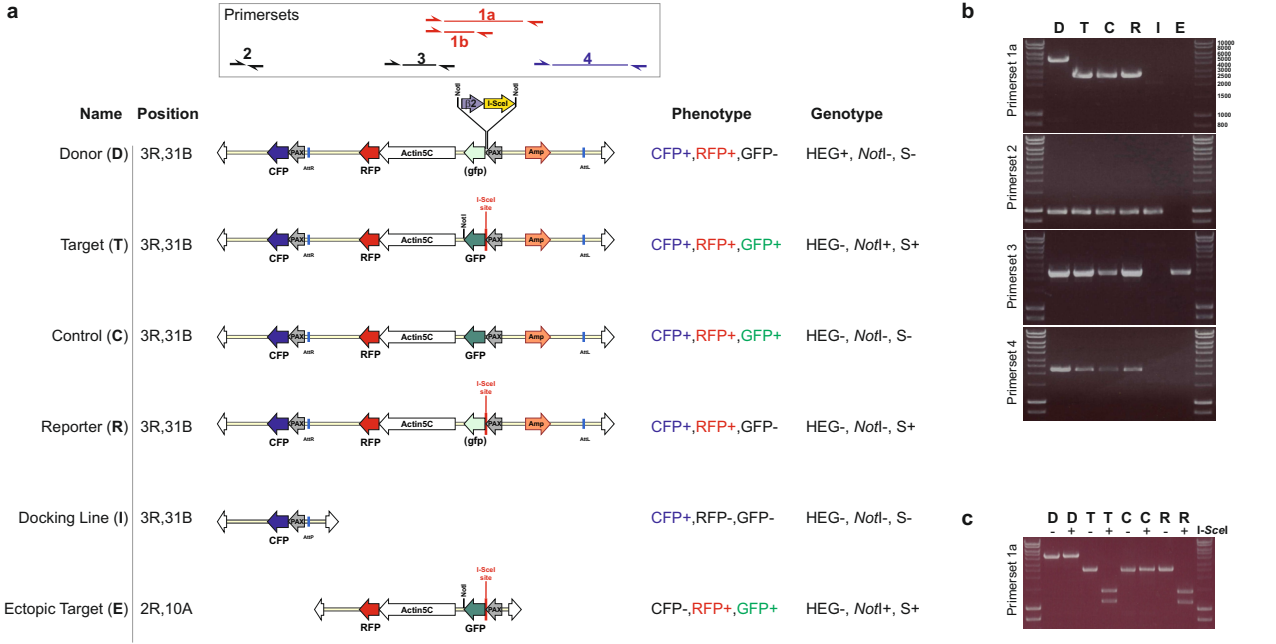
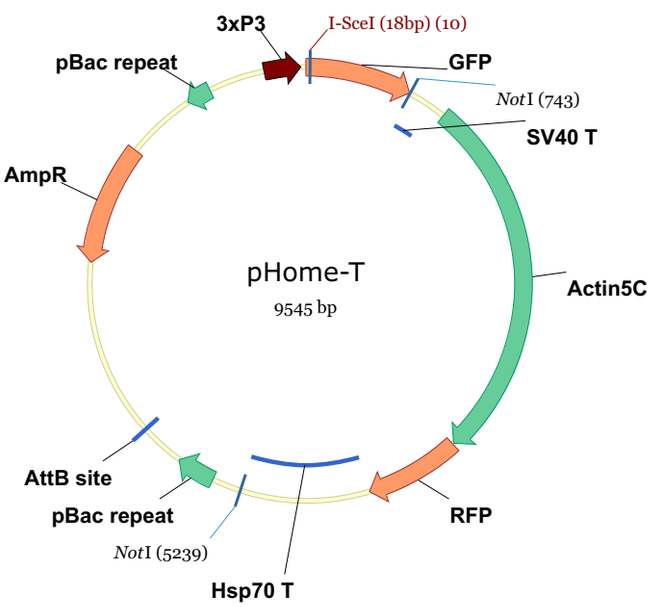
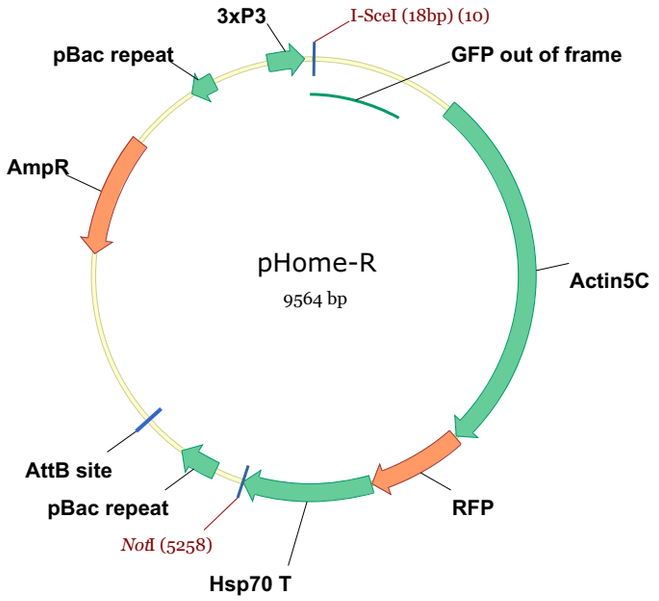
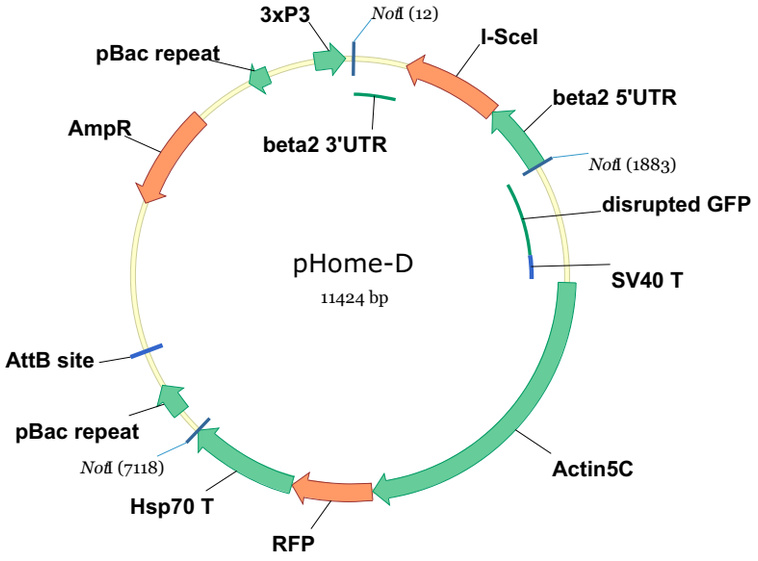

