doi:10.1038/nature09937

A synthetic homing endonuclease–based gene drive system in the human malaria mosquito

Nikolai Windbichler[1], Miriam Menichelli[1], Philippos Aris Papathanos[1], Summer B. Thyme[2,3], Hui Li[4], Umut Y. Ulge[4,5], Blake T. Hovde[6], David Baker[2,3,7], Raymond J. Monnat Jr[4,5,6], Austin Burt[1,8]* & Andrea Crisanti[1,9]*

Genetic methods of manipulating or eradicating disease vector populations have long been discussed as an attractive alternative to existing control measures because of their potential advantages in terms of effectiveness and species specificity[1–3]. The development of genetically engineered malaria-resistant mosquitoes has shown, as a proof of principle, the possibility of targeting the mosquito's ability to serve as a disease vector[4–7]. The translation of these achievements into control measures requires an effective technology to spread a genetic modification from laboratory mosquitoes to field populations[8]. We have suggested previously that homing endonuclease genes (HEGs), a class of simple selfish genetic elements, could be exploited for this purpose[9]. Here we demonstrate that a synthetic genetic element, consisting of mosquito regulatory regions[10] and the homing endonuclease gene *I-SceI*[11–13], can substantially increase its transmission to the progeny in transgenic mosquitoes of the human malaria vector *Anopheles gambiae*. We show that the *I-SceI* element is able to invade receptive mosquito cage populations rapidly, validating mathematical models for the transmission dynamics of HEGs. Molecular analyses confirm that expression of I-SceI in the male germline induces high rates of site-specific chromosomal cleavage and gene conversion, which results in the gain of the *I-SceI* gene, and underlies the observed genetic drive. These findings demonstrate a new mechanism by which genetic control measures can be implemented. Our results also show in principle how sequence-specific genetic drive elements like HEGs could be used to take the step from the genetic engineering of individuals to the genetic engineering of populations.

HEGs encode highly specific endonucleases with recognition sequences that typically occur only once per host genome, and have been identified in unicellular organisms in all three biological domains[14]. HEG-induced DNA double strand breaks (DSB) activate the recombinational repair system of the cell, which uses the homologous chromosome carrying the HEG as a template for repair. As a result the HEG is copied to the broken chromosome in a process referred to as 'homing'. HEGs use this transmission distortion mechanism to spread through populations[15]. To investigate I-SceI activity *in vivo* we have developed an experimental system consisting of three distinct transgenic mosquito lines, the Donor, the Reporter and the Target, carrying either the *I-SceI* gene or its recognition site at identical positions on homologous chromosomes (Supplementary Fig. 1). For this purpose we used an *A. gambiae* docking line[16] that allowed the site-specific integration of three different plasmids carrying the red fluorescent protein (RFP) transformation marker on chromosome 3R (Supplementary Fig. 2). The Donor line was generated using the construct pHome-D, containing a 3×P3-GFP (green fluorescent protein) transcription unit interrupted by a synthetic HEG element consisting of the *I-SceI* gene and the regulatory regions of the male testis-specific *A. gambiae β2-tubulin* gene[10]. The Reporter line was developed using the construct pHome-R, containing an I-SceI cleavage site that shifts out of frame the coding sequence of the *GFP* gene. The Reporter locus allows the scoring of I-SceI cleavage activity by monitoring the frequency of GFP^+ individuals in which the GFP reading frame was restored via non-homologous end joining (NHEJ) in the progeny of Donor/Reporter trans-heterozygous males (Fig. 1a, b). Finally, the Target line was developed using pHome-T, containing the I-SceI cleavage site within the coding sequence of a functional *GFP* gene. This construct contains a diagnostic NotI recognition site that facilitates the molecular genotyping of homing events. The Target locus allows the assessment of I-SceI homing activity in the progeny of Donor/Target trans-heterozygous males crossed with wild-type females by measuring the increase in the frequency over a 1:1 ratio of GFP^- to GFP^+ individuals arising from the insertion of the HEG gene into the GFP open reading frame (Fig. 1d, e).

When Donor/Reporter trans-heterozygous females were crossed to wild-type males all progeny showed the expected GFP^- phenotype, as the *β2-tubulin* promoter regulating I-SceI is not active in females. By contrast 3% of the progeny from Donor/Reporter trans-heterozygous males and wild-type females showed a GFP^+ phenotype (Fig. 1c). Sequencing of PCR products from the region around the I-SceI site showed that in 5 out of 20 GFP^+ individuals the correct reading frame had been restored by NHEJ repair events. The remaining 15 GFP^+ mosquitoes showed in place of the I-SceI site a sequence that resembled the region joining the 3×P3 promoter and the *CFP* gene (cyan fluorescent protein), which lacks a unique restriction site present in the 3×P3-GFP cassette (Supplementary Fig. 2). We established from one such GFP^+ individual the HEG-resistant Control strain, containing all three fluorescent marker genes but lacking the I-SceI site within the GFP sequence (Supplementary Fig. 1). The remaining 97% of the progeny from Donor/Reporter trans-heterozygous males and wild-type females were GFP^- and the majority of these mosquitoes (93%) showed a $GFP^-RFP^+CFP^+$ phenotype expected to arise either from an intact GFP^- parental locus, NHEJ events that did not restore GFP expression or I-SceI homing events (Fig. 1c).

