

The Population Genetics of Using Homing Endonuclease Genes in Vector and Pest Management

Anne Deredec,* Austin Burt* and H. C. J. Godfray^{†,1}

*NERC Centre for Population Biology, Department of Biology, Imperial College London, Ascot, Berks SL5 7PY, United Kingdom and [†]Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom

Manuscript received March 7, 2008
Accepted for publication April 28, 2008

ABSTRACT

Homing endonuclease genes (HEGs) encode proteins that in the heterozygous state cause double-strand breaks in the homologous chromosome at the precise position opposite the HEG. If the double-strand break is repaired using the homologous chromosome, the HEG becomes homozygous, and this represents a powerful genetic drive mechanism that might be used as a tool in managing vector or pest populations. HEGs may be used to decrease population fitness to drive down population densities (possibly causing local extinction) or, in disease vectors, to knock out a gene required for pathogen transmission. The relative advantages of HEGs that target viability or fecundity, that are active in one sex or both, and whose target is expressed before or after homing are explored. The conditions under which escape mutants arise are also analyzed. A different strategy is to place HEGs on the Y chromosome that cause one or more breaks on the X chromosome and so disrupt sex ratio. This strategy can cause severe sex-ratio biases with efficiencies that depend on the details of sperm competition and zygote mortality. This strategy is probably less susceptible to escape mutants, especially when multiple X shredders are used.

THE possibility of controlling man's major pests, pathogens, and disease vectors using genetic manipulation has long been discussed (HAMILTON 1967; CURTIS 1968) and is of great current interest (TURELLI and HOFFMANN 1999; ALPHEY *et al.* 2002; JAMES 2005; SINKINS and GOULD 2006). A broad spectrum of possible strategies has been explored. Organisms can be manipulated to be conditionally sterile or lethal and released into the environment to disrupt mating or to reduce the fecundity of the wild population (THOMAS *et al.* 2000; ATKINSON *et al.* 2007; PHUC *et al.* 2007). With these inundative techniques the manipulated construct is not required to persist in the environment. A different approach is to introduce a beneficial genetic construct into a wild population with a drive mechanism that causes it to increase in frequency. The construct might impose a fitness load on the population, reducing its density or causing it to go extinct. Alternatively, it may alter the phenotype of the organism with no or minor changes to its fitness. The latter is of particular relevance to disease vectors where it may be possible to reduce or eliminate transmission. Recent advances in molecular genetics have demonstrated that knocking out certain Anopheles mosquito genes, or inserting new constructs, prevents the insect from transmitting Plasmodium, the malaria pathogen (ITO

et al. 2002; MOREIRA *et al.* 2002), while RNAi techniques have been used to prevent Aedes mosquitoes from transmitting the dengue virus (FRANZ *et al.* 2006). Enthusiasm for these control strategies is tempered by the realization that any method involving genetic manipulation will require the highest scrutiny and investigation prior to implementation and that support from the public will be essential for any project to go ahead (ALPHEY *et al.* 2002; JAMES 2005; KNOLS *et al.* 2006).

A variety of different mechanisms for driving genes through a population have been considered, most of them based on elements with non-Mendelian inheritance that have been discovered in nature (BURT and TRIVERS 2006). Some genes cause the chromosomes on which they reside to be overrepresented in the gamete pool and thus could be used to increase the frequency of an introduced linked gene (BURT and TRIVERS 2006). Genetic constructs can be designed that show underdominance—heterozygote inferiority—and hence will increase in frequency once their abundance passes a certain threshold (DAVIS *et al.* 2001; MAGORI and GOULD 2006). Elements that jump between chromosomes can be used as vectors for beneficial constructs, and transposable elements in particular have received a lot of attention (COATES *et al.* 1998). Heterozygote females carrying *medea* elements modify their eggs such that they survive only if they carry the *medea* gene or are fertilized by sperm that carry the element. This disadvantages wild-type alleles and allows *medea* to spread

¹Corresponding author: Department of Zoology, University of Oxford, S. Parks Rd., Oxford OX1 3PS, United Kingdom. E-mail: charles.godfray@zoo.ox.ac.uk

(WADE and BEEMAN 1994). Artificially engineered *medea* elements have recently been developed and offer an important new potential drive mechanism (CHEN *et al.* 2007). Certain symbiotic microorganisms with vertical inheritance spread by manipulating host reproduction such that infected individuals produce more daughters than uninfected individuals. Introducing the beneficial gene into the symbiont could then lead to its spread, though with the disadvantage that the gene may not be expressed in the correct tissue. The intracellular bacterium *Wolbachia* that is present in a very large fraction of insects and that spreads through cytoplasmic incompatibility (noninfected females are at a disadvantage because they cannot use the sperm of infected males) is the most important candidate drive mechanism of this type (WERREN 1997; TURELLI and HOFFMANN 1999). Finally, a variant of these techniques is to use the drive mechanism to impose a fitness cost on the organism and then to link the beneficial gene to a construct that mitigates the cost and hence is selected to spread (SINKINS and GODFRAY 2004). In comparing drive mechanisms the most important factors likely to influence success or acceptability include the evolutionary stability of the construct, the degree to which the within- and between-individual spread of the element can be predicted, whether the construct can increase in frequency from rare or if a threshold frequency must be exceeded before spread occurs, and whether it is possible to reverse the manipulation.

An exciting potential drive strategy is to use site-specific selfish genes such as homing endonuclease genes (HEGs) (BURT 2003). A HEG codes for a protein that recognizes and cuts DNA containing a specific 20- to 30-bp sequence (STODDARD 2005). Critically this sequence is found only on chromosomes not containing the HEG and at the precise location where the HEG occurs. After a double-strand break in a heterozygote, the cell's recombinational repair mechanism uses the chromosome carrying the HEG as a template and the HEG is thus copied from one chromosome to the other, converting a heterozygote to a homozygote. If there are no fitness costs to the HEG, it spreads until it reaches fixation. Other elements such as group II introns and certain LINE-like transposable elements have similar strategies for spread, though with a more complicated mechanism involving RNA intermediaries (BURT and TRIVERS 2006). Below we concentrate just on HEGs that offer the most straightforward site-specific selfish genes for exploitation.

HEGs are found in nature in single-celled fungi, plants, protists, and bacteria, but not in higher animals. They tend to reside in noncoding regions (especially introns) and so have little effect on fitness because they are spliced out prior to translation into protein. Due to their low fitness costs they are expected to spread to fixation, but then decay because once they are fixed there is no selection for their maintenance. Compara-

tive studies have shown that HEGs probably survive by jumping from species to species and that maintenance requires that the rate of species jumps must exceed HEG "death" in a lineage (GODDARD and BURT 1999; BURT and KOUFOPANOU 2004). It is likely that it is easier for HEGs to jump among single-cell organisms than among animals with segregated germlines, which may explain their absence from the latter.

The aim of this article is to describe the different ways in which HEGs might be used as part of a genetic control strategy and to develop and analyze the population genetic models that will be required to assess their relative advantages and disadvantages. It builds on the analyses of BURT (2003), who derived the equilibrium frequency and genetic load of HEGs with different homing frequencies that were either lethal or sterile to one or both sexes. He also discussed alternative strategies such as the use of multiple HEGs and their use as "X chromosome shredders" that are analyzed formally here for the first time. We also study the population genetics of mutations that might nullify the action of the HEG.

We first treat "classical" HEGs that spread by copying themselves into homologous chromosomes after double-strand breaks. We derive equations for (i) the spread and equilibria of HEGs that are active after gene expression and (ii) HEGs that are active before. We then explore (iii) the possible advantages of sex-specific expression and (iv) the risk of mutations arising that prevent HEG spread. Second, we study HEGs on the Y chromosome that cause X chromosome breaks—X shredders. We (i) derive equilibrium sex ratios for different numbers of shredders, (ii) analyze the effects of reduced sperm number and competition for zygotes, and (iii) study the evolution of escape mutants.

THEORETICAL RESULTS: DRIVING HEGs

HEG active after gene expression: Consider an engineered HEG that is introduced into a chromosome opposite a functional gene. Let the homing rate (the probability of a successful gene conversion) be e and the fitness costs of disrupting gene function be s for the homozygote and sh for the heterozygote. We begin by assuming fitness costs are equal in males and females and that homing occurs at meiosis after gene expression (so any costs of being homozygote are not experienced by the individual in which homing occurs). If q and p are the gametic frequencies of the HEG and wild-type alleles, respectively, the recurrence for q is

$$q' = \frac{(1-s)q^2 + (1-sh)pq(1+e)}{1-sq^2 - 2shpq}. \quad (1)$$

The equilibrium frequency for the HEG, q^* , is

$$q^* = 0; \quad s > \frac{e}{1-h+eh} \quad \text{and} \quad s > \frac{e}{h(1+e)}, \quad (2)$$

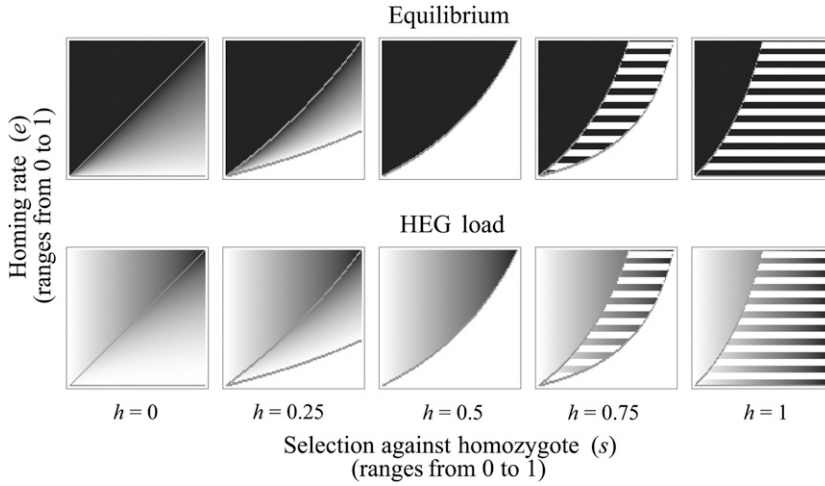


FIGURE 1.—Equilibrium frequency and load of a HEG that is active (homes) after gene expression. Top row: equilibrium HEG frequency as a function of homing rate and fitness costs. The gene is fixed in the solid region and lost in the open region. Where an interior equilibrium is possible the darkness of the shading is proportional to the equilibrium frequency. In the striped region the gene is either fixed or lost depending on its initial frequency. Bottom row: load imposed by the HEG in the same regions of parameter space as in the top row. The darkness of the shading is proportional to the HEG load.

