The Impact of Dissociation on Transposon-mediated Disease Control Strategies

John M. Marshall[*]

[*]Department of Biomathematics, UCLA School of Medicine, Los Angeles, CA 90095-1766

1

Running head: Dissociation and Transposon Spread

Keywords: transposable element, internal deletion, malaria, dengue, *Anopheles gambiae*

Corresponding author:

John M. Marshall

Department of Biomathematics

UCLA School of Medicine

University of California at Los Angeles

Los Angeles, California 90095-1766

Voice: (310) 825-1602

Fax: (310) 825-8685

Email: johnmm@ucla.edu

ABSTRACT

Vector-borne diseases such as malaria and dengue fever continue to be a major health concern through much of the world. The emergence of chloroquine-resistant strains of malaria and insecticide-resistant mosquitoes emphasize the need for novel methods of disease control. Recently, there has been much interest in the use of transposable elements to drive resistance genes into vector populations as a means of disease control. One concern that must be addressed before a release is performed is the potential loss of linkage between a transposable element and resistance gene. Transposable elements such as *P* and *hobo* have been shown to produce internal deletion derivatives at a significant rate, and there is concern that a similar process could lead to loss of the resistance gene from the drive system following a transgenic release. Additionally, transposable elements such as *Himar1* have been shown to transpose significantly more frequently when free of exogenous DNA. Here, we show that any transposon-mediated gene drive strategy must have an exceptionally low rate of dissociation if it is to be effective. Additionally, the resistance gene must confer a large selective advantage to the vector in order to surmount the effects of a moderate dissociation rate and transpositional handicap.

INTRODUCTION

The creation of transgenic mosquitoes that have a fitness advantage when feeding on *Plasmodium*-infected blood (Marrelli *et al.* 2007) has greatly renewed interest in the use of genetically modified vectors as a means of disease control. Resistance genes are unlikely to reach fixation in a wild vector population on their own, partly because disease prevalence tends to be relatively low in vector populations (Beier *et al.* 1999), and partly because the selective advantage of a resistance gene diminishes as the disease becomes less prevalent (Boëte and Koella 2003). Consequently, a series of gene drive systems have been proposed to bias the disease-resistance gene in favor of fixation (Craig 1963; Curtis 1968). One such drive system is a transposable element (TE), which is able to spread through a population by virtue of its ability to replicate within a genome (Charlesworth *et al.* 1994).

Any successful gene drive strategy requires tight linkage between the drive system and effector gene (James 2005). A general concern of drive systems is that rare recombination events can lead to loss of linkage between the drive system and effector gene (Riehle *et al.* 2003; Knols and Scott 2003; Curtis 2003). For TEs, there is an additional concern that internal deletion of DNA sequences within elements can occur during transposition. In the class of transposons being considered for genetic modification, this is thought to occur by an abortive gap repair mechanism (Rubin and Levy 1997). Following excision or transposition, a double-stranded gap is introduced into the host chromosomal DNA. This gap is sometimes filled by copying information from a homologous chromosome, sister chromatid or ectopic chromosomal site still containing the TE. If this process is interrupted, then the central portions of the element will not be copied, leading to an internal deletion event (Figure 1).

The rate of internal deletion is largely dependent on the TE and host species being considered. In *Drosophila melanogaster*, *P* (Engels 1989) and *hobo* (Gelbart and Blackman 1989) elements produce deletion derivatives at a significant frequency due to abortive gap repair. The same mechanism is primarily responsible for the formation of nonautonomous *Ds* elements in maize (Rubin and Levy 1997) and the prevalence of nonautonomous *mariner* elements in many natural populations (Lohe *et al.* 2000). Despite this, there are other TEs, such as the *Herves* element in *Anopholes gambiae*, that remain in an almost exclusively intact form throughout evolutionary history (Subramanian *et al.* 2007).

Variability in the rate of internal deletion poses the question: what dissociation rate can be tolerated by an effective disease control strategy? If the dissociation rate is exceptionally slow, then the TE will drive the resistance gene into the vector population before the resistance gene has a chance to become detached. However, if the dissociation rate is exceptionally fast, then the resistance gene will be lost from the construct as the transposon spreads through the population, preventing the resistance trait from reaching a sufficiently high prevalence for effective disease control.

The situation is complicated by the tendency of TEs to transpose less frequently as their size is increased. Experiments in which *Himar1 mariner* elements were increased in size by adding various fragments of exogenous DNA (Lampe *et al.* 1998) found that the mobility of these elements is inversely proportional to their size. A similar relationship was observed for *P* elements in *D. melanogaster* (Spradling 1986). These results are particularly relevant to the use of genetically modified vectors since it is projected that any useful gene drive system will carry at least two effector genes, a marker gene and regulatory elements (Marcelo Jacobs-Lorena,

personal communication). Of concern is the possibility that elements no longer carrying their transgenic load will spread more quickly than their intact counterparts.