To test for the occurrence of homing we analysed the progeny of crosses between Donor/Target trans-heterozygous and wild-type mosquitoes (Fig. 1f). As expected, the ratio of GFP^+:GFP^- phenotypes in the offspring of Donor/Target trans-heterozygous females crossed to wild-type males was about 50:50. By contrast, in the reciprocal cross of trans-heterozygous males and wild-type females the ratio was 14:86. The excess of GFP^- progeny, the majority of which were RFP^+CFP^+, could originate either from NHEJ events or as a result of homologous repair involving the HEG^+ chromosome (that is, homing). To investigate the molecular nature of GFP inactivation we performed a PCR

[1]Imperial College London, Department of Life Sciences, South Kensington Campus, London, SW7 2AZ, UK. [2]Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. [3]Graduate Program in Biomolecular Structure and Design, University of Washington, Seattle, Washington 98195, USA. [4]Department of Pathology, University of Washington, Seattle, Washington 98195, USA. [5]Graduate Program in Molecular and Cellular Biology, University of Washington, Seattle, Washington 98195, USA. [6]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. [7]Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA. [8]Imperial College London, Department of Life Sciences, Silwood Park Campus, Ascot, SL5 7PY, UK. [9]Department of Experimental Medicine, University of Perugia, Via Del Giochetto, 06122 Perugia, Italy.
*These authors contributed equally to this work.

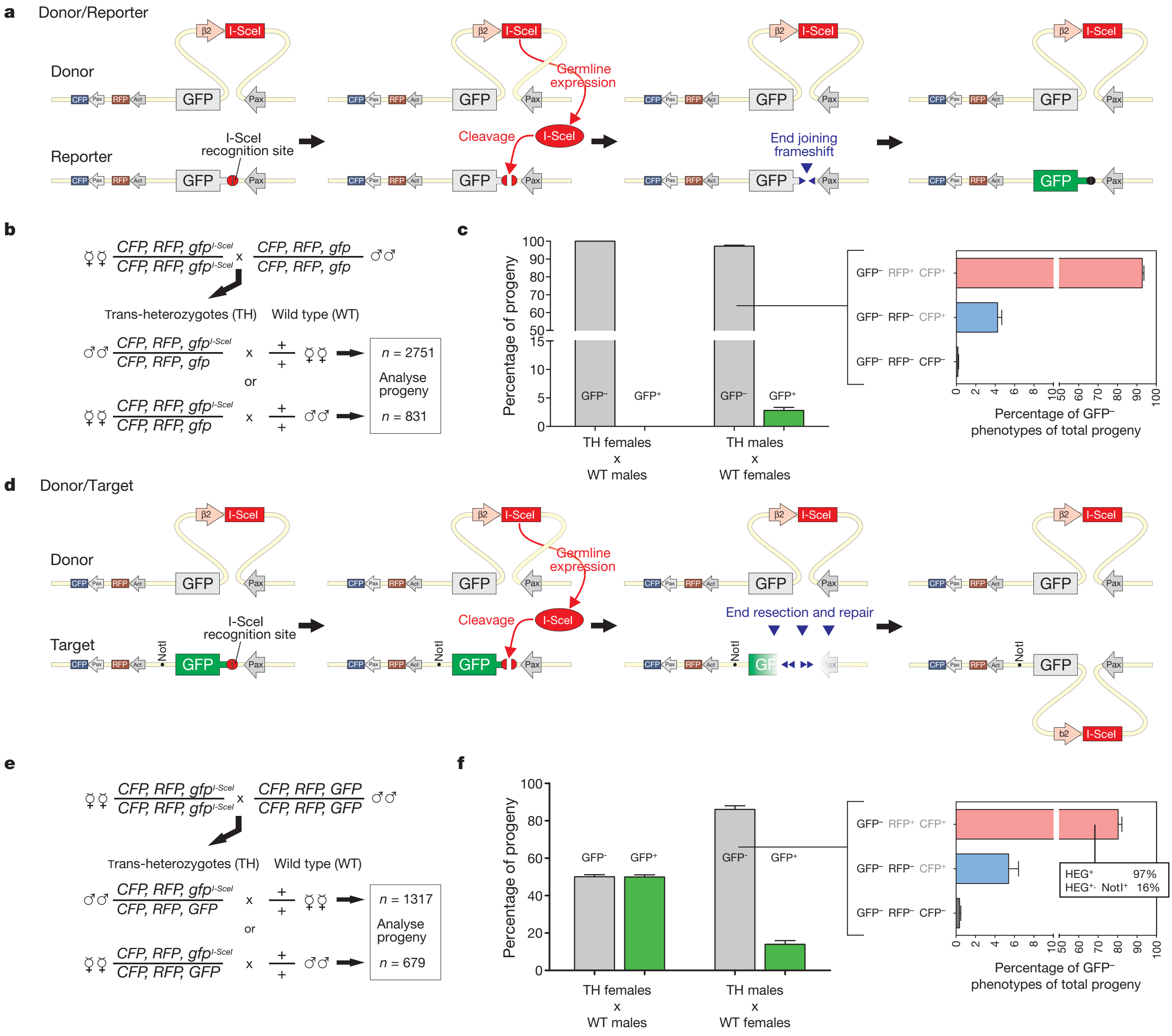

Figure 1 | Analysis of HEG activity in transgenic mosquitoes. **a**, **d**, Anticipated molecular events unfolding in Donor/Reporter (**a**) or Donor/Target (**d**) trans-heterozygous (TH) males. The Donor locus expresses I-SceI under the control of the male germline promoter *β2-tubulin*. The Reporter and the Target loci contain an I-SceI recognition site within the *GFP* gene. **a**, In Donor/Reporter TH males I-SceI activity is detected by scoring events that restore the *GFP* reading frame upon cleavage of the I-SceI recognition site. **d**, In Donor/Target TH males cleavage of the Target locus is followed by end resection and homing of the *I-SceI* gene from the homologous chromosome. This leads to the inactivation of the GFP reporter gene and can also lead to co-conversion of the NotI molecular marker (Pax, 3×P3 promoter; Act, *Actin5C* promoter). **c**, **f**, Phenotypic analysis of progeny from crosses of Donor/Reporter (**b**, **c**) or Donor/Target (**e**, **f**) trans-heterozygote with wild-type (WT) mosquitoes. The column graphs show the percentage of GFP^- and GFP^+ individuals. The bar graphs on the right show, as a percentage of the total progeny, the $GFP^-RFP^+CFP^+$, $GFP^-RFP^-CFP^+$ and $GFP^-RFP^-CFP^-$ individuals observed. The inset (**f**, right panel) shows the molecular genotype of $GFP^-RFP^+CFP^+$ individuals analysed for the presence of the HEG and the NotI molecular markers.