a
Primersets
1a
1b
2
3
4
Name
Position
Phenotype
Genotype
Donor (D)
3R,31B
CFP+,RFP+,GFP-
HEG+, NotI-, S-
Target (T)
3R,31B
CFP+,RFP+,GFP+
HEG-, NotI+, S+
Control (C)
3R,31B
CFP+,RFP+,GFP+
HEG-, NotI-, S-
Reporter (R)
3R,31B
CFP+,RFP+,GFP-
HEG-, NotI-, S+
Docking Line (I)
3R,31B
CFP+,RFP-,GFP-
HEG-, NotI-, S-
Ectopic Target (E)
2R,10A
CFP-,RFP+,GFP+
HEG-, NotI+, S+
I-SceI site
NotI
I-SceI
Actin5C
PAX
Amp
CFP
RFP
GFP
(gfp)
AttR
AttL
AttP
b
D T C R I E
Primerset 1a
Primerset 2
Primerset 3
Primerset 4
10000
5000
3000
2500
2000
1500
1000
800
c
D D T T C C R R
- + - + - + - +
I-SceI
Primerset 1a

Supplementary Figure 1. Structure of transgenic loci.
a, Predicted structure of the transgenic loci following integration of the transformation constructs within the genomic location of the φC31 integrase docking strain on chromosome 3R. Information concerning the genomic location, the fluorescent phenotype (CFP+/-; RFP+/- and GFP +/-) and the genotype (HEG+/-; I-SceI site, S+/-; and *Not*I site, *Not*I+/-) of the Donor (D), Target (T), Control (C), Reporter (R), and Ectopic Target (E) lines as well as the original Docking line (I) is provided. Marker genes and their regulatory elements are drawn to scale. The relative position of the PCR primer sets were used to genotype the constructs are shown on top (1a, 1b, 2, 3 and 4). **b**, The molecular profile of homozygous transgenic individuals for each primerset. **c**, I-SceI in vitro digestion reactions of PCR products obtained from genomic DNA of the Donor, Target, Control and Reporter genomic constructs using primerset 1a.