$$q^* = 1; \quad s < \frac{e}{1-h+eh} \quad \text{and} \quad s < \frac{e}{h(1+e)}. \quad (3)$$

When these inequalities do not hold there is an interior equilibrium

$$q^* = \frac{e - (1+e)hs}{s(1-2h)}, \quad (4)$$

which is stable if

$$\frac{e}{1-h+eh} < s < \frac{e}{h(1+e)} \quad (5)$$

and unstable if

$$\frac{e}{1-h+eh} > s > \frac{e}{h(1+e)}. \quad (6)$$

The latter implies low heterozygote fitness, $h > \frac{1}{2}$, in which case the HEG either goes extinct or reaches fixation depending on whether the initial frequency is less than or greater than the unstable equilibrium (Figure 1).

We define the “HEG load” (L) to be the relative reduction in the growth rate of the population in the presence of the HEG. Assume a population with discrete generations and let λ_0 be the rate at which the population increases in the absence of any density-dependent effects (equivalent here to per capita female fecundity). Then define $L = 1 - \lambda'_0/\lambda_0$, where λ'_0 is the population growth rate when the HEG is present. HEG load is thus a quantity very similar to genetic load as usually interpreted in classical population genetics, except that it does not include effects on males and can include processes that bias the sex ratio (see below). In using HEGs to drive down vector and pest densities it is important to note that a HEG load of L does not necessarily mean that the population density is reduced by a factor L . The observed reduction will depend on the precise form of density dependence operating in the

population, as well as on the relative ordering of homing, target gene expression, and density dependence in the life cycle. This article is concerned only with genetic dynamics and so we cannot predict absolute population reductions. Nevertheless, calculating L provides a useful comparative measure of potential population reductions. In this case, the HEG load experienced by the population at equilibrium is

$$L = \frac{e^2(1-hs)^2 - s^2h^2}{s(1-2h)} \quad (7)$$

for $0 < q^* < 1$, no load when $q^* = 0$, and $L = s$ for $q^* = 1$.

Consider first the special case in which HEGs are employed to knock out an essential gene with the aim of maximizing HEG load and driving a population to extinction. For a fully recessive homozygote lethal ($s = 1$, $h = 0$) the equilibrium HEG frequency is e and the load e^2 . Thus very substantial loads, and hence potential reductions in population numbers, are possible as homing frequencies approach unity. However, the highest equilibrium HEG loads do not occur when the HEG is invariably lethal in the homozygote. For a given value of the homing rate (e), the greatest load occurs when the fitness costs are at the maximum that still allows the HEG to become fixed ($s = e$), in which case the load equals the selection pressure ($L = s$) (Figure 1).

If we assume that the heterozygote also has reduced fitness ($h > 0$), then a greater range of equilibrium behaviors may be observed (Figure 1). The HEG is always fixed when fitness costs (s) are low and homing rates (e) are high, but away from this region of parameter space decreasing heterozygote fitness first sees a reduction in the parameter combinations where the HEG equilibrium frequency is less than one but greater than zero (an internal equilibrium). Then, when heterozygote fitness is closer to the homozygote than to the wild type ($h > \frac{1}{2}$), a region of bistability appears (the HEG goes to extinction) or fixation depending on

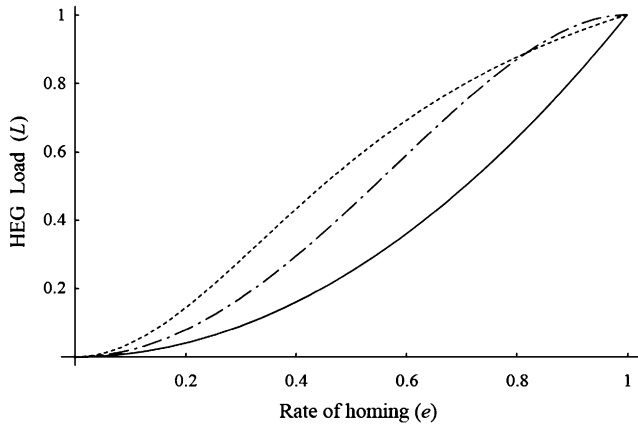


FIGURE 2.—HEG load (L) as a function of homing frequency (e) when the target gene is essential and the wild type is fully dominant ($s = 1$, $h = 0$). We plot three cases: in the first two the effects of the HEG are experienced equally by both sexes, the homozygote either being lethal (solid line) or having no postmating fertility through either sperm or ova (dotted-and-dashed line). The third case is a female-specific HEG where effects on either viability or postmating fertility lead to the same load (dotted line).

its initial frequency) that increases as the HEG becomes fully dominant. In the absence of stochastic effects, for most parameter combinations a HEG with $h > \frac{1}{2}$ will not spread from rare. As before, for fixed e , load is generally maximized at the highest value of s that allows fixation.

A slightly different strategy is to engineer a HEG to target a gene that is required for reproduction rather than survival. In the simplest case, if the knockout prevents the individual from participating in mating, then the dynamics and HEG load are exactly the same as described above. But suppose the knockout acts later, such that mating occurs normally but is less productive (for example, the male makes defective sperm that fertilize the eggs normally but result in inviable progeny), so that any matings involving either a male or a female carry the HEG lead to fewer offspring (a postmating fertility effect). Then while the dynamics of spread and equilibrium would still be as described above, the genetic load would be greater. Only those matings not involving an infertile carrier of the HEG, a fraction $(1 - q^2s - 2pqhs)^2$, would produce offspring and hence the genetic load would be $1 - (1 - q^2s - 2pqhs)^2$. The load is thus always greater than or equal to the equivalent load ($q^2s + 2pqhs$) for a HEG targeting survival. If the knockout mutation is recessive and abolishes reproductive success completely ($s = 1$, $h = 0$), the HEG equilibrium frequency is $q = e$ and the load is $L = e^2(2 - e^2) > e^2$ (Figure 2).

The second special case is when a HEG is employed to knock out a gene required for an insect to vector a pathogen. Ideally the gene would have no fitness costs to the host ($s = 0$), in which case it would always spread and cause no HEG load. A HEG whose fitness effects are

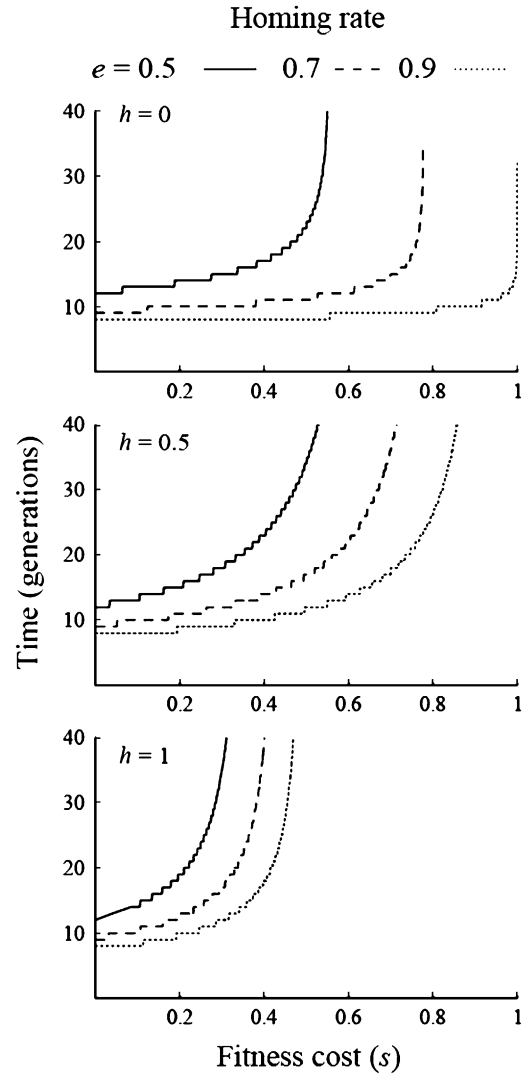


FIGURE 3.—The number of generations taken for a HEG to increase in frequency from 0.05 to 0.9 as a function of fitness costs (s), homing frequency (e), and dominance (h).

manifest only in homozygotes becomes fixed provided $e > s$ and causes a load of $L = s$ when it affects survival or fecundity. When fertility is affected and determined after mating by the genotype of both partners, then $L = s(2 - s)$.

Were a HEG to be used in a vector or pest control program, not only the ultimate outcome but also the rate at which it is attained would be significant. In Figure 3 we plot the number of generations that it takes for a HEG to increase in frequency from 0.05 to 0.9. For those HEGs that can reach fixation, spread is faster for high homing rates and for genes with recessive fitness costs. For much of this parameter space rapid spread occurs within 10–15 generations, which for many insect species is just a couple of years and so is highly relevant to pest and vector control on relatively short timescales.