analysis of the region spanning the GFP locus and encompassing the *I-SceI* gene or its recognition site. The results showed that 97% of $GFP^-RFP^+CFP^+$ individuals contained the HEG cassette (Fig. 1f). The estimated cleavage rate for I-SceI was therefore about 95%, and the overall homing rate 56%. Importantly, the diagnostic NotI marker present only on the Target locus allowed the identification of recombinant GFP^- HEG^+ $NotI^+$ chromosomes that were generated as a result of homing events (Supplementary Fig. 3). We were able to detect the NotI site, located ~0.7 kilobases from the I-SceI cleavage site, in 16% of HEG^+ chromosomes analysed, indicating that this marker was retained in 45% of all homing events (and lost due to co-conversion in the remaining 55%). In both sets of male trans-heterozygous to wild-type crosses about 4–5% of the progeny were $GFP^-RFP^-CFP^+$, and a small number of larvae lacked all three visible markers (Fig. 1c, f). These phenotypes were not observed in progeny of trans-heterozygous females, suggesting that they were the result of I-SceI activity accompanied by deletions encompassing parts of the RFP gene or the entire locus. These experiments are summarized in Supplementary Table 1.

Another independent transgenic line, referred to as Ectopic Target, was generated by transposase-mediated integration of the pHome-T plasmid on chromosome 2 (Supplementary Fig. 1). When Donor/Ectopic Target trans-heterozygous males were crossed to wild-type females the frequency of the GFP^- phenotype in the progeny was 88%, compared to approximately 50% in the female trans-heterozygous

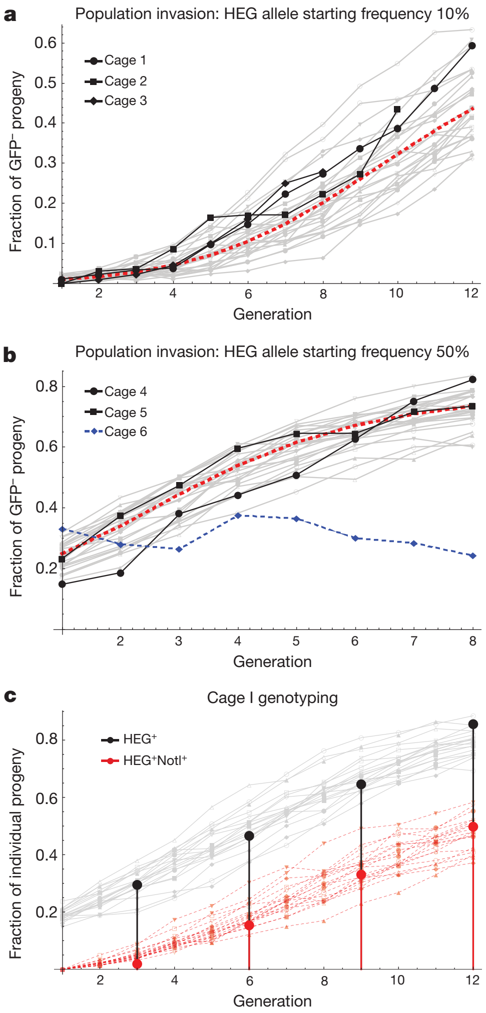

Figure 2 | HEG invasion in mosquito cage populations. a, **b**, Temporal dynamics of GFP$^-$ mosquitoes in populations in which the HEG Donor allele was seeded at a frequency of 10% (**a**) or 50% (**b**) into a background of GFP$^+$ mosquitoes carrying the HEG Target allele. The experimental points (black) are overlaid onto predicted dynamics derived from a deterministic population genetic model (dashed red line) and from 20 iterations of a stochastic model (grey lines). The dynamics of a cage population in which the HEG Donor and Control alleles were combined at a frequency of 50% is also shown (dashed blue line). **c**, Molecular genotyping performed on individuals randomly collected from cage 1 at generations 3,6,9 and 12 using a set of PCR primers that specifically amplifies the HEG cassette (Primer set 1b, Supplementary Fig. 2). Presence of the NotI marker was determined by *in vitro* digestion of PCR products using NotI. The graph shows the fraction of mosquitoes carrying the HEG (black) and the fraction carrying the HEG and the NotI marker on the same chromosome (red) overlaid onto predictions from 20 stochastic simulations (grey lines and dashed red lines, respectively).

control cross (Supplementary Fig. 4). However none of 94 GFP$^-$RFP$^+$CFP$^-$ individuals, the phenotypic class expected to contain non-parental HEGs, carried the HEG sequence. This experiment indicates that, in the absence of a repair template on the homologous chromosome, I-SceI cleavage activity does not induce detectable homing. Finally, we observed no significant deviation from a 1:1 ratio of GFP$^-$ and GFP$^+$ progeny from crosses of trans-heterozygous mosquitoes in which the Donor locus was combined with the Control locus (data not shown).