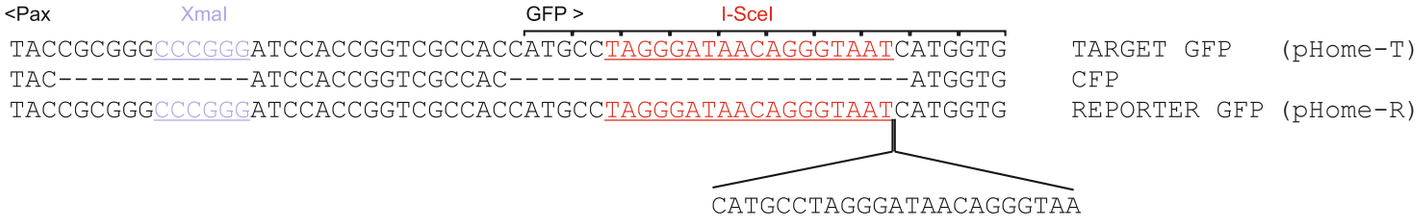
a


3xP3
I-SceI (18bp) (10)
pBac repeat
GFP
NotI (743)
SV40 T
AmpR
pHome-T
9545 bp
Actin5C
AttB site
pBac repeat
NotI (5239)
RFP
Hsp70 T

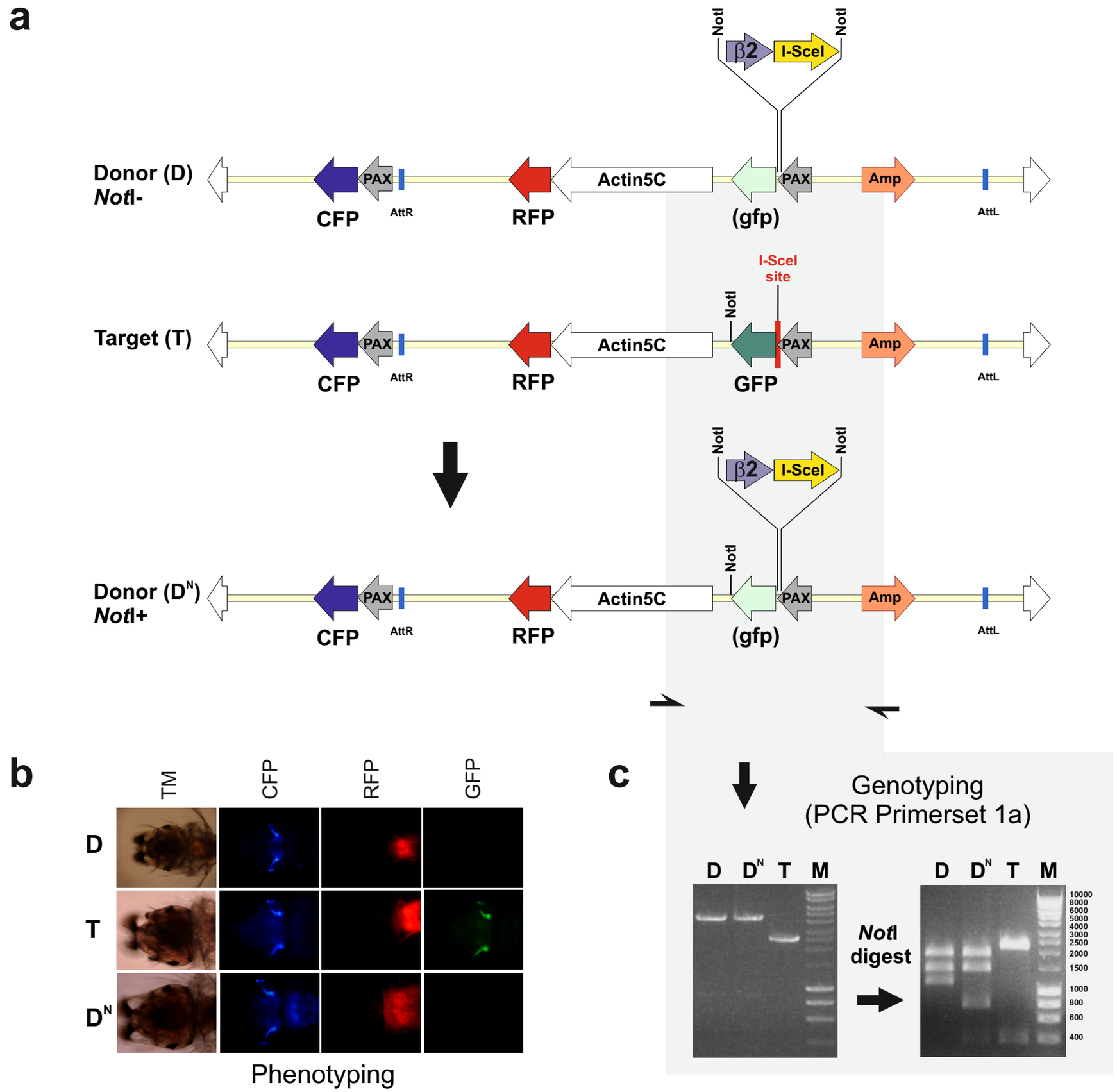

3xP3
I-SceI (18bp) (10)
pBac repeat
GFP out of frame
AmpR
pHome-R
9564 bp
Actin5C
AttB site
pBac repeat
NotI (5258)
RFP
Hsp70 T


