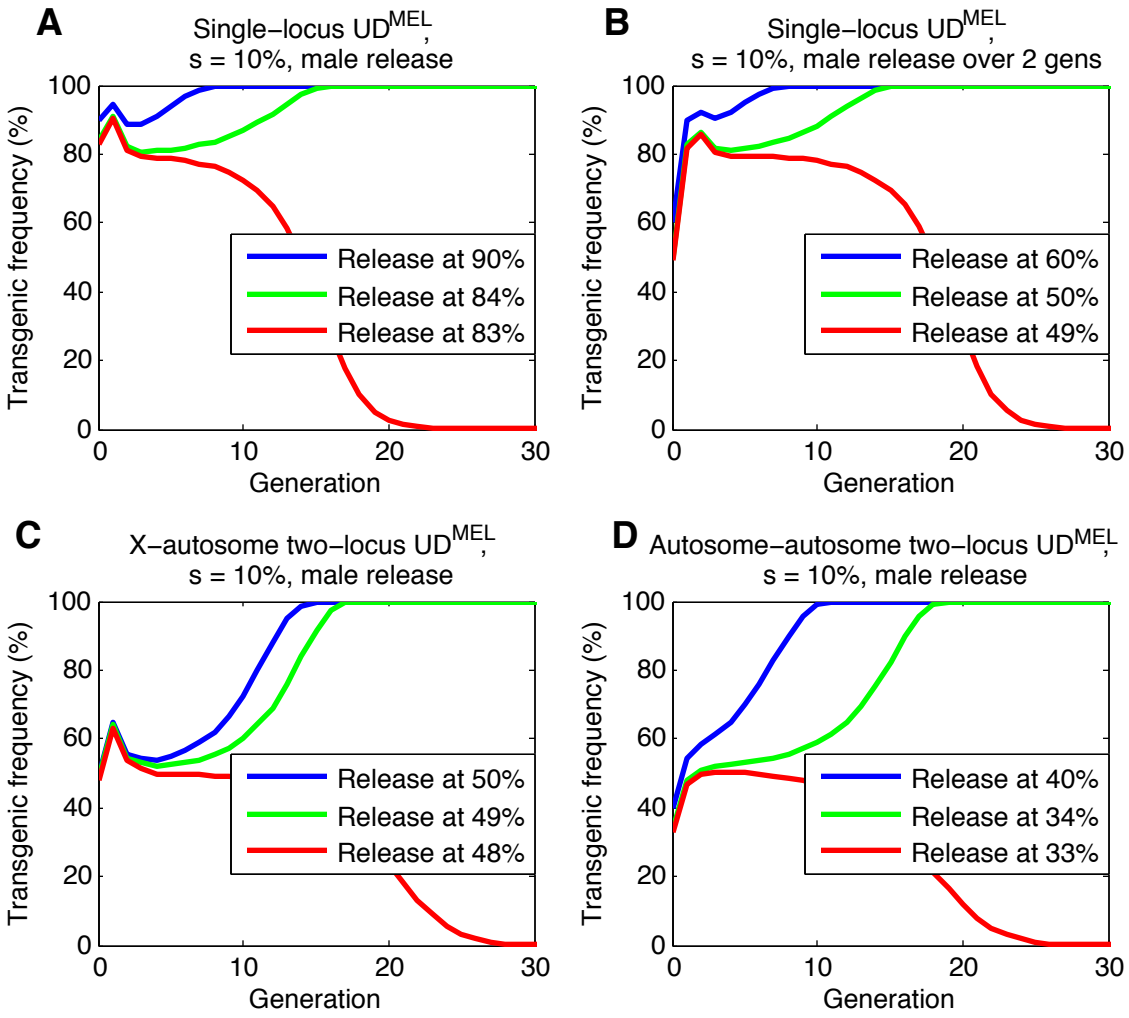
SIngle Locus UD^{MEL} 16 dihybrid punnet Square