HEG active before gene expression: We now assume that homing and gene conversion occur prior to the expression of the gene containing the HEG recognition

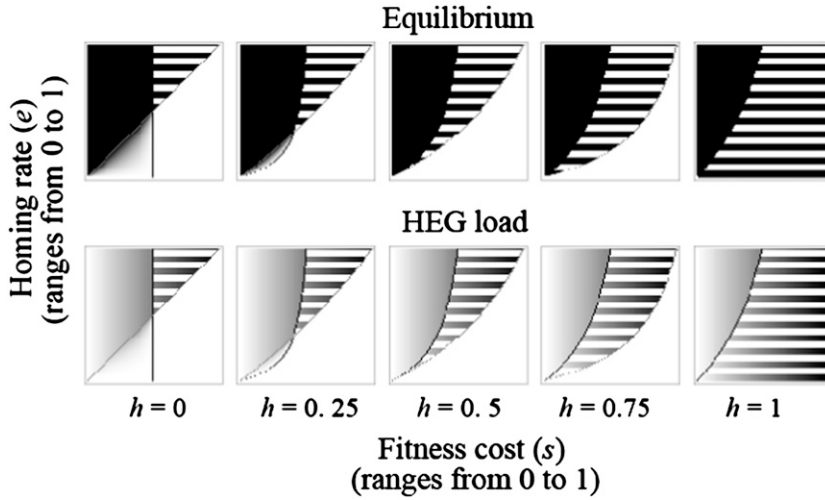


FIGURE 4.—Equilibrium frequency and load of a HEG that is active (homes) before gene expression. Drawing conventions and parameters are the same as in Figure 1.

sequence. Any fitness consequences of disrupting the gene are now experienced both by homozygotes and by the “transformed” heterozygotes. The recurrence for HEG frequency is now

$$q' = \frac{(1-s)(q^2 + 2pqe) + (1-sh)pq(1-e)}{1 - s(q^2 + 2pqe) - 2pqsh(1-e)}. \quad (8)$$

The equilibrium frequency for the HEG, q^* , is

$$q^* = 0; \quad s > \frac{e}{1-h+eh} \quad \text{and} \quad s > \frac{e}{2e+h-eh}, \quad (9)$$

$$q^* = 1; \quad s < \frac{e}{1-h+eh} \quad \text{and} \quad s < \frac{e}{2e+h-eh}. \quad (10)$$

When these inequalities do not hold, there is an interior equilibrium

$$q^* = \frac{e(1-2s) - hs(1-e)}{s(1-2e-2h(1-e))}, \quad (11)$$

which is stable if

$$\frac{e}{1-h+eh} < s < \frac{e}{2e+h-eh} \quad (12)$$

and unstable if

$$\frac{e}{1-h+eh} > s > \frac{e}{2e+h-eh}. \quad (13)$$

In the last case the HEG either goes extinct or reaches fixation depending on whether the initial frequency is less than or greater than the unstable equilibrium (Figure 4).

The HEG load is s when the HEG becomes fixed and

$$L = \frac{(e - e(2-h)s - hs)(e + (1-e)hs)}{s(1-2e(1-h) - 2h)} \quad (14)$$

for the interior equilibrium.

A HEG targeting a fully recessive gene with a significant effect on fitness ($h=0$, $s > \frac{1}{2}$) can invade only if $e > s$ and if the initial HEG frequency exceeds a threshold of $e(2s-1)/s(2e-1)$. Thus the strategy of using HEGs to create a recessive lethal ($s=1$, $h=0$) will not work if homing occurs prior to gene expression. The benefits of gene conversion in the heterozygote are nullified by the fitness costs of creating a homozygote.

In the limit, when there are no costs to carrying a HEG, it makes no difference when homing occurs and the HEG will always spread. For moderate fitness costs, the condition for the fixation of the HEG is the same irrespective of the order of homing and expression. However, in the regions where both fixation and extinction of the HEG can occur depending on initial conditions, fixation now requires higher gene frequencies compared to the case where the HEG is active after gene expression. Where the HEG is not fixed, its equilibrium frequency and load are always lower when homing occurs prior to expression, and the rate of spread of the gene is also relatively slower (data not shown).

Sex-specific expression: Returning to our original model of the HEG being active after gene expression, we now assume that it targets a gene that has different effects on males and females and/or that the HEG has different rates of homing in the two sexes. Let q_x , e_x , s_x and h_x have the same meanings as before except now we assume their values may be different in males ($x=m$) or females ($x=f$). The dynamics are given by the coupled recurrence equations

$$q'_x = \frac{(1-s_x)q_x q_{x'} + (1-s_x h_x)(1/2)(p_x q_{x'} + p_{x'} q_x)(1+e_x)}{1 - s_x q_x q_{x'} - s_x h_x (p_x q_{x'} + p_{x'} q_x)}, \quad (15)$$

$$\{x, x'\} = \{m, f\} \quad \text{and} \quad \{f, m\}.$$

The general solution to these equations is too complex to be helpful and we focus on a few specific cases. First, assume the knockouts are recessive and affect only females ($h_f=0$, $s_f > 0$, $s_m=0$). Then

$$q_m^* = \frac{(e_f + e_m)(1 + e_m)}{s_f(1 - e_m^2) + 2e_m(e_f + e_m)}, \quad q_f^* = \frac{e_f + e_m}{s_f(1 + e_m)}; \\ \frac{e_f + e_m}{1 + e_m} < s_f, \quad (16)$$

$$q_m^* = q_f^* = 1, \quad \frac{e_f + e_m}{1 + e_m} > s_f. \quad (17)$$

The HEG load is $s_f q_m q_f$ or

$$L = \frac{(e_f + e_m)^2}{s_f(1 - e_m^2) + 2e_m(e_f + e_m)} \quad (18)$$

for $q_m^*, q_f^* < 1$ and $L = s_f$ otherwise. If we assume homing frequency is the same in the two sexes ($e_f = e_m = e$) and the knockout is a female-specific recessive lethal or sterile ($s_f = 1$), then the load is $L = 4e^2/(3e^2 + 1)$ (Figure 2). This load is always greater than that when both sexes are killed or removed from the mating pool, for which the load is $L = e^2$. Killing males is counterproductive because it reduces the frequency of the HEG without reducing population productivity. We can also compare the load caused by a female-specific lethal or sterile with that of a HEG that disrupts both male and female fertility so that only zygotes produced by parents neither of whom are homozygous HEG carriers survive ($L = e^2(2 - e^2)$). The female-specific HEG is superior unless homing rates (e) are large (Figure 2). As before, the maximum HEG load, for a given homing rate, occurs for the highest homozygote fitness cost at which the HEG can become fixed. Here this is found when $s_f = 2e/(1 + e)$ in which case $L = s_f$. Larger loads, for the same homing rate, are thus possible for sex-specific fitness effects. The rate of spread of sex-specific HEGs is similar to that of nonspecific genes.

Second, assume that the fitness effects of the HEG are the same for both males and females, but that homing rates are different ($s_f = s_m$, $h_f = h_m = 0$, $e_f \neq e_m$). Recall that when the homing rate is the same in the two sexes, fixation requires $s < e$: in the present case s must be less than the average rate of homing in the two sexes. If $s_f = s_m = 1$, then fixation cannot occur and the load is $L = \frac{1}{2}(1 + e_f e_m - \sqrt{(1 - e_f^2)(1 - e_m^2)})$. If the average homing rate is kept constant, the load is at a minimum when rates are the same in the two sexes and increases as the difference gets larger.

Third, consider the case where both homing and costs are sex specific. If the HEG homes only in females and is a female-specific lethal or sterile ($e_m = 0$, $e_f > 0$, $s_m = 0$, $s_f = 1$), then the loads produced are identical to the non-sex-specific case. But if homing is restricted to males rather than females, then a female lethal or sterile HEG ($e_m > 0$, $e_f = 0$, $s_m = 0$, $s_f = 1$) causes substantially lower loads to occur ($L = e_m^2/(1 + e_m^2)$), with the maximum obtainable load (as homing frequency approaches one) being $L = \frac{1}{2}$ rather than 1. The reason for this is

that when homozygous females are rendered dead or sterile, then heterozygous females make a relatively larger contribution to the next generation, and hence the spread of the HEG is particularly influenced by the homing that occurs in these heterozygotes. The case of non-sex-specific lethality but homing only in a single sex provides an even worse outcome in terms of load than female lethality and male homing.

Consider now using HEGs to knock out a gene essential for vector transmission. Suppose there are mild costs ($s \ll e$) to the knockout that may be experienced by males, females, or both sexes equally. The HEG will always go to fixation and there are only minor differences in the speed at which this happens (fastest when only one sex suffers the fitness cost).

We have also explored sex-specific fitness costs when the HEG is active prior to gene expression. Qualitatively the conclusions are very similar to the comparison of the two situations in the non-sex-specific case: HEG activity prior to expression always tends to reduce the rate of spread, genetic load, and equilibrium frequency.

Dynamics of HEG-escape mutants: When a homing endonuclease cuts a chromosome it is normally repaired using the second chromosome as a template, the mechanism through which the HEG increases in frequency. But it is possible that the chromosome is rejoined in a different way. Possibly an incomplete copy of the HEG is transferred, or alternatively the ends of the chromosome may be ligated without the use of a template. If the wild-type chromosome is precisely reconstituted, then the initial cut leaves no trace and the only dynamic effect is a reduction in the efficiency of homing, a lower value of the parameter e . But if the repair destroys the HEG recognition site, then non-standard repair can be very important. A critical issue is whether the repaired chromosome, without a HEG recognition site, contains a functioning gene.

