Supplementary Material

Engineered reciprocal chromosome translocations drive high threshold, reversible population replacement in Drosophila

Anna B. Buchman*[#], Tobin Ivy*, John M. Marshall[$], Omar S. Akbari*[#&], and Bruce A. Hay*[&]

*Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA. [$]School of Public Health, University of California Berkeley, Berkeley CA. [#] Division of Biological Sciences, University of California, San Diego, San Diego, CA.

[&]Authors for correspondence
haybruce@caltech.edu
oakbari@ucsd.edu

Supplementary Text S1:

Theoretical framework:

As described in the Materials & Methods, we apply a modified version of the model of Curtis and Robinson [1] to describe the spread of reciprocal translocations through a population. We denote the first chromosome with a translocated segment by "*T*" and the wild-type version of this chromosome by "*t*." Similarly, we denote the second chromosome with a translocated segment by "*R*" and the wild-type version of this chromosome by "*r*." As a two-locus system, there are nine possible genotypes; however, only individuals carrying the full chromosome complement are viable, which corresponds to the genotypes *TTRR*, *TtRr* and *ttrr*, the proportion of the *k*th generation of which are denoted by p_k^{TTRR}, p_k^{TtRr} and p_k^{ttrr}. The four haplotypes that determine the genotype frequencies in the next generation – *TR*, *tR*, *Tr* and *tr* – are described by the following frequencies:

$$f_k^{TR} = p_k^{TTRR}(1-s) + 0.25 p_k^{TtRr}(1-hs)$$
$$f_k^{tR} = f_k^{Tr} = 0.25 p_k^{TtRr}(1-hs)$$
$$f_k^{tr} = p_k^{ttrr} + 0.25 p_k^{TtRr}(1-hs)$$

Here, *s* denotes the reduced fecundity of *TTRR* individuals and *hs* denotes the reduced fecundity of *TtRr* individuals relative to wild-type individuals, where $h \in [0,1]$ and the maximum and minimum fitness costs are also bounded by 1 and 0. By considering all possible mating pairs, the genotype frequencies in the next generation are:

$$p_{k+1}^{TTRR} = (f_k^{TR})^2 / \sigma_k$$
$$p_{k+1}^{TtRr} = 2(f_k^{TR} f_k^{tr} + f_k^{tR} f_k^{Tr}) / \sigma_k$$
$$p_{k+1}^{ttrr} = (f_k^{tr})^2 / \sigma_k$$

where σ_k is a normalizing term given by,

$$\sigma_k = (f_k^{TR})^2 + 2(f_k^{TR} f_k^{tr} + f_k^{tR} f_k^{Tr}) + (f_k^{tr})^2$$

Fitness cost models:

In order to compare our laboratory drive data to theoretical predictions, we investigated a number of different fitness cost models. In all cases, the parents in the first generation were not subject to a fitness cost. The simplest model is one in which the fitness of each genotype stays constant over time. Another model considers fitness costs that depend on the population frequency of the genotype. For linear frequency-dependence, this is given by,

$$s = (s_0 - s_1) p_k^{ttrr} + s_1$$

Here, s_0 represents the fitness cost of a translocation homozygote in an almost fully wild-type population, and s_1 represents the fitness cost in an almost fully transgenic population. An alternative model is that fitness is time-dependent, as could be explained by introgression of introduced genotypes. For linear time-dependence, this is given by,

$$s = \left(\frac{s_1 - s_0}{t_f} \right) t + s_0$$

Here, s_0 represents the fitness cost in the second generation and s_1 represents the fitness cost in the final generation, denoted by t_f. For sigmoidal time-dependence, it is given by,

$$s = (s_0 - s_1) \left(1 - \frac{1}{1 + e^{-\alpha(1-\tau)}} \right) + s_1$$

Here, s_0 and s_1 are as before, τ denotes the time of intermediate fitness cost, and α denotes the speed of transition between the two fitness costs.