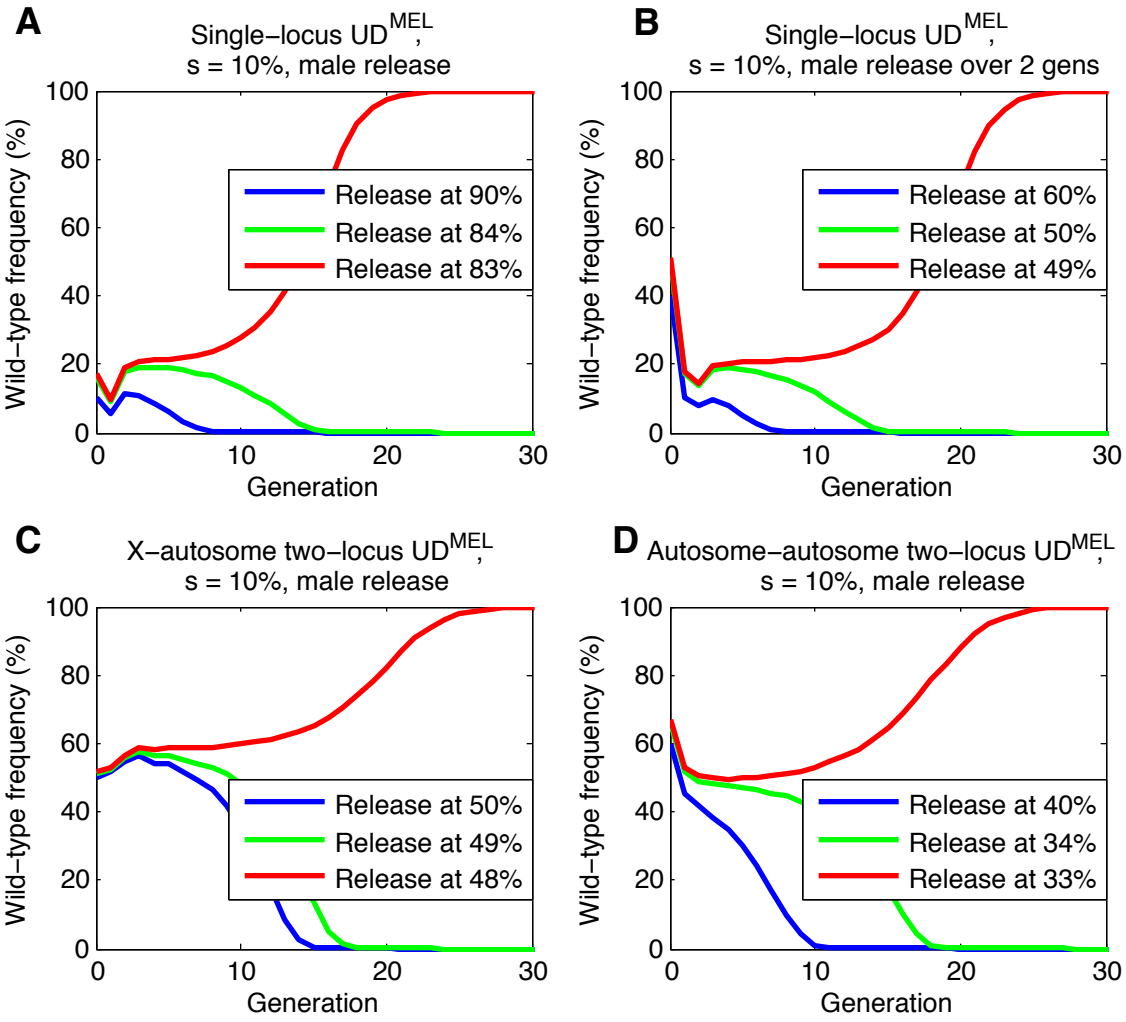
B

Female

Male	A/A ; B/B	A/+ ; B/B	+/+ ; B/B	A/A ; B/+	A/A ; +/+	+/+ ; B/+	A/+ ; B/+	A/+ ; +/+	+/+ ; +/+
A/A ; B/B	A/A ; B/B	A/A ; B/B A/+ ; B/B	A/+ ; B/B	A/A ; B/B A/A ; B/+	A/A ; B/+	A/+ ; B/B A/+ ; B/+	A/A ; B/B A/+ ; B/B A/A ; B/+ A/+ ; B/+	A/A ; B/+ A/+ ; B/+	A/+ ; B/+
A/+ ; B/B	A/A ; B/B A/+ ; B/B	A/+ ; B/B A/A ; B/B +/+ ; B/B	A/+ ; B/B	A/A ; B/B A/A ; B/+ A/+ ; B/B A/+ ; B/+	A/A ; B/+	A/+ ; B/B +/+ ; B/B A/+ ; B/+ +/+ ; B/+	A/A ; B/B A/A ; B/+ A/+ ; B/B A/+ ; B/+ +/+ ; B/B +/+ ; B/+	A/+ ; B/+ A/A ; B/+ +/+ ; B/+	A/+ ; B/+ +/+ ; B/+
+/+ ; B/B	A/+ ; B/B	A/+ ; B/B +/+ ; B/B	+/+ ; B/B	A/+ ; B/+ A/+ ; B/B	A/+ ; B/+	+/+ ; B/B +/+ ; B/+	A/+ ; B/B A/+ ; B/+ +/+ ; B/B +/+ ; B/+	A/+ ; B/+ +/+ ; B/+	+/+ ; B/+
A/A ; B/+	A/A ; B/B A/A ; B/+	A/A ; B/B A/+ ; B/B A/A ; B/+ A/+ ; B/+	A/+ ; B/B A/+ ; B/+ A/+ ; +/+	A/A ; B/B A/A ; B/+	A/A ; B/+ A/A ; +/+	A/+ ; B/+ A/+ ; B/B A/+ ; +/+	A/A ; B/B A/+ ; B/B A/A ; B/+ A/+ ; B/+ A/A ; +/+ A/+ ; +/+	A/A ; B/+ A/A ; +/+ A/+ ; B/+ A/+ ; +/+	A/+ ; B/+ A/+ ; +/+
A/A ; +/+	A/A ; B/+	A/A ; B/+ A/+ ; B/+	A/+ ; B/+	A/A ; B/+ A/A ; +/+	A/A ; +/+	A/+ ; B/+ A/+ ; +/+	A/A ; B/+ A/A ; +/+ A/+ ; B/+ A/+ ; +/+	A/A ; +/+ A/+ ; +/+	A/+ ; +/+
+/+ ; B/+	A/+ ; B/B A/+ ; B/+	A/+ ; B/B +/+ ; B/B A/+ ; B/+ +/+ ; B/+	+/+ ; B/B +/+ ; B/+ +/+ ; B/+	A/+ ; B/B A/+ ; +/+ A/+ ; B/+ +/+ ; B/+	A/+ ; B/+ A/+ ; +/+	+/+ ; B/B +/+ ; B/+ +/+ ; +/+	A/+ ; B/B +/+ ; B/B A/+ ; B/+ +/+ ; B/+ A/+ ; +/+ +/+ ; +/+	A/+ ; B/+ A/+ ; +/+ +/+ ; B/+ +/+ ; +/+	+/+ ; B/+ +/+ ; +/+
A/+ ; B/+	A/A ; B/B A/+ ; B/B A/A ; B/+ A/+ ; B/+	A/A ; B/B A/A ; B/+ A/+ ; B/B A/+ ; B/+ +/+ ; B/B +/+ ; B/+	A/+ ; B/B A/+ ; B/+ +/+ ; B/B +/+ ; B/+	A/A ; B/B A/+ ; B/B A/A ; B/+ A/+ ; B/+ A/A ; +/+ A/+ ; +/+	A/A ; B/+ A/+ ; B/+ A/A ; +/+ A/+ ; +/+	A/+ ; B/B +/+ ; B/B A/+ ; B/+ +/+ ; B/+ A/+ ; +/+ +/+ ; +/+	A/A ; B/B A/+ ; +/+ A/A ; B/+ A/+ ; +/+ A/+ ; B/B +/+ ; B/B A/+ ; B/+ +/+ ; B/+ A/+ ; +/+ +/+ ; +/+	A/A ; B/+ A/A ; +/+ A/+ ; B/+ A/+ ; +/+ +/+ ; B/+ +/+ ; +/+	A/+ ; B/+ A/+ ; +/+ +/+ ; B/+ +/+ ; +/+