Another complicating factor is the impact of disease-resistance genes on host fitness. Mounting an immune response is generally associated with an evolutionary cost in insects (Kraaijeveld and Godfray 1997) and hence any resistance gene is likely to be associated with a selective disadvantage. Despite this, transgenic mosquitoes have been created that have no noticeable fitness cost when fed on *Plasmodium*-free blood (Moreira *et al.* 2004) and are in fact more fit when fed on *Plasmodium*-infected blood (Marrelli *et al.* 2007). Whether the resistance gene confers a fitness benefit or cost to the host will significantly influence the fate of the resistance gene in the vector population.

Successful disease control requires that the disease-resistance trait reaches a high prevalence in the vector population within a reasonable timeframe. Throughout our modeling, we make recommendations regarding the dissociation rates, transpositional handicaps and fitness costs or benefits required to achieve this goal.

MATERIALS AND METHODS

We use a system of ordinary differential equations to describe element spread through a randomly mating population incorporating dissociation (Figure 2). To simulate a transgenic release, we initiate the model with a proportion of the vector population having one copy of the intact TE construct and the rest having none. Element spread occurs as long as replicative transposition exerts a greater influence on element copy number than element deletion and host fitness costs. Dissociation leads to the number of intact constructs decreasing by one and the

number of elements free of their transgenic load increasing by one. This reduces the number of TEs having a copy of the disease-resistance gene.

Mathematical model: Particles in the model may have between 0 and T copies of the TE, where T is the maximum element copy number attainable during the period of spread. Of these copies, m are attached to the resistance gene and n are dissociated from their transgenic load. We keep track of the proportion of vectors having each genotype, $x_{(m,n)}$, using a separate differential equation for each one. The process of element spread can then be modeled by a system of $((T+1)^2 + T + 1)/2$ simultaneous ordinary differential equations:

$$
\begin{aligned}
\frac{dx_{(m,n)}(t)}{dt} &= \theta(t) \sum_{i,j,k,l} p_{(i,j),(k,l),(m,n)} x_{(i,j)}(t) x_{(k,l)}(t) - \mu_{(m,n)}(t) x_{(m,n)}(t) \\
&+ (m-1)(1-\delta)u_{m+n-1} x_{(m-1,n)}(t) + (n-1)u_{m+n-1} x_{(m,n-1)}(t) \\
&+ (m+1)v x_{(m+1,n)}(t) + (n+1)v x_{(m,n+1)}(t) \\
&+ (m+1)w_{(m+1,n-1)} x_{(m+1,n-1)}(t) \\
&- \left\{ m(1-\delta)u_{m+n} + nu_{m+n} + mv + nv + mw_{(m,n)} \right\} x_{(m,n)}(t)
\end{aligned}
\qquad (1)
$$

where $n, m \geq 0$ and $n + m \leq T$.

The first two terms in Equation 1 describe the birth and death rates of a vector having the genotype (m,n). The death rate of a type-(m,n) vector, $\mu_{(m,n)}$, is given by the equation:

$$
\mu_{(m,n)}(t) = 1 + d(t)(m+n)^p + m\sigma_- - 1_{\{m>0\}}\sigma_+ \sum_{n=0}^{T} x_{(0,n)}(t) .
\qquad (2)
$$

The first term in this equation represents the death rate of a type-$(0,0)$ vector, the second term represents the fitness cost of $m+n$ TEs, the third term represents the fitness cost of m resistance genes, and the fourth term represents the selective advantage of the resistance trait. The fitness cost of a single TE, $d(t)$, is assumed to decrease over time according to the equation:

$$
d(t) = (d_0 - d_\infty)2^{-t/t_{1/2}} + d_\infty .
\qquad (3)
$$

Here, d_0 represents the mean fitness cost of a new TE insert, d_∞ represents the mean fitness cost of a TE segregating in the population, and $t_{1/2}$ represents the number of vector generations after which the mean fitness cost of a TE is $(d_0 + d_\infty)/2$. Following Charlesworth and Charlesworth (1983), we assume that the fitness cost of the TE is a log-concave function of element copy number. The power p determines the rate at which this fitness cost increases with total TE copy number (Charlesworth and Charlesworth 1983). The fitness cost of the resistance gene, σ_-, is assumed to increase linearly with dosage, m; while the fitness benefit of the resistance trait, σ_+, is conditional upon the presence of one or more resistance genes, and is proportional to the fraction of the vector population that is susceptible to the disease at time t.

The total birth rate of disease vectors, $\theta(t)$, is equal to the total death rate, normalizing for the sum of $x_{(i,j)}(t)x_{(k,l)}(t)$ terms in the birth rate term. This leads to the equation:

$$\theta(t) = \frac{\sum_{m,n} \mu_{(m,n)}(t) x_{(m,n)}(t)}{\sum_{i,j,k,l} x_{(i,j)}(t) x_{(k,l)}(t)} \ . \tag{3}$$

The birth rate of a type-(m,n) vector is then given by summing over every pair of genotypes in the population and weighting each pair by the probability that, when a type-(i,j) vector mates with a type-(k,l) vector, the offspring will be of type-(m,n). Assuming that all TEs are fairly well-spaced and hence segregate independently, this probability is given by:

$$P_{(i,j),(k,l),(m,n)} = \sum_{x=0}^{m}\sum_{y=0}^{n} \frac{1}{2^{i+j+k+l}} \binom{j}{y}\binom{l}{n-y}\binom{i}{x}\binom{k}{m-x} \ . \tag{4}$$

The tendency of some elements to transpose locally (Tower *et al.* 1993; Zhang and Spradling 1993; Guimond *et al.* 2003) and display homing behavior (Guimond *et al.* 2003), though important, is beyond the scope of this study.