For exponential time-dependence, it is given by,

$$s = a 2^{-t/t_{1/2}} + (s_0 - a)$$

Here, s_0 represents the fitness cost in the second generation, s_1 represents the fitness

cost after many generations, $t_{1/2}$ denotes the time at which the fitness cost is halfway between the two, and *a* is given by,

$$a = \frac{s_0 - s_1}{1 - 2^{-t_f / t_{1/2}}}$$

Finally, we considered a model in which lab-reared individuals homozygous for the translocation and their translocation homozygote offspring have reduced fitness if they are not the result of outbreeding with wild-type individuals. We accordingly subdivided translocation homozygotes into those that are outbred, denoted by the genotype *TTRRo*, and those that are not, denoted by the genotype *TTRRn*. Adapting the above modeling framework, the proportions of the *k*th generation corresponding to these genotypes are denoted by p_k^{TTRRo} and p_k^{TTRRn}, and the haplotypes that determine the frequencies of outbred and non-outbred translocation homozygotes in the next generation – *TRo* and *TRn* – are described by the frequencies:

$$f_k^{TRo} = p_k^{TTRRo}(1 - s_o) + 0.25 p_k^{TtRr}(1 - hs_o)$$

$$f_k^{TRn} = p_k^{TTRRn}(1 - s_n)$$

Here, s_o denotes the reduced fecundity of outbred *TTRRo* individuals relative to wild-type individuals, and s_n denotes the reduced fecundity of non-outbred *TTRRn* individuals, where $0 \le s_o \le s_n < 1$. By considering all possible mating pairs, the genotype frequencies in the next generation are:

$$p_{k+1}^{TTRRo} = ((f_k^{TRo})^2 + 2 f_k^{TRo} f_k^{TRn}) / \sigma_k$$
$$p_{k+1}^{TTRRn} = (f_k^{TRn})^2 / \sigma_k$$
$$p_{k+1}^{TtRr} = 2(f_k^{TRo} f_k^{tr} + f_k^{TRn} f_k^{tr} + f_k^{tR} f_k^{Tr}) / \sigma_k$$
$$p_{k+1}^{ttrr} = (f_k^{tr})^2 / \sigma_k$$

where σ_k is a normalizing term given by,

$$\sigma_k = (f_k^{TRo})^2 + (f_k^{TRn})^2 + 2(f_k^{TRo} f_k^{TRn} + f_k^{TRo} f_k^{tr} + f_k^{TRn} f_k^{tr} + f_k^{tR} f_k^{Tr}) + (f_k^{tr})^2$$

Model fitting and model comparison:

We estimated fitness parameters for each model and compared models according to their Akaike Information Criterion (AIC) values. Model fitting was performed using population count data for the 18 drive experiments conducted for each translocation system (three for each of the 80%, 70%, 60%, 40%, 30% and 20% release frequencies). AIC was calculated as $2k - 2\log L$, where k denotes the number of model parameters, and the preferred model is the one with the smallest AIC value. The likelihood of the data was calculated, given the fitness cost parameters, assuming a binomial distribution of the two phenotypes (individuals homozygous or heterozygous for the translocation were considered as the same phenotype to match the experimental counts). Model predictions were used to generate expected genotype proportions over time for each fitness cost, and the log likelihood had the form,

$$\log L(h,s) = \sum_{i=1}^{18}\sum_{k=1}^{14} \log\binom{TTRR_{i,k} + TtRr_{i,k} + ttrr_{i,k}}{TTRR_{i,k} + TtRr_{i,k}} + ttrr_{i,k} \log(p_{i,k}^{ttrr}(h,s)) + (TTRR_{i,k} + TtRr_{i,k}) \log(1 - p_{i,k}^{ttrr}(h,s))$$

Here, $TTRR_{i,k}$, $TtRr_{i,k}$ and $ttrr_{i,k}$ represent the number of *TTRR*, *TtRr* and *ttrr* individuals at generation k in experiment i, and the corresponding expected genotype frequencies are fitness cost-dependent. In the case of the introgression model, the log likelihood had the form,

$$\log L(s_n, s_o, h) = \sum_{i=1}^{18}\sum_{k=1}^{14} \log\binom{TTRRn_{i,k} + TTRRo_{i,k} + TtRr_{i,k} + ttrr_{i,k}}{TTRRn_{i,k} + TTRRo_{i,k} + TtRr_{i,k}} + ttrr_{i,k} \log(p_{i,k}^{ttrr}(s_n, s_o, h)) + (TTRRn_{i,k} + TTRRo_{i,k} + TtRr_{i,k}) \log(1 - p_{i,k}^{ttrr}(s_n, s_o, h))$$

Here, $TTRRn_{i,k}$ and $TTRRo_{i,k}$ represent the number of *TTRRn* and *TTRRo* individuals at generation k in experiment i. The best estimate of the fitness cost is that having the highest log-likelihood. A 95% credible interval was estimated using a Markov Chain Monte Carlo sampling procedure. Matlab and R code implementing these equations is

available upon request.