To test whether the observed transmission ratio distortion allows for efficient genetic drive of I-SceI in receptive *A. gambiae* populations, we monitored its transmission dynamics in five cage populations of 600 individuals over 8 to 12 generations. Cage populations containing the I-SceI Target allele at initial frequencies of 90% or 50% were seeded with the I-SceI Donor allele at a frequency of 10% or 50%, respectively. GFP dominance results in an initial frequency of GFP$^-$ individuals of 1% or 25% in the two experimental conditions. In subsequent generations GFP$^+$ individuals are expected to carry at least one allele of the original GFP$^+$ target gene or a misrepaired GFP$^+$ allele, whereas GFP$^-$ individuals contain two alleles in which GFP has been inactivated either by insertion of the HEG or NHEJ. At each generation a random sample of the progeny was visually analysed for the GFP marker at the larval stage. In all populations the frequency of GFP$^-$ individuals increased rapidly over time (Fig. 2). The frequency rose from about 1% to 60% in 12 generations (cage 1), and from about 1% to 40% in 10 generations (cage 2). In the two populations seeded with higher initial HEG frequencies GFP$^-$ individuals reached about 75–80% after 8 generations. By contrast the frequency of GFP$^-$ individuals did not change significantly in a population (cage 6) in which the HEG Donor line was used in combination with the non-receptive Control line (Fig. 2b), indicating that the absence of GFP expression in GFP$^-$RFP$^+$CFP$^+$ mosquitoes did not result in a measurable fitness advantage over GFP$^+$RFP$^+$CFP$^+$ mosquitoes. We generated deterministic and stochastic population genetic models, using as parameters the experimentally derived rates of cleavage, homologous repair and NHEJ, assuming no fitness differences among genotypes (Supplementary Fig. 5a). The observed dynamics in the population cages fall well within the stochastic variation expected from the model (Fig. 2), indicating a quantitative match between the experimental data and our theoretical understanding of HEG transmission dynamics. If I-SceI had any effect on mosquito fitness, it was not large enough to significantly affect this concordance.

Detailed phenotypic and molecular analyses were carried out at different generations on individuals sampled from the mosquito population of cage 1. More than 90% of all GFP$^-$ mosquitoes were RFP$^+$CFP$^+$ for 12 generations (Supplementary Fig. 5b). To confirm that the rise of the GFP$^-$RFP$^+$CFP$^+$ phenotype reflected a parallel increase in the HEG allele we performed a PCR assay on randomly chosen mosquitoes to determine the presence of the *I-SceI* gene. The frequency of individuals positive for the HEG cassette rose from about 19% to 86% by generation 12 (Fig. 2c). Moreover, NotI digests of the PCR products showed that the frequency of individuals with chromosomes carrying both the HEG and the NotI marker, a combination that was absent at the beginning of the experiment, increased to 50% by generation 9 (Fig. 2c). The dynamics of both HEG$^+$ and NotI$^+$ allele frequencies matched expectations from stochastic simulations (Fig. 2c). We conclude that the rise in the frequency of GFP$^-$ individuals reflected the corresponding increase in the frequency of the HEG allele. The increase in the frequency of the NotI marker in the Donor allele pool indicates that homing is the cause for the observed rise in the frequency of HEG$^+$ individuals.

Our results demonstrate that homing can occur at appreciable frequencies in the germline of *A. gambiae* and therefore address a fundamental uncertainty that previously had been associated with proposals to use HEGs for pest control, namely whether HEGs would function in animals as they do in microbes. HEGs do not occur naturally in the nuclear genomes of metazoans; our results indicate that this absence is not because homing cannot occur, and instead supports alternative explanations such as that the segregated germline of animals prevents the horizontal transmission amongst species that these selfish genes need to persist over long evolutionary timespans[17]. The transmission dynamics of HEGs in cage populations provide the first evidence of the potential of these genetic elements to serve as synthetic gene drive systems in insect pests and add a promising candidate to those under development[18,19].

The sequence-specific activity of HEGs could be exploited to develop vector control strategies aimed at either disrupting the mosquito genes that contribute to its vectorial capacity or introducing at selected chromosomal locations novel genes that impair the mosquito's ability to function as vector for malaria[9]. Any use of HEGs in natural

A. gambiae populations will depend on the ability to re-engineer their specificity[20–23] towards native mosquito sequences. We identified in the *A. gambiae* genome within intergenic regions of the left (2L) and right arms (2R) of chromosome 2 two sequences that show similarities to the recognition sites of the two HEGs I-AniI and I-CreI that have previously been shown to be amenable to re-engineering to target novel human and plant sequences[23–29]. A previously described HEG engineering strategy was then used to generate an I-AniI variant to selectively cleave the 2L site, and a variant of monomerized I-CreI (termed mCre[30]) to cleave the 2R site selectively (Supplementary Fig. 6). The change in specificity of these enzymes demonstrates that HEGs can be designed to recognize new mosquito sequences and opens the possibility to investigate the biology of HEGs in wild-type mosquito populations. Although technical hurdles in HEG engineering technology must still be addressed to reach the flexibility required to target specific mosquito genes essential for viability or disease transmission, our results suggest how these genetic elements could overcome a major scientific roadblock in developing genetic control measures targeting species like the main vector of human malaria: the genetic manipulation of entire field populations starting from a few laboratory individuals.

METHODS SUMMARY

The generation of transgenic lines and population cage experiments are described in Methods. To monitor homing mosquitoes were subjected to fluorescent microscopy at the larval stage to detect the presence of the marker genes or subjected to PCR to detect the presence of the HEG gene at the adult stage.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 September 2010; accepted 16 February 2011.
Published online 20 April 2011.