3xP3
NotI (12)
I-SceI
pBac repeat
beta2 5'UTR
AmpR
beta2 3'UTR
NotI (1883)
disrupted GFP
pHome-D
11424 bp
SV40 T
AttB site
pBac repeat
NotI (7118)
Actin5C
Hsp70 T
RFP

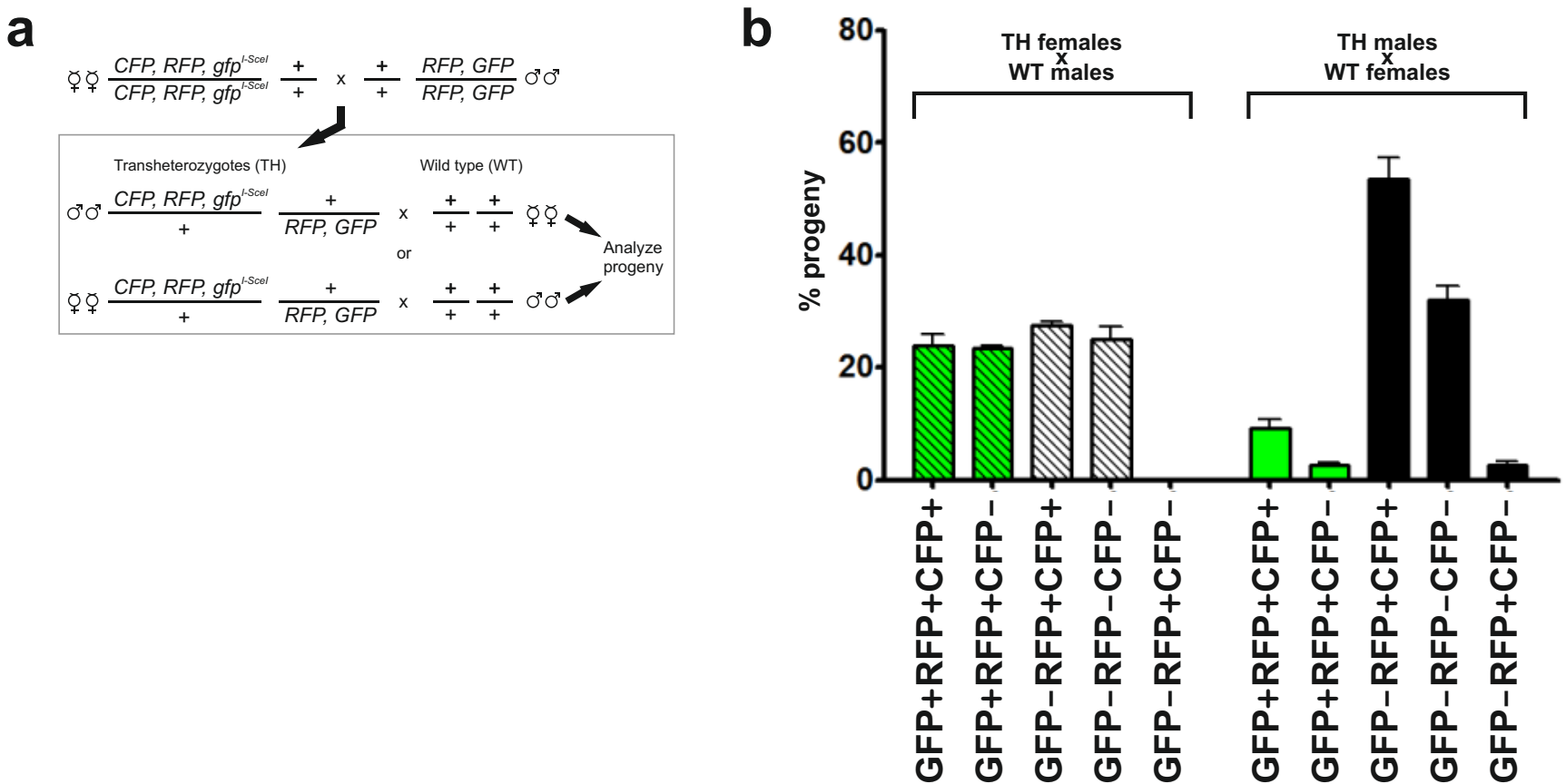
b


<Pax
XmaI
GFP >
I-SceI
TACCGCGGGCCCGGGATCCACCGGTCGCCACCATGCCTAGGGATAACAGGGTAATCATGGTG TARGET GFP (pHome-T)
TAC----------ATCCACCGGTCGCCAC----------ATGGTG CFP
TACCGCGGGCCCGGGATCCACCGGTCGCCACCATGCCTAGGGATAACAGGGTAATCATGGTG REPORTER GFP (pHome-R)
CATGCCTAGGGATAACAGGGTAA