Results:

The AIC values for each of the fitness cost models are shown in the table below:

Fitness cost model:	**AIC (Translocation system 1):**	**AIC (Translocation system 2):**
Constant fitness costs	6800.6	7781.9
Linear, frequency-dependent fitness costs	6094.8	6955.6
Linear, time-dependent fitness costs	6412.6	7267.2
Sigmoidal, time-dependent fitness costs	5748.5	6814.6
Exponential, time-dependent fitness costs	6059.3	6490.4
Introgression model	5271.5	5883.3

In summary, the best fitting model for the observed population dynamics is one in which non-outbred individuals homozygous for the translocation have reduced fitness as compared to individuals who are the result of at least one cross with a wild-type individual. Calculations of fitness parameters for translocation system 1 suggest a fitness cost for non-outbred translocation homozygotes of 0.466 (95% CrI: 0.459-0.474) relative wild-types, and essentially no fitness cost among outbred individuals – a much reduced fitness cost is estimated for outbred translocation homozygotes of 0.00016 (95% CrI: 0.00001-0.00092), which is reduced by a fraction 0.36 (95% CrI: 0.01-0.85)

among translocation heterozygotes. For translocation system 2, a similar dynamic is seen. The fitness cost for non-outbred translocation homozygotes is 0.465 (95% CrI: 0.459-0.471) relative wild-types, and for outbred translocation homozygotes is 0.00014 (95% CrI: 0.00001-0.00073), reduced by a fraction 0.24 (95% CrI: 0.01-0.95) among translocation heterozygotes.

The resulting population dynamics for the best fitting model (the introgression model) are shown in Supplemental Figure S1. Here, we see that, even for the best fitting model, the observed dynamics are not accurately captured. Marked individuals are expected to increase in frequency in the second generation before stabilizing for super-threshold releases and decreasing for sub-threshold releases. This is largely due to the fact that translocation homozygotes are released, and in the second generation, both translocation homozygotes and heterozygotes are seen as marked. The introgression model attempts to compensate for the discrepancy in model predictions at generation two and beyond through the incorporation of fitness costs; however, among other discrepancies, this prevents the model from predicting drive to fixation for 60% releases, as observed in the experimental results.

Of note, the expected threshold dynamics were seen for both sets of reciprocal translocations in laboratory drive experiments; however, further experimental and modeling work will be required to understand the discrepancies between the observed and predicted population dynamics, and any mechanisms that may be responsible for them. In particular, it will be interesting to see whether similar dynamics are seen for subsequent translocation systems, and whether further insight can be gained from marking procedures capable of discerning between translocation heterozygotes and homozygotes.

References:

(1) Curtis, C. F.; Robinson, A. S. Computer Simulation of the Use of Double Translocations for Pest Control. *Genetics* **1971**, *69* (1), 97–113.