To explore this assume there are three classes of allele, the wild type (+), the HEG (H), and a misrepaired allele (M) at frequencies of p , q_H , and q_M , respectively. In reality there are likely to be several classes of misrepaired allele, although we simplify the situation by allowing just a single type. The genotype frequencies and fitnesses are given in the table,

Genotypes	++	+H	+M	HH	HM	MM
Frequency	p^2	$2pq_H$	$2pq_M$	q_H^2	$2q_M q_H$	q_M^2
Fitness	1	$1 - h_H s_H$	$1 - h_M s_M$	$1 - s_H$	$1 - s_f$	$1 - s_M$

where the subscripted s and h parameters describe the fitness costs of the different alleles and their pattern of dominance. Population fitness, \bar{w} , is the average genotype fitness weighted by frequency (and so the HEG load is $1 - \bar{w}$) and γ is the frequency of misrepair [and hence $(1 - \gamma)$ is the frequency of repair resulting in a functional HEG].

With these assumptions we can write recurrences for q_H and q_M ,

$$q'_H = \frac{1}{W} (q_H^2(1 - s_H) + pq_H(1 + e(1 - \gamma))(1 - h_H s_H) + q_M q_H(1 - s_I)), \quad (19)$$

$$q'_M = \frac{1}{W} (q_M^2(1 - s_M) + pq_M(1 - h_M s_M) + pq_H(1 - h_H s_H)e\gamma + q_M q_H(1 - s_I)). \quad (20)$$

The general solution of these equations is too complex to be useful so we explore some special cases.

Assume first that the wild-type allele is fully dominant ($h_H = h_M = 0$) and that both the HEG and the misrepair allele cause nonfunctional gene products leading to death in the absence of a wild-type allele ($s_H = s_M = s_I = 1$). The equilibrium frequency of the two alleles is

$$q_H^* = e(1 - \gamma)^2, \quad q_M^* = e\gamma(1 - \gamma), \quad (21)$$

and the HEG load is $e^2(1 - \gamma)^2$. Thus if the aim of the program is to reduce population densities by targeting a lethal gene, the effect of misrepair leading to nonfunctional alleles is simply to make homing less efficient.

Now assume that the misrepair allele produces a gene product that is at least partially functional. Specifically let $s_M = s_I = s$ while the HEG remains homozygous lethal ($s_H = 1$) and all other parameters are as above. In the case of $s = 0$ the misrepaired allele is completely functional and it can be shown formally that there is no stable equilibrium with the HEG present, no matter how small is the rate with which escape arises (γ). For $s > 0$ an equilibrium exists with both the HEG and the escape mutant present. As the costs of the escape mutant rise, the equilibrium frequency of the HEG also increases. Higher homing rates (e) and higher probabilities of legitimate repair ($1 - \gamma$) lead to both greater frequencies of the HEG and greater load (Figure 5).

Recall that in the absence of misrepair a HEG that has no fitness costs ($s_H = 0$) inexorably becomes fixed. If less than half the time ($\gamma < \frac{1}{2}$) misrepair generates nonfunctional or partially functional alleles ($s_M = s_I > 0$), then fixation of a cost-free HEG still occurs, though it may happen more slowly. If functional alleles are produced that have no fitness costs ($s_M = s_I = 0$), then both the HEG and the misrepair allele increase in frequency until the wild-type allele disappears. The equilibrium frequency of the HEG ($q_H^* = 1 - \gamma(1 - q_{H0})$) depends on the frequency with which these mutant alleles arise (γ) as well as the initial HEG frequency q_{H0} . After this the HEG and misrepair alleles will show neutral dynamics affected only by drift.

Complex dynamics may occur if the HEG targets a gene whose knockout is neither lethal nor of no fitness consequence to the organism ($0 < s_H < 1$). We have not performed a full analysis of all possible scenarios but

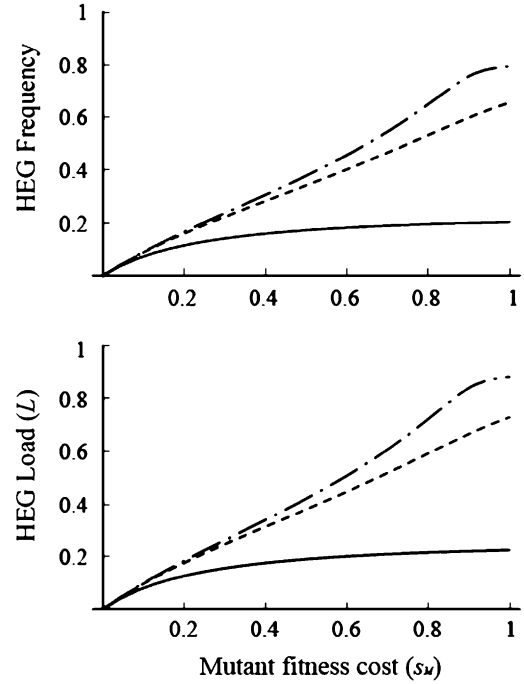


FIGURE 5.—The equilibrium HEG frequency (top) and HEG load (bottom) when the repair of the cut chromosome can produce mutant alleles with intermediate fitness costs, $0 < s_M < 1$. A homing rate of $e = 0.9$ is assumed and the three lines represent different probabilities of misrepair: $\gamma = 0.5$ (solid line), 0.1 (dotted line), and 0.01 (dashed line).

HEGs are more likely to become fixed if they have high fitness in the homozygote and lost if the homozygote is costly. High homing rates (e) tend to favor HEG fixation and low rates loss; high misrepair rates (γ) also increase the chance of loss, while low relative escape mutant fitness ($s_I > s_H$) tends to favor fixation. For intermediate values of s_H polymorphisms may occur with the HEG frequency depending on the relative costs of the HEG and escape mutant, as well as the frequency with which the latter arises. Simulations of the dynamics show complex behaviors including dependence on initial conditions and cycles where the frequency of the HEG at times gets so close to zero that in a real population it could be lost due to stochastic effects. We note that similar, complex dynamics have been observed in other meiotic drive systems (CHARLESWORTH and HARTL 1978; NAUTA and HOEKSTRA 1993).

THEORETICAL RESULTS: X CHROMOSOME SHREDDERS

Assume that in a species with heterogametic males k HEGs are inserted into the Y chromosome and each recognizes a specific, different site on the X chromosome causing a break with probability e . The chance that a particular X chromosome survives this assault is thus $(1 - e)^k$ and the fraction of gametes carrying the Y

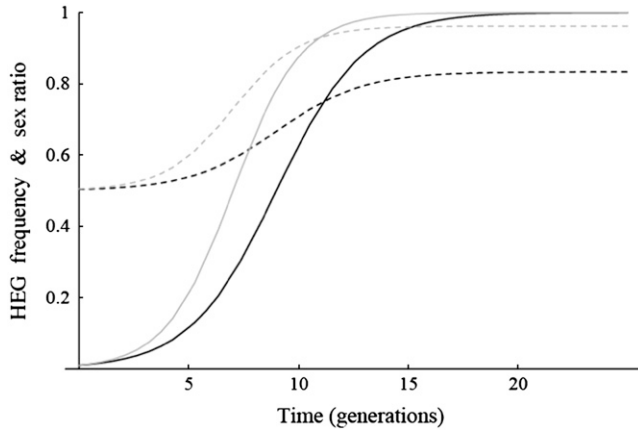


FIGURE 6.—The spread of an X-shredding HEG (solid lines) and its consequences for the population sex ratio (dashed lines). For each pair of lines the bottom (solid) curve represents the case of a single HEG recognition site on the X chromosome ($k = 1$) and the top (shaded) curve represents five recognition sites ($k = 5$). A cutting frequency of $e = 0.8$ is assumed. At equilibrium the HEG loads are 0.67 and 0.99 for $k = 1$ and 5, respectively.

chromosome increases from $\frac{1}{2}$ to $1/[1 + (1 - e)^k]$. If there are no fitness costs to reduced sperm volume and the frequency of HEG-bearing Y chromosomes (as a fraction of all Y chromosomes) in the current generation is q , then in the next generation it will be

$$q' = \frac{q(1/(1 + (1 - e)^k))}{q(1/(1 + (1 - e)^k)) + (1 - q)(1/2)}. \quad (22)$$

The HEG spreads to fixation on the Y chromosome (Figure 6) and hence the sex ratio (proportion of males) is equal to the fraction of Y-bearing gametes $1/[1 + (1 - e)^k]$. Note that in the special case of a single HEG the equilibrium sex ratio is $1/(2 - e)$. We define the HEG load when there is a biased sex ratio to be $L = 1 - 2(1 - r)(\lambda'_f/\lambda_f)$, where r is the sex ratio and λ'_f and λ_f are the average population growth rates (excluding the sex-ratio effect) in the presence and the absence of the HEG, respectively. In this case, once the HEG has become fixed, $L = (1 - (1 - e)^k)/(1 + (1 - e)^k)$.

The equilibrium sex ratio as a function of the chromosome break frequency (e) and the number of HEGs are shown in Figure 7. If breaks occur with a high probability, then the insertion of a single HEG can skew the sex ratio so strongly to males that the population is unlikely to persist. If breaks occur with lower probability, then high skew and population extinction can still be achieved using multiple HEGs. The spread of the gene is relatively fast. For example, a single X shredder with a cutting rate of $e = 0.9$ can increase in frequency from 0.01 to 0.99 in ~ 15 generations. The same speed of spread for lower cutting rates can be achieved with multiple X shredders (Figure 6).