1. Curtis, C. F. Possible use of translocations to fix desirable genes in insect pest populations. *Nature* **218,** 368–369 (1968).
2. Hamilton, W. D. Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science* **156,** 477–488 (1967).
3. Alphey, L. *et al.* Malaria control with genetically manipulated insect vectors. *Science* **298,** 119–121 (2002).
4. Corby-Harris, V. *et al.* Activation of *Akt* signaling reduces the prevalence and intensity of malaria parasite infection and lifespan in *Anopheles stephensi* mosquitoes. *PLoS Pathog.* **6,** e1001003 (2010).
5. Ito, J., Ghosh, A., Moreira, L. A., Wimmer, E. A. & Jacobs-Lorena, M. Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. *Nature* **417,** 452–455 (2002).
6. Moreira, L. A. *et al.* Bee venom phospholipase inhibits malaria parasite development in transgenic mosquitoes. *J. Biol. Chem.* **277,** 40839–40843 (2002).
7. Li, F., Patra, K. P. & Vinetz, J. M. An anti-chitinase malaria transmission-blocking single-chain antibody as an effector molecule for creating a *Plasmodium falciparum*-refractory mosquito. *J. Infect. Dis.* **192,** 878–887 (2005).
8. Sinkins, S. P. & Gould, F. Gene drive systems for insect disease vectors. *Nature Rev. Genet.* **7,** 427–435 (2006).
9. Burt, A. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc. R. Soc. Lond. B* **270,** 921–928 (2003).
10. Catteruccia, F., Benton, J. P. & Crisanti, A. An *Anopheles* transgenic sexing strain for vector control. *Nature Biotechnol.* **23,** 1414–1417 (2005).
11. Jacquier, A. & Dujon, B. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell* **41,** 383–394 (1985).
12. Bellaiche, Y., Mogila, V. & Perrimon, N. I-SceI endonuclease, a new tool for studying DNA double-strand break repair mechanisms in *Drosophila*. *Genetics* **152,** 1037–1044 (1999).
13. Windbichler, N. *et al.* Homing endonuclease mediated gene targeting in *Anopheles gambiae* cells and embryos. *Nucleic Acids Res.* **35,** 5922–5933 (2007).
14. Stoddard, B. L. Homing endonuclease structure and function. *Q. Rev. Biophys.* **38,** 49–95 (2005).
15. Goddard, M. R., Greig, D. & Burt, A. Outcrossed sex allows a selfish gene to invade yeast populations. *Proc. R. Soc. Lond. B* **268,** 2537–2542 (2001).
16. Meredith, J. M. *et al.* Site-specific integration and expression of an anti-malarial gene in transgenic *Anopheles gambiae* significantly reduces *Plasmodium* infections. *PLoS ONE* **6,** e14587 (2011).
17. Burt, A. & Koufopanou, V. Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Curr. Opin. Genet. Dev.* **14,** 609–615 (2004).
18. Chen, C. H. *et al.* A synthetic maternal-effect selfish genetic element drives population replacement in *Drosophila*. *Science* **316,** 597–600 (2007).
19. McMeniman, C. J. *et al.* Stable introduction of a life-shortening *Wolbachia* infection into the mosquito *Aedes aegypti*. *Science* **323,** 141–144 (2009).
20. Ashworth, J. *et al.* Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* **441,** 656–659 (2006).
21. Jarjour, J. *et al.* High-resolution profiling of homing endonuclease binding and catalytic specificity using yeast surface display. *Nucleic Acids Res.* **37,** 6871–6880 (2009).
22. Ashworth, J. *et al.* Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. *Nucleic Acids Res.* **38,** 5601–5608 (2010).
23. Thyme, S. B. *et al.* Exploitation of binding energy for catalysis and design. *Nature* **461,** 1300–1304 (2009).
24. Gao, H. *et al.* Heritable targeted mutagenesis in maize using a designed endonuclease. *Plant J.* **61,** 176–187 (2010).
25. Grizot, S. *et al.* Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. *Nucleic Acids Res.* **37,** 5405–5419 (2009).
26. Munoz, I. G. *et al.* Molecular basis of engineered meganuclease targeting of the endogenous human RAG1 locus. *Nucleic Acids Res.* **39,** 729–743 (2010).
27. Redondo, P. *et al.* Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. *Nature* **456,** 107–111 (2008).
28. Arnould, S. *et al.* Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J. Mol. Biol.* **371,** 49–65 (2007).
29. Rosen, L. E. *et al.* Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucleic Acids Res.* **34,** 4791–4800 (2006).
30. Li, H., Pellenz, S., Ulge, U., Stoddard, B. L. & Monnat, R. J. Jr. Generation of single-chain LAGLIDADG homing endonucleases from native homodimeric precursor proteins. *Nucleic Acids Res.* **37,** 1650–1662 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Ashburner, S. Russell, D. Huen and S. Chan for comments, assistance and for plasmids. We thank M. P. Calos for providing the pET11phiC31polyA plasmid. We thank M. J. Fraser Jr for providing the pBSII-IFP2-orf plasmid. We thank J. Meredith and P. Eggleston for providing the docking strain. We thank A. Hall, T. Nolan, K. Magnusson, D. Rogers and S. Fuchs for assistance. We thank S. Arshiya Quadri and M. Szeto for experimental support and the members of the laboratories of D. Baker, R. Monnat, A. Scharenberg and B. Stoddard for their collective support of HEG engineering. A. F. M. Hackmann provided graphics support. Funded by a grant from the Foundation for the National Institutes of Health through the Vector-Based Control of Transmission: Discovery Research (VCTR) program of the Grand Challenges in Global Health initiative and by NIH RL1 awards GM084433 to D.B. and CA133831 to R.J.M.

Author Contributions N.W. designed the experiments. N.W., M.M. and P.A.P. performed the experiments. N.W. and P.A.P. generated the transgenic lines. M.M. maintained mosquito populations. N.W. analysed the data. A.B. and N.W. generated the population dynamic models. A.C. and A.B. inspired the work and wrote the paper together with N.W. HEG redesign and target site cleavage analyses were performed by S.B.T., H.L., U.Y.U. (contributed equally) and B.T.H. with guidance from D.B. and R.J.M. All authors read and approved the final manuscript.

Author Information The plasmids pHome-T and pHome-D have been deposited to GenBank under the accession numbers HQ159398 and HQ159399. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.C. (a.drcrisanti@imperial.ac.uk).