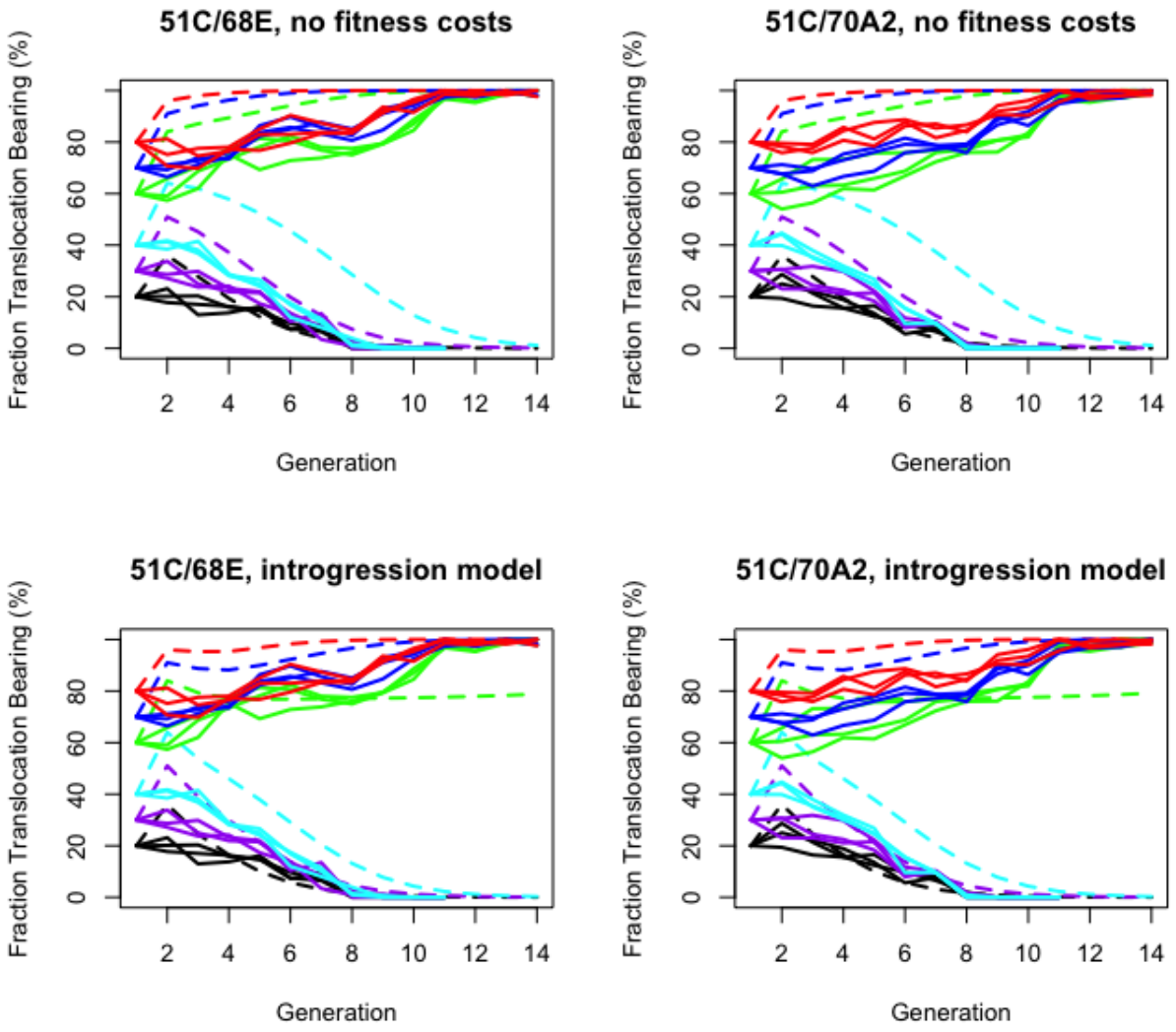

Figure S1. Dynamics of translocation-based population replacement, and predictions from zero fitness cost, and best fit models. (A, B) Population frequency of the adult population having the indicated translocation is plotted versus generation number for a number of homozygous translocation release ratios: 80%, 70%, 60%, 40%, 30% and 20%. Solid lines indicate observed population frequencies, and dashed lines indicate predicted translocation-bearing genotype frequencies for an element with no fitness cost. (C, D). The same data as in (A, B) but plotted along with dynamics predicted based on a best fit model described in the methods and text.

Supplementary Table 1. List of primer sequences used in this study.