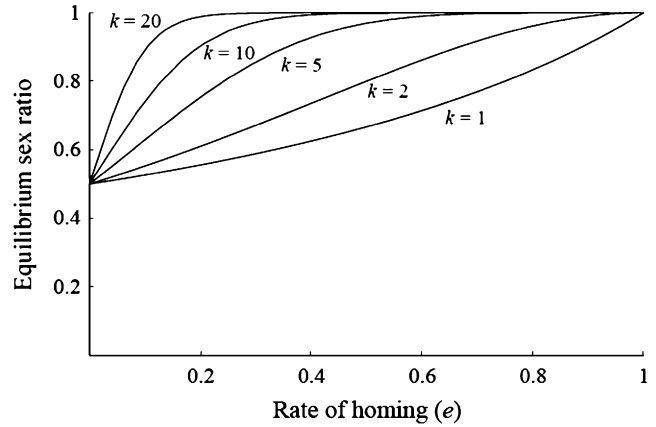


FIGURE 7.—Equilibrium sex ratio in the presence of X-shredding HEGs as a function of the chromosome break frequency (e) and the number of HEGs or HEG recognition sites (k). In all cases the HEG-bearing Y chromosome goes to fixation at equilibrium.

Note that the mathematics is unaltered if the parameter k refers not to the number of unique target sites of multiple HEGs, but to the multiple targets of a single HEG. Theoretically, very high values of k might thus be achieved if the HEG is engineered to recognize common sequences of a tandem array of genes or other multiple-gene families. Note that at some point the concentration of the homing endonuclease might become limiting, depending upon how strongly it is expressed.

Costs due to lower sperm numbers: The arguments above assume that males with reduced sperm production suffer no fitness penalties. The simplest way to relax this assumption is to let the relative cost paid by engineered males be a function, $f(x)$, of relative sperm volume ($x = 1/2[1 + (1 - e)^k]$), where $f(x)$ decreases from one to zero as x varies over the same range. Intuitively, males are penalized by producing fewer sperm. If the costs are sufficiently high that HEG-bearing sperm fertilize fewer eggs than they would have in the absence of X shredding, then the spread of the HEG can be prevented ($f(x) > 1 - x$). Below this threshold the HEG still spreads to fixation and the final sex ratio is the same, but the costs can considerably slow the rate at which this is attained.

A more mechanistic model of the costs of reduced sperm production can also be constructed. Multiple mating will reduce the advantage of the Y chromosome bearing the HEG as some of the benefits of the smaller number of X gametes will be shared with wild-type Y chromosomes from other males. To model this, assume that when a female is mated by x males carrying the HEG and y carrying the wild-type allele, her eggs are fertilized randomly by the pool of sperm produced by the $x + y$ males. The recurrence for the frequency of the HEG becomes

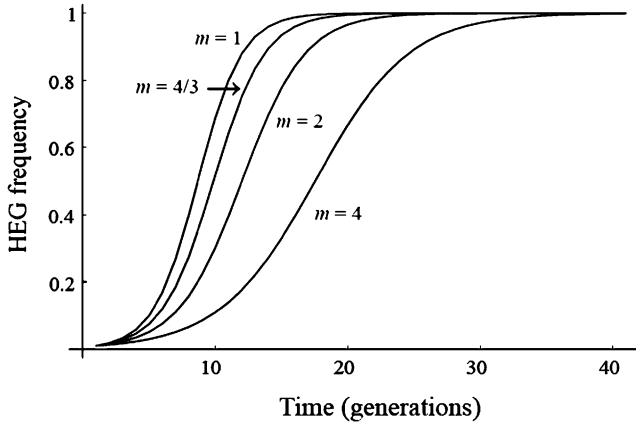


FIGURE 8.—The effect of multiple mating and sperm competition on the rate at which an X-shredder HEG spreads through the population. The precise assumptions made about the distribution of mating frequencies are described in the text and the average number of matings per female is m . We assumed a cutting frequency (e) of 0.9, a HEG with one recognition site ($k = 1$), and a HEG initial frequency of 0.01.

$$q' = \frac{\sum_{x,y} p(x,y)(x/((1 + (1 - e)^k)x + 2y))}{\sum_{x,y} p(x,y)((x+y)/((1 + (1 - e)^k)x + 2y))}, \quad (23)$$

where $p(x, y)$ is the probability of the combination $[x, y]$. If females mate only once at random with the two types of males [so $p(1, 0) = q$, $p(0, 1) = 1 - q$, and all other $p(x, y) = 0$], then Equation 23 simplifies to Equation 22. Another limit is the case when all females are mated by a large number of males ($x + y \rightarrow \infty$), in which case $q' = q$: the HEG allele has exactly the same fitness as the wild-type allele because it gets no special benefit from the reduction in X gametes it brings about.

Figure 8 explores intermediate cases assuming $p(x, y)$ is a compound distribution with the total number of matings ($x + y$) determined by a geometric distribution [the probability that a female is mated $x + y$ times is $(1 - \phi)\phi^{x+y-1}$ with $\phi < 1$] and the numbers of x and y conditional on their sum described by the binomial distribution. Increased frequencies of multiple mating reduce the advantages of X shredding and slow the spread of the HEG, without affecting the final outcome.

Costs due to lower zygote numbers: A different type of cost occurs when sperm that carry shredded and hence inviable X chromosomes can pseudofertilize zygotes. The latter die and are not available for fertilization by the HEG-carrying Y. Let $u = (1 - e)^k$ be the fraction of X chromosomes that avoid shredding. Assume that a fraction z of the remaining $(1 - u)$ shredded X chromosomes pseudofertilize a zygote. In the absence of X shredding the Y chromosome could expect to fertilize $\frac{1}{2}$ the zygotes, and with X shredding but not pseudofertilization this fraction goes up to $1/(1 + u)$. However, if pseudofertilization occurs the fraction drops to $1/(1 + u + (1 - u)z)$. Clearly the ad-

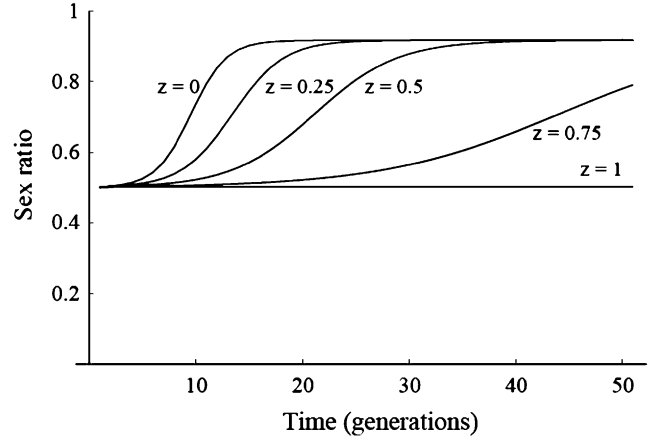


FIGURE 9.—The effect of “pseudofertilization” on the rate at which an X-shredder HEG spreads through the population. It is assumed that a fraction z of sperm with cut X chromosomes fertilize ova that subsequently die. We assumed a cutting frequency (e) of 0.7, a HEG with two recognition sites ($k = 2$), and a HEG initial frequency of 0.01.

vantage of X shredding disappears if $z = 1$ while for $z < 1$ the HEG still spreads to fixation with the final sex ratio unaffected, though the speed of spread declines as z increases (Figure 9).

X-shredder escape mutants: Breaks in the X chromosome may be repaired by ligation of the two ends (repair using the homologous chromosome as a template is not of course possible in the heterogametic sex). If the repair regenerates the HEG recognition site, then this is simply equivalent to a reduction in the efficiency of shredding. But if the recognition site is lost, and if the repaired chromosome suffers no or mild fitness costs after repair, then an X-shredder escape mutant will have been generated.

To explore this for the case of an X shredder with a single recognition site ($k = 1$) we model the dynamics of four kinds of chromosome: the wild-type sex chromosomes (Y and X), Y chromosomes bearing a HEG (Y_h), and X chromosomes bearing the escape mutant (X_m). Among Y chromosomes the frequency of Y_h is q while the frequency of X_m among X chromosomes is ξ_m in sperm and ξ_f in eggs. The genotype frequencies and fitnesses are given in the following tables:

Male genotypes	XY	X_mY	XY_h	X_mY_h
Frequency	$(1 - q)(1 - \xi_f)$	$(1 - q)\xi_f$	$q(1 - \xi_f)$	$q\xi_f$
Fitness	1	$1 - s_m$	$1 - s_H$	$1 - s_m$

Female genotypes	XX	XX_m	X_mX_m
Frequency	$(1 - \xi_m)(1 - \xi_f)$	$\xi_m(1 - \xi_f) + (1 - \xi_m)\xi_f$	$\xi_m\xi_f$
Fitness	1	$1 - s_f h_f$	$1 - s_f$

The subscripted s and h parameters describe the fitness costs of the different alleles and their pattern of dominance. We assume that the costs of bearing a HEG arise directly from the X-shredding process.

Assuming that escape mutants arise with frequency g , the recurrences for gamete frequencies are

$$\begin{aligned} q' &= \frac{q(2 - 2s_H(1 - \xi_f) - 2s_m\xi_f - e(1 - s_m)(1 - \gamma)\xi_f)}{2 - 2qs_H(1 - \xi_f) - 2s_m\xi_f - e(1 - \gamma)(1 - \xi_f) - s_m\xi_f} \\ \xi_m' &= \frac{2(1 - s_m)\xi_f + e(2q\gamma(1 - s_H)(1 - \xi_f) - (1 - s_m)(1 - \gamma)\xi_f)}{2 - 2qs_H(1 - \xi_f) - 2s_m\xi_f - e(1 - \gamma)(1 + q(1 - 2s_H)(1 - \xi_f) - s_m\xi_f)} \\ \xi_f' &= \frac{(1 - h_f s_f)(\xi_f + \xi_m) - 2(1 - h_f)s_f\xi_f\xi_m}{2(1 - s_f\xi_f\xi_m - s_f h_f(\xi_m + \xi_f - 2\xi_m\xi_f))}. \end{aligned} \quad (24)$$

The general solution is very complex and we focus on some special cases. First, assume that there are no costs to HEG carriage ($s_H = 0$). If the escape mutant also has no costs ($s_m = s_f = 0$), then it will always spread to fixation. If an escape mutant suffers fitness costs s_M in the homozygote and the hemizygote ($s_m = s_f = s_M$) and $h_M s_M$ in the heterozygote, then fixation of the escape mutant generally occurs below a critical threshold given by the two inequalities,

$$\begin{aligned} s_M &< \frac{e(1 + \gamma)}{2 - e + e\gamma} \quad \text{and} \\ \frac{e(1 + \gamma) - s_M(3 - 2s_M)(2 - e + e\gamma)}{s_M(e(1 + \gamma) - (2 - s_M)(2 - e + e\gamma))} &< h_M. \end{aligned} \quad (25)$$

Fixation of the mutant restores equal sex ratios and the remaining HEG alleles in the population will then have neutral dynamics because they have identical fitness to the wild type. If the costs are above this threshold, then the HEG will spread to fixation with the population remaining polymorphic for the escape mutant (Figure 10). Thus a single HEG is least affected by resistant escape mutants when its target sequence is an essential gene whose function is lost after repair ($s_M = 1$, $h_M = 0$). Similar dynamics are observed when the gene targeted by the HEG is essential only in males or in females.