METHODS

Development of transgenic lines. The pHome-R, pHome-T (GenBank HQ159398) and pHome-D (GenBank HQ159399) plasmids are derived from the same parent backbone and are identical apart from differences explained below. All three plasmids contain the DsRed fluorescent protein (RFP) reporter gene driven by the *Drosophila* Actin5C promoter as well as piggyBac inverted repeats for transposase mediated integration and an AttB sequence for site specific integration using the φC31 integrase. They also contain a 3×P3 (artificial promoter element binding 3 Pax-6 homodimers[31]) driven *GFP* marker gene that is modified the following way: the pHome-T construct contains the 18-base pairs I-SceI target site (The *A. gambiae* genome does not contain an I-SceI site) within the open reading frame of the *GFP* reporter gene (Supplementary Fig. 2b). The GFP coding sequence (CDS) containing the I-SceI recognition sequence remains functional but can be inactivated by cleavage followed by certain types of non-homologous end joining (NHEJ) DNA repair events (in particular when followed by a frameshift as the GFP N terminus is generally tolerant to amino acid changes) or homing. In addition the pHome-T construct carries a NotI recognition site immediately downstream of the *GFP* open reading frame (ORF) that is not present in the other two plasmids. The pHome-R construct also contains an I-SceI site within the *GFP* gene. In this case the *GFP* ORF of this plasmid is out of frame but cleavage and NHEJ followed by a frameshift is expected to restore GFP expression (Supplementary Fig. 1b). Finally the pHome-D construct contains the HEG expression cassette which consists of the *Anopheles β2-tubulin* promoter[10] and terminator flanking the *I-SceI* ORF, which contains an SV40 nuclear localization signal. This cassette is located within two I-SceI half sites disrupting the *GFP* gene at the same position as the I-SceI recognition site in the Reporter and Target vectors. This setup resembles the natural occurrence of HEGs and their target sites as both the HEG expression cassette and the HEG target site are flanked by identical homologous regions. Transgenic lines were developed as previously described[10,32]. Briefly, *A. gambiae* embryos of the φC31 integrase docking line[16] (or wild-type embryos) were injected using a Femtojet Express injector and sterile Femtotips (Eppendorf) with a mixture of 0.2 mg ml^{-1} of plasmid and 0.8 mg ml^{-1} of φC31 integrase or piggyback helper RNA, respectively. The 5′-capped helper RNAs were produced using the mMESSAGEmachine kit (Ambion) from linearized vectors pBSII-IFP2-orf (transposase) and pET11phiC31polyA (integrase). The hatched larval survivors were screened for transient expression of either the 3×P3-GFP if present or the Actin5C-RFP marker. In the presence of the 3×P3-GFP marker (pHome-T) only transients were grown up and crossed to wild-type mosquitoes whereas in the case of Actin5C-RFP (pHome-R, pHome-D) all survivors were crossed as this promoter drives no expression in the most posterior segments of the larvae where, due to the way embryos are injected, most or all of the transient fluorescence is usually observed. The progeny of these crosses were analysed for fluorescence to identify transgenic individuals. We have previously shown that founder effects and inbreeding can be determinants of the fitness of transgenic mosquitoes[33]. To minimize these effects the progeny of each transgenic founder was backcrossed to wild-type mosquitoes for at least three generations before homozygote strains were established. Transgenic mosquitoes at different developmental stages were analysed on a Nikon inverted microscope (Eclipse TE200) to detect GFP, RFP and CFP expression. Digital images were captured on a Nikon inverted microscope (Eclipse TE200) with an attached Nikon DXM1200 digital camera.

PCR and restriction analysis. PCRs (Phusion HF polymerase, Finnzymes) were performed on genomic DNA (Wizard Genomic DNA purification kit, Promega) prepared from single transgenic adult mosquitoes. We extracted DNA from single hemizygous virgin female or male mosquitoes in the mating experiments. Each hemizygous offspring analysed allows the scoring of a single chromosome from the double transgenic parent. We extracted DNA from single virgin female adults in the case of the population cage experiments and the PCR assay was performed on 60 to 96 randomly chosen individuals per generation. The following primers as shown in Supplementary Fig. 1 were used:

Primer set 1a, forward 5′-TGGAAATGAGAAGTAGGTGCATCTGCA-3′, reverse 5′-GGAATAAGGGCGACACGGAAATGTTG-3′; primer set 1b, forward 5′-TGTGACAGTGGAAATGAGAAGTAGGTGC-3′, reverse 5′-TCTCAACGTAGTCCACAAAGCATCAA-3′; primer set 2, forward 5′-GCGATGACGAGCTTGTTGGTG-3′, reverse 5′-CGTGCACAGGCTTTGATAACTCCT-3′; primer set 3, forward 5′-CTCTCCGCTCTCAAGTCGCGTTCA-3′, reverse 5′-TGCAGATGCACCTACTTCTCATTTCCA-3′; primer set 4, forward 5′-ATCGCTGAGATAGGTGCCTCACTGA-3′, reverse 5′-CTCATGTAACAGTTCATAGTTCTCGC-3′.

In vitro digestions using NotI (Roche) and I-SceI (NEB) were performed according to manufacturer's recommendations. We used primer set 1a on hemizygous progeny of Donor/Target and Donor/Reporter crosses. Primer set 1b was used in Donor/Ectopic Target crosses and all population cage experiments.

Population dynamic modelling. To model the cage populations we assumed that the two starting alleles, Target (T; GFP$^+$, HEG$^-$, NotI$^+$), and Donor (D; GFP$^-$, HEG$^+$, NotI$^-$), give rise to three classes of novel alleles: (1) D^N, products of homing that retain the NotI site and can themselves act as donors (GFP$^-$, HEG$^+$, NotI$^+$); (2) M^+, products of misrepair (for example, homologous repair from a different template, or non-homologous end-joining) that are GFP$^+$ (that is, GFP$^+$, HEG$^-$, NotI$^{+/-}$); and (3) M^-, products of misrepair that are GFP$^-$ (GFP$^-$, HEG$^-$, NotI$^{+/-}$). All three novel alleles are resistant to further cleavage, and the two products of misrepair are not able to home. In the germlines of male D/T or D^N/T mosquitoes the T allele is cleaved with probability c. In D^N/T males these cleaved alleles are then converted into a D^N allele with probability h, into an M^+ allele with probability $(1-h)r$, or into an M^- allele with probability $(1-h)(1-r)$, where h is the rate of canonical homologous repair in males and r is the probability that other forms of repair maintain GFP expression. In D/T males, cleaved T alleles are converted into D^N alleles with probability hn and into D alleles with probability $h(1-n)$, where n is the probability that new products of homing retain the NotI site, and $1-n$ is the probability that it is lost by co-conversion. The overall net homing rate in this model is $e_m = ch$.