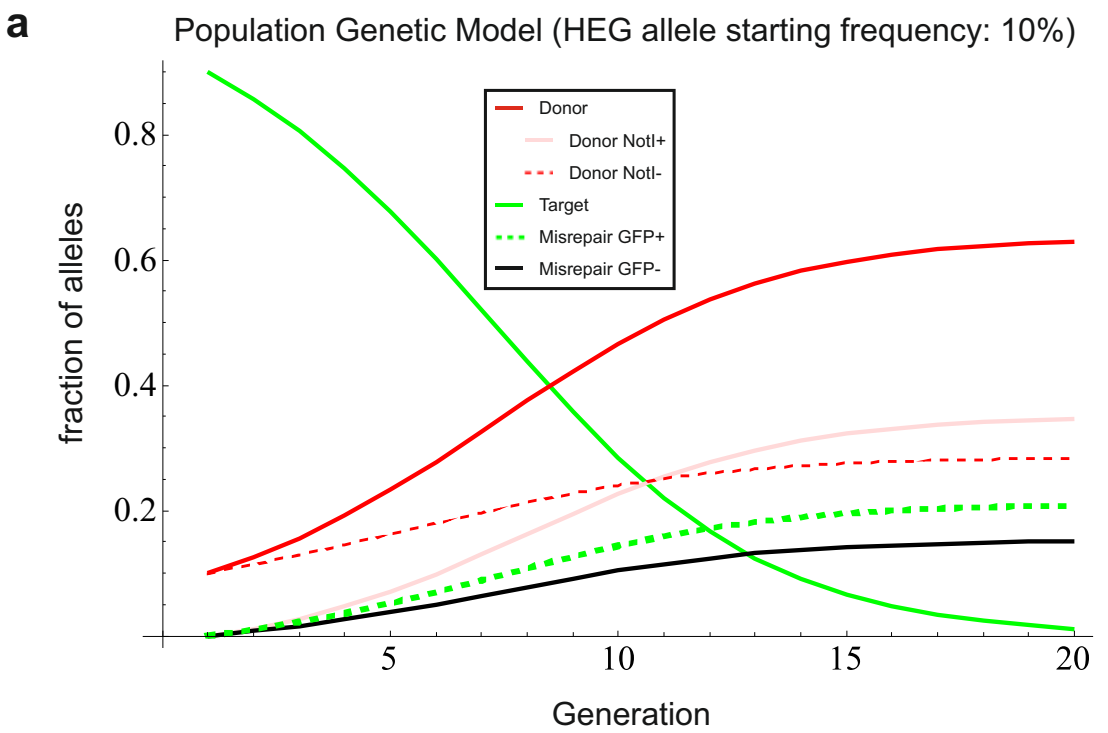
Supplementary Figure 2. Structure of transformation vectors.
a, Schematic representation of the plasmid vectors pHome-T, pHome-R and pHome-D used in this study. **b**, Nucleotide sequence of the I-SceI cleavage sites in pHome-T and pHome-R and the corresponding region joining the 3xP3 promoter and the CFP gene in the AttP docking line. The GFP CDS in pHome-T construct contains the I-SceI site in frame and remains functional. The construct pHome-R contains a partially duplicated I-SceI generating an out of frame GFP gene. The region between the 3xP3 and the CFP marker lacks both the I-SceI and the *Xma*I site.