Now suppose that there is a cost to X shredding ($s_H > 0$), perhaps because sperm volume is reduced or through pseudofertilization (see above), and focus on cases where the target gene cut on the X chromosome may be essential. If HEG costs exceed a threshold ($s_H > e(1 - \gamma)/2$) that depends on the frequency with which escape mutants arise, then the HEG always goes extinct and the sex ratio returns to equality. If the cost of the HEG is below this threshold, then the HEG becomes fixed and the predicted sex ratio depends on the fitness of the escape mutant. If the escape mutant has wild-type fitness it spreads to fixation and equal sex ratio is restored. But if repair gives rise to an escape mutant that at least in some circumstances is lethal, then a polymorphism may result where the sex ratio shows an intermediate bias toward males. For instance, if the mutant is dominant and lethal in females ($s_f = 1$, $h_f = 1$,

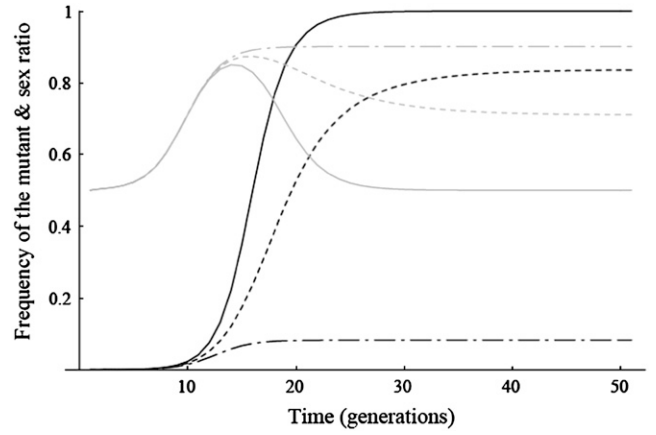


FIGURE 10.—Examples of the spread of an X-shredder escape mutant. Three different assumptions about the fitness of the escape mutant are made (solid lines, $s_M = 0$; dashed lines, $s_M = 0.5$; dotted-and-dashed lines, $s_M = 1$), and the sex ratio (shaded lines) and escape mutant frequency (among males, solid lines) are plotted. We assumed a cutting frequency (e) of 0.9, a HEG with a single recognition site ($k = 1$), that the rate at which escape mutants are generated (γ) is 0.01, and a HEG initial frequency of 0.01.

$s_m = 0$), the sex ratio evolves toward $1/(2 - e + e\gamma)$, and at each generation a proportion $e\gamma/(1 - e + e\gamma)$ of the females die before reproducing.

This analysis suggests that single X shredders, or X shredders with single recognition sites ($k = 1$), are likely to fail if cost-free escape mutants occur. However, the probability of escape mutants of this type arising can be markedly reduced by using multiple HEGs or HEGs that cut at multiple targets on the X chromosome. An escape mutant requires that the X chromosome be rejoined with the loss of the HEG recognition site at all k sites before it can avoid further attack. We have not modeled this scenario in detail but the main effect of using a HEG with multiple shredding sites can be captured by reducing g , the rate at which escape mutants are generated. Though formally for all $g > 0$ long-term persistence of the HEG is not possible when escape mutants have no costs, for small g this may take a very long time. A more detailed model of escape mutant evolution would need to take into account the frequency of X chromosomes that had lost the HEG recognition site at some but not all sites. Also, for HEGs targeting recognition sites in multicopy genes, the possibility of gene conversion and molecular drive would need to be considered.

DISCUSSION

Parasitic or selfish genetic elements have clear attractions as potential tools for population genetic engineering and control since they are able to spread through a population even if they do not increase the fitness of organisms carrying them—indeed, they may spread even if they cause some harm. Moreover, our ever-

increasing understanding of the molecular basis of drive means that it is becoming feasible to synthesize artificial elements ourselves (HAN and BOEKE 2004; ADELMAN *et al.* 2007; CHEN *et al.* 2007). This article is a theoretical investigation of the conditions under which one class of parasitic genetic element, HEGs, might be useful for population genetic engineering and in particular for applications to control disease vectors or pests. The aim of the work presented here is to provide guidance in designing efficient constructs with the lowest likelihood of resistance arising.

Two logically different uses of HEGs are considered. First, we investigated how their intrinsic homing ability might be put to use. A HEG could be designed to target an essential gene and hence to impose a genetic load on a population or in a vector to knock out a gene that is essential for disease transmission but nonessential for the host. Second, we consider a HEG placed on the Y chromosome that recognizes and cuts one or more sites on the X chromosome. By reducing competition from X-carrying sperm the driving Y chromosome spreads, causing a male-biased sex ratio that can severely reduce population growth rate.

Our models show that a HEG targeting an essential gene can potentially cause a substantial reduction in population fitness, while one that targets a gene whose knockout has minor consequences for host fitness quickly goes to fixation. These strategies rely on the break in the chromosome caused by the HEG being repaired by gene conversion with the homologous chromosome as the template. In our models the frequency of homing by this mechanism is described by the parameter e that typically should be as high as possible. But repair by gene conversion is not the only possible pathway, and experiments have shown that the frequency of different types of repair can depend upon the genomic context. For example, if the cleavage site is flanked by direct repeats, then the single-strand annealing (SSA) pathway usually seems to predominate, at least in the premeiotic male germline. In this process all DNA sequence between the repeats is removed: thus the HEG site is lost but a functional gene is unlikely to be reconstituted. Thus SSA reduces the efficiency of homing rather than giving rise to resistant alleles. In one set of experiments with *Drosophila* the SSA pathway was used two-thirds of the time in the premeiotic male germline with conversion accounting for only 10–20% of repairs (PRESTON *et al.* 2006). In other experiments, without direct repeats flanking the cleavage site, conversion repair was observed ~85% of the time [excluding repair of the sister chromatid, which precisely regenerates the target site, producing a chromosome that can simply be cut again (RONG and GOLIC 2003)]. These experiments indicate that it will be important to avoid targets that are flanked by direct repeats. The timing of HEG expression can also be important: if cleavage is late in spermatogenesis, then nonhomolo-

gous end joining (NHEJ) can predominate (PRESTON *et al.* 2006). Here the ends of the chromosome are directly ligated without involvement of the homolog (and, as discussed below, with a high probability of the loss of the HEG recognition sequence), again not the desired result. There may also be differences in the frequency of conversion repair (e), depending upon where the target is located in the genome.

If the aim of the intervention is to impose a load on the population, the models suggest that it is better to target a gene that affects fertility rather than viability and that it is in general better for the target to be recessive (though in some cases maximum load is obtained when heterozygote fitness is somewhat reduced, $0 < h < 0.5$). In *Drosophila* there are thought to be ~3000 genes essential for viability and 50–100 needed for female fertility, most of which are recessive (ASHBURNER *et al.* 1999, 2005). It is also possible to impose a load if one targets male fertility. In *Drosophila* there are two genes (*misfire* and *sneaky*) that, when disrupted, have the effect of impairing embryo development after fertilization (OHSAKO *et al.* 2003; WILSON *et al.* 2006; SMITH and WAKIMOTO 2007). *Misfire* is needed for correct sperm head decondensation and *sneaky* is required for the proper breakdown of the sperm plasma membrane. Homologs of these genes in pest and vector species are potential targets. In principle, if a gene were to affect both male fertility (in this way) and female fertility, then targeting it could be more efficient than knocking out an essential gene, but it is not clear if any such genes exist in insects.

It may also be possible to target genes that have fitness consequences in only one sex or that have different homing frequencies in the two sexes. Targeting only females is advantageous as the reduced overall costs allow the HEG to reach higher equilibrium frequencies (and spread faster) and then impose a greater load on the population. For most human disease vectors it is the female that requires a blood meal prior to oviposition and transmits the pathogen, another advantage for preferentially targeting this sex.

Different control sequences can be used to determine the relative timing of homing and the expression of the gene targeted by the HEG. We find constructs are more invasive if HEG expression is after that of the target gene, so that heterozygotes have more-or-less normal fitness. In many cases HEGs cannot spread if the fitness consequences of gene conversion are experienced by the individuals in which it occurs, and the speed of spread and eventual load of any HEG that has a significant fitness cost is lower if homing occurs before expression of the target gene.