A/+ ; +/+	A/A ; B/+ A/+ ; B/+	A/+ ; B/+ A/A ; B/+ +/+ ; B/+	A/+ ; B/+ +/+ ; B/+	A/A ; B/+ A/A ; +/+ A/+ ; B/+ A/+ ; +/+	A/A ; +/+ A/+ ; +/+	A/+ ; B/+ A/+ ; +/+ +/+ ; B/+ +/+ ; +/+	A/A ; B/+ A/A ; +/+ A/+ ; B/+ A/+ ; +/+ +/+ ; B/+ +/+ ; +/+	A/+ ; +/+ A/A ; +/+ +/+ ; +/+	A/+ ; +/+ +/+ ; +/+
+/+ ; +/+	A/+ ; B/+	A/+ ; B/+ +/+ ; B/+	+/+ ; B/+	A/+ ; B/+ A/+ ; +/+	A/+ ; +/+	+/+ ; B/+ +/+ ; +/+	A/+ ; B/+ A/+ ; +/+ +/+ ; B/+ +/+ ; +/+	A/+ ; +/+ +/+ ; +/+	+/+ ; +/+

2 Locus UD^{MEL} 81 dihybrid punnet Square


A
Single−locus UD^MEL,
s = 10%, male release
Transgenic frequency (%)
Generation
Release at 90%
Release at 84%
Release at 83%
B
Single−locus UD^MEL,
s = 10%, male release over 2 gens
Transgenic frequency (%)
Generation
Release at 60%
Release at 50%
Release at 49%
C
X−autosome two−locus UD^MEL,
s = 10%, male release
Transgenic frequency (%)
Generation
Release at 50%
Release at 49%
Release at 48%
D
Autosome−autosome two−locus UD^MEL,
s = 10%, male release
Transgenic frequency (%)
Generation
Release at 40%
Release at 34%
Release at 33%


A
Single−locus UD^MEL,
s = 10%, male release
Wild−type frequency (%)
Generation
Release at 90%
Release at 84%
Release at 83%
B
Single−locus UD^MEL,
s = 10%, male release over 2 gens
Wild−type frequency (%)
Generation
Release at 60%
Release at 50%
Release at 49%
C
X−autosome two−locus UD^MEL,
s = 10%, male release
Wild−type frequency (%)
Generation
Release at 50%
Release at 49%
Release at 48%
D
Autosome−autosome two−locus UD^MEL,
s = 10%, male release
Wild−type frequency (%)
Generation
Release at 40%
Release at 34%
Release at 33%