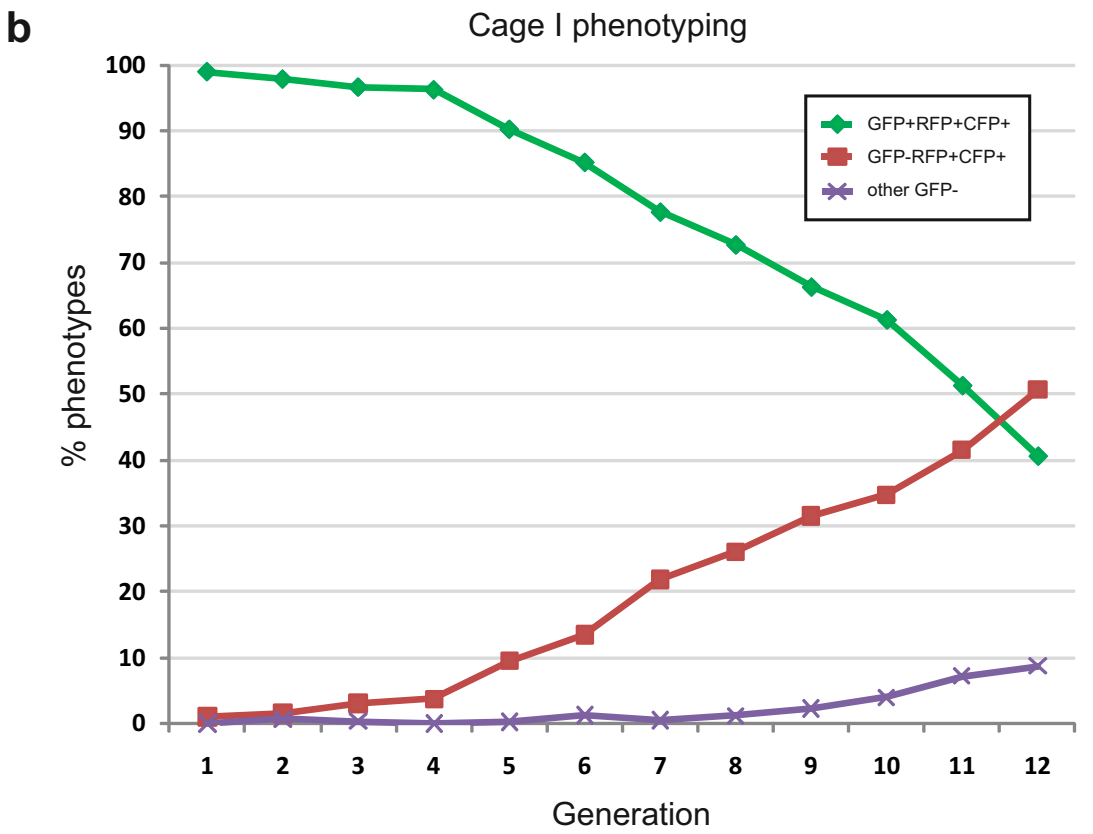

a
NotI
β2
I-SceI
NotI
Donor (D)
NotI-
PAX
Actin5C
PAX
Amp
CFP
AttR
RFP
(gfp)
AttL
I-SceI
site
Target (T)
NotI
PAX
Actin5C
PAX
Amp
CFP
AttR
RFP
GFP
AttL
NotI
β2
I-SceI
NotI
Donor (DN)
NotI+
NotI
PAX
Actin5C
PAX
Amp
CFP
AttR
RFP
(gfp)
AttL
b
TM
CFP
RFP
GFP
D
T
DN
Phenotyping
c
Genotyping
(PCR Primerset 1a)
D
DN
T
M
NotI
digest
D
DN
T
M
10000
8000
6000
5000
4000
3000
2500
2000
1500
1000
800
600
400

Supplementary Figure 3. Phenotyping and Genotyping of Donor and Target alleles.
a, Schematic representation of the Target (T), the Donor *Not*I- (D) and the Donor *Not*I+ (D^N) alleles. **b**, Transmission (TM) fluorescence green (GFP), red (RFP) and blue (CFP) micrographs of hemizygous larvae carrying Target (T), the Donor *Not*I- (D) and the Donor *Not*I+ (D^N) alleles. **c**, Molecular diagnostic test utilized to differentiate between T, D and D^N loci based on PCR amplification of genomic DNA and in-vitro digestion with *Not*I.