There is one circumstance in which the inability of an early-acting HEG to spread may be a positive advantage. Successful control of several pests has been obtained by the mass release of males that have been sterilized using chemicals or radiation (sterile insect techniques,

SIT). A development of this is to engineer condition-dependent dominant lethals that are mass reared in the presence of an artificial repressor that inactivates this gene (THOMAS *et al.* 2000; ATKINSON *et al.* 2007; PHUC *et al.* 2007). If released into the environment in sufficient numbers they bring about population collapse by competing with wild-type males for mates that then fail to produce offspring (in the absence of the artificial repressor in the field). If the gene is only lethal for females (THOMAS *et al.* 2000), this release of insects with a dominant lethal (RIDL) technique has the additional advantage that the male progeny of released insects can transmit the female lethality to the next generation. Although in the absence of further releases the trait quickly disappears, its temporary persistence in the environment is an advantage of RIDL over SIT. A HEG that could not spread through a population could be combined with a condition-dependent female-lethal construct in a “RIDL-with-drive” strategy, which would be more effective than RIDL by itself (THOMAS *et al.* 2000), though again without long-term persistence.

The models show that for broad regions of parameter space, typically involving HEGs with substantial fitness costs in the heterozygote state, the fate of the construct depends on its initial frequency. Although these HEGs cannot spread from rare, were they to be released in sufficient numbers, the constructs would spread to fixation. The ability of many HEG constructs to spread from rare, even in the presence of costs, distinguishes them from a number of potential drive elements—for example, *medea* elements, Wolbachia, and underdominant chromosomes—that in the presence of costs require frequencies to exceed a threshold before spread occurs (SINKINS and GOULD 2006). If a HEG construct was developed and then found to spread only above a modest threshold frequency, it might still be useful as a drive mechanism. Currently, we see no particular advantages to such constructs and efforts should be directed at developing HEGs that can spread from rare.

For insect-transmitted diseases a different strategy involving homing is to target a gene whose disablement has no or relatively low costs to the vector but that is essential for successful parasite transmission. Such HEGs rapidly rise in frequency to fixation in the population and considerations such as whether the HEG acts before or after target gene expression have little or no effect on whether it spreads. Currently, there is intensive study of mosquito gene products required by *Plasmodium* and other disease agents (*e.g.*, DINGLASAN *et al.* 2007; ECKER *et al.* 2007) and measures of the costs to the vector of their knockout would be very interesting.

Any strategy that aims to impose a fitness load on a population inevitably leads to strong selection for resistance. For HEGs, the easiest way for resistance to occur is for an allele to arise that codes for a functioning gene but that does not possess the HEG recognition site. This could arise through point mutation or, most likely,

through the chromosome break caused by the HEG not being rejoined by homologous repair but by NHEJ. Here we show that if an escape mutant has normal fitness, then it will spread and the HEG will eventually disappear. If the HEG has low fitness costs, then the spread of the escape mutant will occur slowly and significant short-term benefits in population management may still occur. But for HEGs whose prime purpose is imposing load any escape mutant that arises will quickly destroy any benefit.

It is thus critical to design a HEG where loss of the recognition site in the target gene also implies loss of gene function. Some HEGs that target protein-coding genes have evolved to “ignore” the third base silent sites in the target sequence, so as to broaden their own specificity (KOUFOPANOU *et al.* 2002; KUOKAWA *et al.* 2005). If this feature can be maintained in the engineered HEGs, then the most obvious sort of resistant mutation can be nullified. Loss of the recognition site that arises from NHEJ normally involves the deletion of a short stretch of DNA, and hence engineering a HEG to recognize the DNA coding for an active site of an enzyme, or equivalent conserved motif, should ensure its loss causes nonfunctionality. Further strategies that might be explored include targeting multiple sites within the same gene simultaneously and targeting multiple genes simultaneously. Akin to multiple-drug resistance, these measures might substantially delay the evolution of resistance, possibly allowing enough time for local extinction to occur.

The second distinct way of using a HEG is to destroy X chromosomes at male meiosis and so bias the sex ratio (BURT 2003), a strategy formally modeled here for the first time. This mechanism does not rely on homing *per se* but on the HEG providing an extra-Mendelian advantage to its Y chromosome carrier. Naturally occurring driving Y chromosomes are known from two medically important genera of mosquito, *Aedes* and *Culex*, which is encouraging for this strategy (NEWTON *et al.* 1976; SWEENEY and BARR 1978; WOOD and NEWTON 1991; CHA *et al.* 2006). Nothing is known at the molecular level of how they work, but cytologically drive is associated with X chromosome breaks at male meiosis (CAZEMAJOR *et al.* 2000; WINDBICHLER *et al.* 2007). The degree to which breaking the X chromosome favors the Y chromosome depends quite critically on the organism’s biology. We show that if there are costs to reduced sperm production, or if multiple mating occurs and fertilization is a lottery, then the spread of the HEG-bearing Y may be slowed. Similarly, if damaged X chromosomes “pseudofertilize” eggs, that is, if they compete with other sperm and create zygotes that then die, the HEG-bearing Y spread may be slowed and in the limit even totally prevented. We lack information on these issues for major vectors and pests and studies on the reproductive biology of a candidate species would need to be done in advance to assess the likely outcome of

adopting this approach. For example, if irradiation of sperm (which causes chromosome breaks) leads to death of zygotes and not sperm, as occurs in some species, then this would suggest pseudofertilization might occur.

An escape mutant on the X chromosome that destroys the HEG recognition site and does not dramatically reduce fitness would increase in frequency and ultimately return the sex ratio to equality. We would also expect a suppressor mutant on an autosome to spread in a similar manner and destroy the sex-ratio bias, though we have not modeled this scenario. Again, as with classical HEGs, the probability of resistance occurring can be reduced by designing the HEG to cut multiple sites on the X chromosome. This is also advantageous as for cutting frequencies (e) less than one multiple site increases the equilibrium sex-ratio bias. An example of how this might be implemented is provided by *Anopheles gambiae*, the most important vector of malaria in sub-Saharan Africa. In this species rDNA repeats are restricted to the X chromosome, and the I-PpoI homing endonuclease of *Physarum* slime molds recognizes and cuts a sequence within the gene (WINDBICHLER *et al.* 2007). Given that there are >100 copies of the rDNA gene (COLLINS and PASKEWITZ 1996), the evolution of resistance will be a very rare event. Note that if the X-shredding HEG was placed on an autosome, it would still disrupt X sperm but not spread. This strategy could be employed in a noninvasive pest or vector management campaign, although it may be necessary to make HEG expression condition dependent (using a RIDL-like strategy) to enable efficient mass rearing.

Hybrids between the two main strategies—classical homing and biasing the sex ratio—might also be devised. In the major agricultural pest, the medfly (*Ceratitis capitata*), XX individuals require at least one functional copy of the autosomal gene *tra* (transformer) to be phenotypic females (PANE *et al.* 2005). If a HEG was engineered to knock out the *tra* gene, it would drive the Y chromosome to extinction so that all males would be XX *tra*⁻ *tra*⁻ and the sex ratio would be $(1 - e)/2$, where e is the homing efficiency (detailed modeling not shown).

Throughout this article we have measured the effect of HEGs designed to reduce pest or vector population densities by what we have called “HEG load,” the decrease in population fitness caused by reductions in viability and fertility. The effect of sex-ratio biases on population fitness can be measured using the same metric. It is critically important to stress that how HEG load translates into real population reductions depends on the particular demography and population dynamics of the species involved and also the context in which pest or vector management is sought. For example, if population control is attempted for a seasonally invasive species or one with boom-and-bust population dynamics, then the most important concern may be to reduce maximum population growth rate. In these circumstances HEG load will map directly onto population

reduction and be a simple means of comparing different strategies. However, if a population is subject to density-dependent mortality, then it is possible for a substantial HEG load to be present but only modest or even no reductions in pest or vector numbers to be observed. In these circumstances comparing HEG load provides a ranking of the effectiveness of different strategies, though not a quantitative metric. In this article we have deliberately not extended the models developed here beyond frequency into the density domain as we believe analyses based on particular systems and specific biologies are the most effective means of exploring these issues. We are also aware that our models have been purely deterministic and have assumed discrete generations: stochastic models with overlapping generations will be valuable in providing guidance about the numbers of HEG-bearing insects that should be released in an actual management program.

The aim of the work described here is primarily to help in the design of HEGs that might be used in pest and vector management, especially involving insects. The success of a HEG-based strategy will depend on (i) the discovery of suitable recognition sites or the successful engineering of HEGs to recognize new sites, (ii) the ability of HEGs to cut and home in insects, (iii) the design of constructs and implementation strategies to avoid or delay resistance, and (iv) the regulatory acceptance of a pest or vector management strategy that involves genetic manipulation. We have mentioned above one significant recognition site in a malaria vector that can be cut by a naturally occurring HEG, while there has been substantial recent progress on reengineering HEGs to recognize novel sites (ASHWORTH *et al.* 2006; ARNOULD *et al.* 2007). Preliminary work suggests artificially introduced HEGs can function in mosquito cells (WINDBICHLER *et al.* 2007). Much remains to be done but we believe HEGs offer exciting prospects for novel approaches to pest and vector management and that consideration of their population genetics will be important in guiding this development.

We are grateful to Michael Ashburner, Andrea Crisanti, Fred Gould, Penny Hancock, Ray Monnat, Samantha O’Loughlin, Steve Russell, Barry Stoddard, and an anonymous referee for helpful discussions. This work was funded by a grant from the Foundation for the National Institutes of Health through the Grand Challenges in Global Health.

LITERATURE CITED

- ADELMAN, Z. N., N. JASINSKIENE, S. ONAL, J. JUHN, A. ASHIKYAN *et al.*, 2007 Nanos gene control DNA mediates developmentally regulated transposition in the yellow fever mosquito *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA* **104**: 9970–9975.
- ALPHEY, L., C. B. BEARD, P. BILLINGSLEY, M. COETZEE, A. CRISANTI *et al.*, 2002 Malaria control with genetically manipulated insect vectors. *Science* **298**: 119–121.
- ARNOULD, S., C. PEREZ, J. P. CABANIOLS, J. SMITH, A. GOUBLE *et al.*, 2007 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.