Primer name	Primer sequence, 5' to 3'	Source
P1 P2	CCTAACAACTCACACCTTGCAGCGCCACCTGGCCCTAGAGATCCACCAACTTTTTTGCACTGC ATTCCTAAGCATCAGTGGTTGAACCTACCTTGTTGGCGTGACCAGAGACAGGTTGCGGCG	pIZ/V5-His/CAT (Invitrogen)
P3 P4	AGGTTCAACCACTGATGCTTAGGAATAGGCCATGTGAAGCTGAAGGAATC TATTACCCTGTTATCCCTACTAGTAGGGATAACAGGGTAATACTAGAATCCCTGGGCACAATTT	pFUSEss-CHIg-mG1 (Invivogen)
P5 P6	CTAGTATTACCCTGTTATCCCTACTAGTAGGGATAACAGGGTAATAGTGGTTGTAAGCCTTGCA AAAGGATAAGAATTAGGGTTAGTCGTTTCGGTGTGCCTAGTTTACCAGGAGAGTGGGAGA	pFUSEss-CHIg-mG1 (Invivogen)
P7	CGCCCACGCCATCCAACCGCCGCCGCAACCTGTCTCTGGTCACGCCAACAAGGTAGGTTC	P3/P4 XYZ PCR
P8	ATGACGTTCTTGGAGGAGCGCACCATTTTGTTGCTAAAGGAAAGGATAAGAATTAGGGTT	P5/P6 UVW PCR
P9 P10	AAACGACTAACCCTAATTCTTATCCTTTCCTTTAGCAACAAAATGGTGCGCTCCTCCAAG AATGGAACTCTTCGCGGCCAGGTGGCGCTGCAAGGCTCGAGGGTCGACTGATCATAATCA	pMos-3xP3-DsRed-attp (addgene plasmid #52904)
P11 P12 P15	GGATCCGGGAATTGGGAATTGGGCAATATTTAAATGGCGGCCTTGCAGCGCCACCTGGCC AGCGTGTTTTTTTGCAGTGCAAAAAAGTTGGTGGATCTCTAGGGCCAGGTGGCGCTGCAA CCAACGCATTTTCCAAGCTTGTTTAAACGTGGATCTCTAGGGCCAGGTGGCGCTGCAAGG	Drosophila genomic DNA
P13 P14	TACAAATGTGGTATGGCTGATTATGATCAGTCGACCCTCGAGCCTTGCAGCGCCACCTGG GAGACCGTGACCTACATCGTCGACACTAGTGGATCTCTAGGGCCAGGTGGCGCTGCAAGG	Drosophila genomic DNA
P16	CCTTGCAGCGCCACCTGGCCCTAGAGATCCA	Drosophila

	CGTTTAAACAAGCTTGGAAAATGCGTTGG	genomic DNA
P17	CGAAGCGCCTCTATTTATACTCCGGCGCTCG TTTAAACAAAGTGGCAGGGCCCATGTGTT	
P18	GAGTGGAGCACAAACACATGGGCCCTGCCA CTTTGTTTAAACGAGCGCCGGAGTATAAAT	Drosophila genomic DNA
P19	AAGCATCAGTGGTTGAACCTACCTTGTTGGC GTGTCTGATGCAGATTGTTTAGCTTGTTC	
P20	GCCAACAAGGTAGGTTCAACCACTGATGCTT AGGAATAGGCGTGGTTGTAAGCCTTGCAT	pFUSEss-CHIg-mG1 (Invivogen)
P21	CCCTGTTATCCCTACTAGTAGGGATAACAGG GTAATACTAGTTTACCAGGAGAGTGGGAG	
P22	TATTACCCTGTTATCCCTACTAGTAGGGATAA CAGGGTAATACATGTGAAGCTGAAGGAA	pFUSEss-CHIg-mG1 (Invivogen)
P23	AAAGGATAAGAATTAGGGTTAGTCGTTTCGG TGTGCCTAGAATCCCTGGGCACAATTTTC	
P24	CAAGCGCAGCTGAACAAGCTAAACAATCTGC ATCAGACACGCCAACAAGGTAGGTTCAAC	P20/P21 UVW PCR
P25	ACCTACATCGTCGACACTAGTGGATCTCTAG CTCGAGCTAAAGGAAAGGATAAGAATTAGGG	P22/P23 XYZ PCR
P26	CCCTAATTCTTATCCTTTCCTTTAGGAATTCC AACAAAATGGTGAGCAAGGGCGAGGAGC	pAAV-GFP (addgene plasmid
P27	TTCACTGCATTCTAGTTGTGGTTTGTCCAAAC TCATCAATGTTTACTTGTACAGCTCGTC	#32395)
P28	GCCGCCGGGATCACTCTCGGCATGGACGAG CTGTACAAGTAAACATTGATGAGTTTGGAC	pMos-3xP3-DsRed-attp (addgene plasmid #52904)