Estimates of c, h, n and r were derived from the Donor/Target experiments as follows.

(1) The fraction of GFP$^+$ progeny from D/T males is 0.14. PCR and sequencing of 36 such individuals showed that 7 had the intact Target sequence (19.4%). The proportion of gametes with uncleaved T alleles is therefore $0.14 \times 0.194 = 0.0272$. Because only half the gametes should carry the T allele or its descendants (the other half being derived from the Donor allele), the cleavage rate in males is $c = 1 - 2 \times 0.0272 = 0.95$ (that is, 95%).

(2) The fraction of progeny from D/T males that are GFP$^-$RFP$^+$CFP$^+$ is 0.803. PCR analysis of 156 such individuals showed that 152 of them were HEG$^+$ (97.4%). The proportion of gametes that are HEG$^+$ is thus $0.803 \times 0.974 = 0.782$. Because 50% of gametes are expected to be HEG$^+$ in the absence of homing, the estimated homing rate is $e = 2(0.782 - 0.5) = 0.564$. Because the homing rate is equal to the product of the cleavage rate and the rate of canonical homologous repair, the latter is estimated to be $h = 0.564/0.95 = 0.60$ (that is, 60%).

(3) Of the 152 HEG$^+$ progeny described above, 25 were also NotI$^+$. Therefore, the overall fraction of HEG$^+$NotI$^+$ progeny is $0.803 \times 25/156 = 0.129$. The fraction of gametes with a newly acquired HEG is $e/2 = 0.281$. Therefore, the probability that homing leads to retention of the NotI site is $n = 0.129/0.282 = 0.45$.

(4) The fraction of progeny from D/T males that were GFP$^+$ and that molecular analysis showed had been cleaved and misrepaired was 0.112. The total fraction of gametes with misrepaired alleles (that is, cleaved and not subject to canonical homing) is $c(1-h)/2 = 0.19$. Therefore the probability that misrepaired alleles remain GFP$^+$ is $r = 0.112/0.19 = 0.58$.

Populations start as a mixture of T/T and D/D homozygotes. T/D heterozygous males produce sperm carrying alleles T, D, M^-, M^+ or D^{N} with probabilities $(1 \times c)/2$, $1/2 + ch(1 \times m)/2$, $c(1-h)(1-r)/2$, $c(1-h)r/2$, and $chm/2$, respectively, and T/D^{N} heterozygous males produce sperm carrying these alleles with probabilities $(1-c)/2$, 0, $c(1-h)(1-r)/2$, $c(1-h)r/2$, and $1/2 + ch/2$, respectively. All other male genotypes and all female genotypes produce gametes in Mendelian proportions. All genotypes have equal survival and fertility: each female mates with a single male, chosen randomly (with replacement), and each offspring is from a randomly chosen mated female (with replacement). Simulations were generated in the Mathematica software suite 7 (Wolfram Research).

Defined crosses and population cage experiments. Crosses were carried out using 25 males and 25 virgin females. A total of 1,996 (Donor/Target), 3,582 (Donor/Reporter) and 720 (Donor/Ectopic Target) offspring were analysed in at least three independent experiments for green (GFP), blue (CFP) and red (RFP) fluorescence. Independently reared cage populations were established and the I-SceI containing allele was seeded at a frequency of 10% or 50%. To achieve a frequency of 10% the cage (BugDorm-1, Megaview) contained a population of 540 homozygote mosquitoes carrying the I-SceI Target construct (270 males and 270 females). In addition the population cage also contained 60 homozygote mosquitoes carrying the I-SceI Donor construct (30 males and 30 females). Each generation mosquitoes were allowed to mate for 5–7 days, and then fed on 2–3 mice to ensure that all females were blood-fed. Larvae were allowed to hatch from the eggs and reared until the L3–L4 stage, at which point a random set of at least 300 was screened for the presence of the fluorescent markers. At the pupal stage mosquitoes were separated according to sex and males and females were allowed to emerge separately for each cage population. After at least 48 h, 300 male and 300 female mosquitoes of each population were added to a fresh cage to establish the next generation.

Identification and cloning of *Anopheles gambiae* genomic target sites. Genomic targets for engineered I-AniI and mCreI protein variants were identified by searching the *Anopheles gambiae* genome with PSSMs (positional specific scoring or search matrices) constructed for each HEG protein. The I-AniI PSSM was constructed from cleavage degeneracy data, computational design results and selected variants isolated by using a modified bacterial selection system[34–36] (additional unpublished results). The mCreI PSSM was constructed from cleavage degeneracy data and the results of a comprehensive computational design analysis of I-CreI design data[37,38] (additional unpublished results). A predicted cleavage-sensitive *Anopheles gambiae* chromosomal target site for an engineered I-AniI protein that combined two unpublished protein variants was identified on chromosome 2L (Agam 2L −3C/+5G site: reverse strand nucleotides 26449203–26449184). A comparable, predicted cleavage-sensitive chromosomal target site for a −5C mCreI design was identified on chromosome 2R (Agam 2R −5C2 site: forward strand nucleotides 33439283–33439302). Both target sites, together with 15-bp of flanking *Anopheles* chromosomal sequence on each side, were synthesized as pairs of complementary oligonucleotides that were annealed and ligated into the NheI/SacII sites of the bacterial plasmid vector *pCcdB* to facilitate cleavage analyses[36]. A native I-AniI site previously cloned into pBluescript[35] and a native I-CreI target site cloned into *pCcdB* were used as positive control sites.