- ASHBURNER, M., S. MISRA, J. ROOTE, S. E. LEWIS, R. BLAZEJ *et al.*, 1999 An exploration of the sequence of a 2.9-Mb region of the genome of *Drosophila melanogaster*: the *Adh* region. *Genetics* **153**: 179–219.
- ASHBURNER, M., K. G. GOLIC and R. S. HAWLEY, 2005 *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- ASHWORTH, J., J. J. HAVRANEK, C. M. DUARTE, D. SUSSMAN, R. J. MONNAT *et al.*, 2006 Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* **441**: 656–659.
- ATKINSON, M. P., Z. SU, N. ALPHEY, L. S. ALPHEY, P. G. COLEMAN *et al.*, 2007 Analyzing the control of mosquito-borne diseases by a dominant lethal genetic system. *Proc. Natl. Acad. Sci. USA* **104**: 9540–9545.
- BURT, A., 2003 Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **270**: 921–928.
- BURT, A., and V. KOUFOPANOU, 2004 Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Curr. Opin. Genet. Dev.* **14**: 609–615.
- BURT, A., and R. TRIVERS, 2006 *Genes in Conflict*. Belknap Press, Cambridge, MA.
- CAZEMAJOR, M., D. JOLY and C. MONTCHAMP-MOREAU, 2000 Sex-ratio meiotic drive in *Drosophila simulans* is related to equational nondisjunction of the Y chromosome. *Genetics* **154**: 229–236.
- CHA, S. J., D. D. CHADEE and D. W. SEVERSON, 2006 Population dynamics of an endogenous meiotic drive system in *Aedes aegypti* in Trinidad. *Am. J. Trop. Med. Hyg.* **75**: 70–77.
- CHARLESWORTH, B., and D. L. HARTL, 1978 Population dynamics of the segregation distorter polymorphism of *Drosophila melanogaster*. *Genetics* **89**: 171–192.
- CHEN, C. H., H. X. HUANG, C. M. WARD, J. T. SU, L. V. SCHAEFFER *et al.*, 2007 A synthetic maternal-effect selfish genetic element drives population replacement in *Drosophila*. *Science* **316**: 597–600.
- COATES, C. J., N. JASINSKIENE, L. MIYASHIRO and A. A. JAMES, 1998 Mariner transposition and transformation of the yellow fever mosquito, *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA* **95**: 3748–3751.
- COLLINS, F. H., and S. M. PASKEWITZ, 1996 A review of the use of ribosomal DNA (rDNA) to differentiate among cryptic *Anopheles* species. *Insect Mol. Biol.* **5**: 1–9.
- CURTIS, C. F., 1968 Possible use of translocations to fix desirable genes in insect pest populations. *Nature* **218**: 368–369.
- DAVIS, S., N. BAX and P. GREWE, 2001 Engineered underdominance allows efficient and economical introgression of traits into pest populations. *J. Theor. Biol.* **212**: 83–98.
- DINGLASAN, R. R., A. ALAGANAN, A. K. GHOSH, A. SAITO, T. H. VAN KUPPEVELT *et al.*, 2007 *Plasmodium falciparum* ookinetes require mosquito midgut chondroitin sulfate proteoglycans for cell invasion. *Proc. Natl. Acad. Sci. USA* **104**: 15882–15887.
- ECKER, A., S. B. PINTO, K. W. BAKER, F. C. KAFATOS and R. E. SINDEN, 2007 *Plasmodium berghei*: *Plasmodium* perforin-like protein 5 is required for mosquito midgut invasion in *Anopheles stephensi*. *Exp. Parasitol.* **116**: 504–508.
- FRANZ, A. W. E., I. SANCHEZ-VARGAS, Z. N. ADELMAN, C. D. BLAIR, B. J. BEATY *et al.*, 2006 Engineering RNA interference-based resistance to dengue virus type 2 in genetically modified *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA* **103**: 4198–4203.
- GODDARD, M. R., and A. BURT, 1999 Recurrent invasion and extinction of a selfish gene. *Proc. Natl. Acad. Sci. USA* **96**: 13880–13885.
- HAMILTON, W. D., 1967 Extraordinary sex ratios. *Science* **156**: 477–488.
- HAN, J. S., and J. D. BOEKE, 2004 A highly active synthetic mammalian retrotransposon. *Nature* **429**: 314–318.
- ITO, J., A. GHOSH, L. A. MOREIRA, E. A. WIMMER and M. JACOBS-LORENA, 2002 Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. *Nature* **417**: 452–455.
- JAMES, A. A., 2005 Gene drive systems in mosquitoes: rules of the road. *Trends Parasitol.* **21**: 64–67.
- KNOLS, B. G. J., R. C. HOOD-NOWOTNY, H. BOSSIN, G. FRANZ, A. ROBINSON *et al.*, 2006 GM sterile mosquitoes—a cautionary note. *Nat. Biotechnol.* **24**: 1067–1068.
- KOUFOPANOU, V., M. R. GODDARD and A. BURT, 2002 Adaptation for horizontal transfer in a homing endonuclease. *Mol. Biol. Evol.* **19**: 239–246.
- KUROKAWA, S., Y. BESSHO, K. HIGASHIJIMA, M. SHIROUZU, S. YOKOYAMA *et al.*, 2005 Adaptation of intronic homing endonuclease for successful horizontal transmission. *FEBS J.* **272**: 2487–2496.
- MAGORI, K., and F. GOULD, 2006 Genetically engineered underdominance for manipulation of pest populations: a deterministic model. *Genetics* **172**: 2613–2620.
- MOREIRA, L. A., J. ITO, A. GHOSH, M. DEVENPORT, H. ZIELER *et al.*, 2002 Bee venom phospholipase inhibits malaria parasite development in transgenic mosquitoes. *J. Biol. Chem.* **277**: 40839–40843.
- NAUTA, M. J., and R. F. HOEKSTRA, 1993 Evolutionary dynamics of spore killers. *Genetics* **135**: 923–930.
- NEWTON, M. E., R. J. WOOD and D. I. SOUTHERN, 1976 Cytogenetic analysis of meiotic drive in the Mosquito, *Aedes aegypti* (L.). *Genetica* **46**: 297–318.
- OHSAKO, T., K. HIRAI and M. T. YAMAMOTO, 2003 The *Drosophila* *misfire* gene has an essential role in sperm activation during fertilization. *Genes Genet. Syst.* **78**: 253–266.
- PANE, A., A. DE SIMONE, G. SACCONI and C. POLITO, 2005 Evolutionary conservation of *Ceratitis capitata* transformer gene function. *Genetics* **171**: 615–624.
- PHUC, H. K., M. H. ANDREASEN, R. S. BURTON, C. VASS, M. J. EPTON *et al.*, 2007 Late-acting dominant lethal genetic systems and mosquito control. *BMC Biol.* **5**: 11.
- PRESTON, C. R., C. C. FLORES and W. R. ENGELS, 2006 Differential usage of alternative pathways of double-strand break repair in *Drosophila*. *Genetics* **172**: 1055–1068.
- RONG, Y. S., and K. G. GOLIC, 2003 The homologous chromosome is an effective template for the repair of mitotic DNA double-strand breaks in *Drosophila*. *Genetics* **165**: 1831–1842.
- SINKINS, S. P., and H. C. J. GODFRAY, 2004 Use of *Wolbachia* to drive nuclear transgenes through insect populations. *Proc. R. Soc. B Biol. Sci.* **271**: 1421–1426.
- SINKINS, S. P., and F. GOULD, 2006 Gene drive systems for insect disease vectors. *Nat. Rev. Genet.* **7**: 427–435.
- SMITH, M. K., and B. T. WAKIMOTO, 2007 Complex regulation and multiple developmental functions of misfire, the *Drosophila melanogaster* ferlin gene. *BMC Dev. Biol.* **7**: 21.
- STODDARD, B. L., 2005 Structure, function and mechanism of bifunctional homing endonuclease-maturase proteins. *Biophys. J.* **88**: 21a–22a.
- SWEENEY, T. L., and A. R. BARR, 1978 Sex-ratio distortion caused by meiotic drive in a mosquito, *Culex pipiens* L. *Genetics* **88**: 427–446.
- THOMAS, D. D., C. A. DONNELLY, R. J. WOOD and L. S. ALPHEY, 2000 Insect population control using a dominant, repressible, lethal genetic system. *Science* **287**: 2474–2476.
- TURELLI, M., and A. A. HOFFMANN, 1999 Microbe-induced cytoplasmic incompatibility as a mechanism for introducing transgenes into arthropod populations. *Insect Mol. Biol.* **8**: 243–255.
- WADE, M. J., and R. W. BEEMAN, 1994 The population dynamics of maternal-effect selfish genes. *Genetics* **138**: 1309–1314.
- WERREN, J. H., 1997 Biology of *Wolbachia*. *Annu. Rev. Entomol.* **42**: 587–609.
- WILSON, K. L., K. R. FITCH, B. T. BAFUS and B. T. WAKIMOTO, 2006 Sperm plasma membrane breakdown during *Drosophila* fertilization requires *sneaky*, an acrosomal membrane protein. *Development* **133**: 4871–4879.
- WINDBICHLER, N., P. A. PAPATHANOS, F. CATTERUCCIA, H. RANSON, A. BURT *et al.*, 2007 Homing endonuclease mediated gene targeting in *Anopheles gambiae* cells and embryos. *Nucleic Acids Res.* **35**: 5922–5933.
- WOOD, R. J., and M. E. NEWTON, 1991 Sex-ratio distortion caused by meiotic drive in mosquitoes. *Am. Nat.* **137**: 379–391.