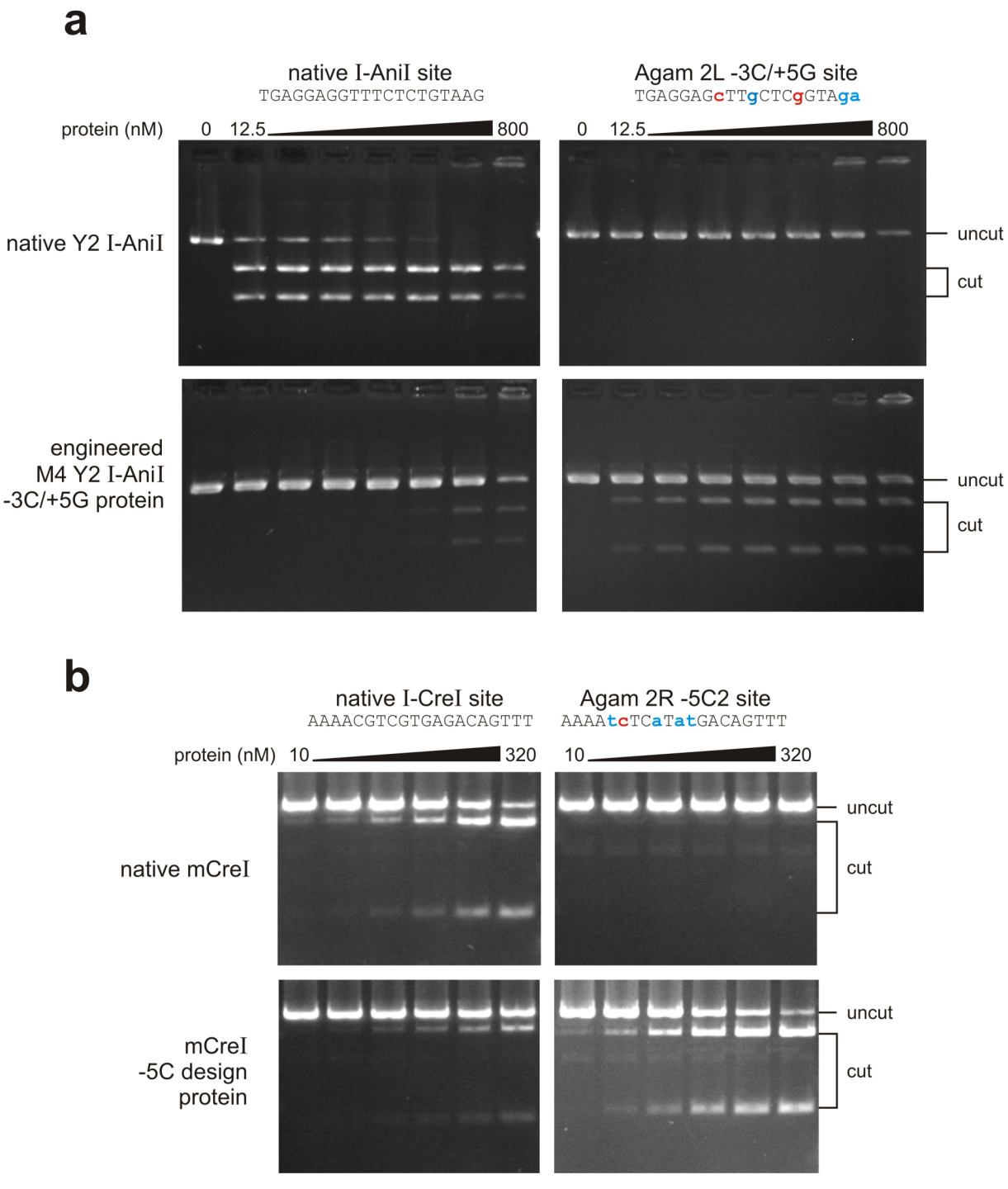

a
CFP, RFP, gfp I-SceI
RFP, GFP
Transheterozygotes (TH)
Wild type (WT)
or
Analyze progeny
b
TH females x WT males
TH males x WT females
% progeny
80
60
40
20
0
GFP+RFP+CFP+
GFP+RFP+CFP−
GFP−RFP+CFP+
GFP−RFP−CFP−
GFP−RFP+CFP−

Supplementary Figure 4. Analysis of HEG activity in Donor/Ectopic Target transgenics.
Phenotypic analysis of the progeny obtained from crosses of Donor/Ectopic Target TH with WT mosquitoes. **a**, Crossing scheme to generate hemizygous individuals that allow the scoring of a single chromosome from the double transgenic parents. **b**, The percentage of individuals identified to fall into the different phenotype classes in the TH female x WT male and TH male x WT female crosses.


a
Population Genetic Model (HEG allele starting frequency: 10%)
Donor
Donor NotI+
Donor NotI-
Target
Misrepair GFP+
Misrepair GFP-
fraction of alleles
0.8
0.6
0.4
0.2
5
10
15
20
Generation


b
Cage I phenotyping
GFP+RFP+CFP+
GFP-RFP+CFP+
other GFP-
% phenotypes
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
11
12
Generation