Engineering of *Anopheles* target site-specific I-AniI and mCreI variants. An engineered I-AniI variant predicted to cleave the Agam 2L −3C/+5G chromosomal target site was generated by combining two previous engineered variants for the −3C and +5G base pair positions in the M4 variant of I-AniI (previously described Y2 variant + two additional residue substitutions, F91I and S92T)[35,39]. The −3C variant was based on a previously published −3C computational design[35] that was further improved by bacterial selection[36]. The residue substitutions in this I-AniI variant were Y18W, E35K, and substitution of the four residue loop PDGM for the native 7-residue loop between I-AniI positions K60 and M66. The +5G variant was identified in a bacterial selection and contained a D168Q substitution. These modifications were combined and incorporated into the open reading frame of the M4 Y2 variant of I-AniI[35,39].

Variants of mCreI specific for the *Anopheles* 2R -5C2 mCreI chromosomal target site were generated using RosettaDesign (RD), a macromolecular modelling and design suite[40]. In brief, the −5G>C base change in the *Anopheles* 2R mCreI target site was modelled, and amino acid residues in close proximity to the −5 position were allowed to mutate *in silico* to accommodate the new −5C design target base pair. Amino acid conformations and associated hydration patterns that improved the energy of mCreI −5C target site complex were accepted more often during design runs, and converged to identify energetically favourable amino acid substitutions predicted to be specific for the −5C design target. The resulting residue substitutions, I24K and R68T, were incorporated into the open reading frame of mCreI by PCR-mediated mutagenesis[38]. The resulting engineered I-AniI and mCreI variant proteins and native control proteins were expressed in *E. coli* and purified as previously described for *in vitro* cleavage analyses[35,38,41]. Bacterial selection[36] has already been used to improve Y2 I-AniI and to generate mCreI[30,35,39]. Thus it should be possible to use sequential positive and negative bacterial selection to rapidly improve further the cleavage efficiency of both engineered proteins on their *Anopheles* chromosomal target sites, and suppress residual cleavage activity on their native target sites if required.

***In vitro* cleavage assays.** pCcdB vector DNA containing the *Anopheles* 2L −3C/+5G I-AniI chromosomal target site was linearized with XbaI, and pBluescript plasmid DNA containing the native I-AniI target site with ScaI, before the cleavage analyses. Cleavage reactions (10 µl final volume) were performed in digest buffer (170 mM KCl, 10 mM $MgCl_2$, 20 mM Tris pH 9.0) containing 5 nmol linearized target site plasmid and serial twofold dilutions of purified enzyme ranging from 800 nM to 12.5 nM. Reactions were incubated at 37 °C for 0.5 h, then stopped by the addition of stop buffer (200 mM EDTA, 30% glycerol, bromophenol blue) before agarose gel electrophoresis to separate substrate and cleavage products.

pCcdB vector DNA containing the *Anopheles* 2R −5C2 mCreI chromosomal target site was linearized by NcoI digestion before cleavage analyses. Cleavage reactions (10 µl final volume) were performed in digest buffer (10 mM $MgCl_{2,}$ 20 mM Tris pH 8.0) containing 10 nmol linearized target site plasmid and serial twofold dilutions of purified enzyme ranging from 320 nM to 10 nM. Reactions were incubated at 37 °C for 1 h, then stopped by the addition of stop buffer (0.5% SDS and bromophenol blue) before agarose gel electrophoresis.

31. Sheng, G., Thouvenot, E., Schmucker, D., Wilson, D. S. & Desplan, C. Direct regulation of *rhodopsin 1* by *Pax-6/eyeless* in *Drosophila*: evidence for a conserved function in photoreceptors. *Genes Dev.* **11,** 1122–1131 (1997).
32. Lobo, N. F., Clayton, J. R., Fraser, M. J., Kafatos, F. C. & Collins, F. H. High efficiency germ-line transformation of mosquitoes. *Nature Protocols* **1,** 1312–1317 (2006).
33. Catteruccia, F., Godfray, H. C. & Crisanti, A. Impact of genetic manipulation on the fitness of *Anopheles stephensi* mosquitoes. *Science* **299,** 1225–1227 (2003).
34. Scalley-Kim, M., McConnell-Smith, A. & Stoddard, B. L. Coevolution of a homing endonuclease and its host target sequence. *J. Mol. Biol.* **372,** 1305–1319 (2007).
35. Thyme, S. B. *et al.* Exploitation of binding energy for catalysis and design. *Nature* **461,** 1300–1304 (2009).
36. Doyon, J. B., Pattanayak, V., Meyer, C. B. & Liu, D. R. Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *J. Am. Chem. Soc.* **128,** 2477–2484 (2006).
37. Argast, G. M., Stephens, K. M., Emond, M. J. & Monnat, R. J. Jr. I-*Ppo*I and I-*Cre*I homing site sequence degeneracy determined by random mutagenesis and sequential *in vitro* enrichment. *J. Mol. Biol.* **280,** 345–353 (1998).
38. Ulge, U. Y., Baker, D. A. & Monnat, R. J. Comprehensive computational design of mCreI homing endonuclease cleavage specificity for genome engineering. *Nucleic Acids Res.* doi:10.1093/nar/gkr022 (1 February 2011).
39. McConnell Smith, A. *et al.* Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. *Proc. Natl Acad. Sci. USA* **106,** 5099–5104 (2009).
40. Das, R. & Baker, D. Macromolecular modeling with Rosetta. *Annu. Rev. Biochem.* **77,** 363–382 (2008).
41. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 (2005).