Supplementary Figure 5. HEG invasion dynamics.
a, Predicted frequencies of Donor (Red), Target (Green), *Not*I+ Donor (Light Red), *Not*- Donor (Dotted Red), GFP- misrepaired (Black) and GFP+ misrepaired alleles (Dotted Green) alleles derived from a deterministic model for HEG spread in an idealized population. The model was parameterized assuming that the starting frequency of the HEG allele of 10%; the rate of cleavage of Targets in the presence of Donors is c=0.95; the rate of homing in males is e_m=0.56, when homing occurs the probability that the *Not*I site is not co-converted is n=0.45; and when cleavage occurs but homing does not the probability the repaired allele is GFP+ is r=0.58. **b**, Temporal dynamics of GFP+RFP+CFP+ (Green) and GFP-RFP+CFP+ (Red) phenotypes in mosquitoes from cage 1. The percentage of other GFP- phenotypes including RFP-GFP-CFP+, RFP-GFP-CFP- and occasional observed phenotypes such as 3xP3-RFP or Actin5C-GFP are collectively shown as a blue line.


a
native I-AniI site
TGAGGAGGTTTCTCTGTAAG
Agam 2L -3C/+5G site
TGAGGAGcTTgCTCgGTAga
protein (nM)
0
12.5
800
native Y2 I-AniI
uncut
cut
engineered
M4 Y2 I-AniI
-3C/+5G protein
b
native I-CreI site
AAAACGTCGTGAGACAGTTT
Agam 2R -5C2 site
AAAAtcTCaTatGACAGTTT
protein (nM)
10
320
native mCreI
mCreI
-5C design
protein

Supplementary Figure 6. Cleavage of *Anopheles gambiae* chromosomal target sites by engineering variants of the I-AniI and mCreI homing endonucleases.

a, Cleavage of the native I-AniI target site (left panels) and the *Anopheles* 2L -3C/+5G chromosomal target site (right panels) by native Y2 I-AniI (top) and by an engineered -3C/+5G variant of Y2 I-AniI containing two additional residue substitutions (M4 Y2 I-AniI: Y2 I-AniI + F91I and S92T substitutions; bottom). Native Y2 I-AniI protein cleaved the native I-AniI target site to completion at 400 nM, whereas the 2L *Anopheles* -3C/+5G chromosomal target site was cleavage-resistant even at 800 nM, the highest protein concentration tested (top row panels). In contrast, engineered M4 Y2 I-AniI -3C/+5G protein cleaved the *Anopheles* 2L -3C/+5G target site at the lowest protein concentration tested (12.5 nM; lower right panel), but was able to cleave the native I-AniI target site only at substantially higher protein concentrations (≥100 nM; lower row panels). Upper case base pairs in all target site sequences represent native target site base pairs. Lower case base pairs in red are design/ engineering target base pairs, and lower case base pairs in blue are predicted to be cleavage-sensitive from prior single base pair scan cleavage degeneracy analyses. **b**, Cleavage of the native I-CreI target site (left panels) and the *Anopheles* 2R -5C chromosomal target site (right panels) by native mCreI protein (top) and a designed -5C-specific mCreI protein variant (bottom). Native mCreI protein cleaved the native I-CreI target site at protein concentrations >20 nM, whereas the *Anopheles* 2R -5C2 chromosomal target site was cleavage-resistant at 320 nM, the highest protein concentration tested (top row panels). In contrast, the mCre -5C design protein cleaved the *Anopheles* 2R -5C2 chromosomal target site at the lowest protein concentration tested, 10 nM, but was able to cleave the native I-CreI site only at higher protein concentrations (≥40 nM). The difference in mCre -5C activity on design and native target sites is readily apparent in digests that used higher protein concentrations (≥160nM; lower row panels).

Molecular characterization of transgenic loci from 3 phenotypic classes

			Donor/Target	Donor/Reporter
GFP+ RFP+ CFP+	Sequencing of PCR products (Primerset 1a)		36	20
	GFP WT sequence	(uncut)	7	0
	GFP Deletions	(<10bp)	7	0
	GFP Deletions	(>10bp)	3	4
	GFP Insertions	(<10bp)	4	1
	GFP Insertions	(>10bp)	1	0
	CFP like sequence*		14	15
GFP- RFP+ CFP+	PCR (Primerset 1a) and *Not*I digest of PCR product		156	
	Donor sized (HEG+)		152	
		*Not*I marker present	25	
		*Not*I marker absent	127	
	Target sized (HEG-)		4	
GFP- RFP- CFP+	Sequencing of PCR products (Primers: 2 fwd-1a rev)		15	
	GFP/CFP recombination**		5	
	no PCR product		10	

* Synthesis-dependent strand annealing using the 3xP3-CFP locus as template could account for these repair events

** The structure of these events indicates that they might have originated through intramolecular recombination via single strand annealing between the homologous regions of the 3xP3-GFP and 3xP3-CFP genes leading to the loss of the interjacent RFP marker.

Supplementary Table 1. Molecular characterization of transgenic loci from 3 phenotypic classes.