The third and fourth terms in Equation 1 describe the rate at which replicative transposition increases the proportion of type-(m,n) vectors in the population. This rate is proportional to the rate of replicative transposition and, due to the presence of repression of transposition (Charlesworth and Charlesworth 1983; Townsend and Hartl 2000; Le Rouzic and Capy 2005), is also a function of total element copy number. In a genome having a total of i element copies, the replicative transposition rate can be described as,

$$u_i = u_1 2^{-c(i-1)} \; ,$$ (5)

where c determines the rate at which the replicative transposition rate falls off with additional element copies, and u_1 is the replicative transposition rate when there is only a single element in the genome. Transposition is slowed down by the presence of a resistance gene, and so the replicative transposition rate of an intact TE construct is given by $(1-\delta)u_i$, where δ represents the fractional decrease in transposition rate due to the effector gene.

The fifth and sixth terms in Equation 1 describe the rate at which element deletion increases the proportion of type-(m,n) vectors in the population. The rate of element deletion, v, is very small and we assume this to be constant, independent of the total TE copy number or transgenic load. The seventh term describes the rate at which dissociation increases the proportion of type-(m,n) vectors in the population. Internal deletion is hypothesized to occur during the act of transposition, and hence its rate is modeled as proportional to the rate of replicative transposition:

$$w_{(m,n)} = \frac{u_{m+n}}{u_1} w^* \; .$$ (6)

Here, w^* is the dissociation rate when there is only a single intact construct in the genome. Finally, the eighth term describes the rate at which replicative transposition, element deletion and dissociation decrease the proportion of type-(m,n) vectors in the host population.

Parameter values: Table 1 contains the parameter values and ranges used in the implementation of our model. At present there is little or no data regarding the post-integration behavior of the candidate elements in human disease vectors such as *An. gambiae* and *Aedes aegypti*, so most of these values were taken from measurements in other species, in particular *D. melanogaster* (Nuzhdin *et al.* 1997; Maside *et al.* 2000; Guimond *et al.* 2003; Pasyukova *et al.* 2004).

We consider a baseline replicative transposition rate of $u_1 = 0.1$ per element per generation (TE^{-1}gen^{-1}). Although this is a fairly high transposition rate, it is realistic (Seleme *et al.* 1999; Vasilyeva *et al.* 1999) and several modeling approaches have recommended it as a minimum requirement for gene drive to occur in a timeframe acceptable to public health goals (Rasgon and Gould 2005; Le Rouzic and Capy 2006). We chose the transpositional regulation parameter to be $c = 2.866$ per element (TE^{-1}) in order to achieve an average equilibrium element copy number of four. This was motivated by the observation that *Herves*, a functional Class II TE in *An. gambiae*, is present across many geographical locations at an average copy number of four (Subramanian *et al.* 2007). *Herves* is of particular interest to transposon-mediated gene drive strategies since it rarely undergoes internal deletion and maintains its structural integrity over long periods of time (Subramanian *et al.* 2007).

Element deletion events are so rare that they are often not observed in laboratory line experiments (Nuzhdin and Mackay 1994). We consider a deletion rate of $v = 4 \times 10^{-6}$ TE^{-1}gen^{-1}

as suggested by pooled estimates from several laboratory line experiments (Nuzhdin *et al.* 1997; Maside *et al.* 2000).

Following Mackay *et al.* (1992), we estimate the mean fitness cost of a new TE insertion by the average fitness cost of a spontaneous mutation. This has been estimated as $d_0 = 0.02$ TE^{-1} (Mukai *et al.* 1972; Ohnishi 1977; Crow and Simmons 1983), and is a reasonable estimate in the early stages of TE spread when insertional mutagenesis is the dominant fitness cost and selection has not yet eliminated TEs with higher fitness costs (Charlesworth 1991). For TEs already segregating in the population, the mean fitness cost has been estimated from simple models of population dynamics by sampling chromosomes from natural populations and assuming that selection and transposition have an equal and opposite influence on TE copy number at equilibrium (Nuzhdin *et al.* 1997; Nuzhdin 2000). This has led to an estimate of fitness cost per element between 10^{-5} and 10^{-3} (Nuzhdin *et al.* 1997). We consider a fitness cost of $d_\infty = 10^{-3}$ TE^{-1} since studies on laboratory lines of *D. melanogaster* suggest a fitness cost toward the upper end of this range (Pasyukova *et al.* 2004).

Charlesworth (1991) has proposed models to understand the abundance of TEs in nature when chromosomal sites vary with respect to the effect of element insertions on host fitness. These models predict that TEs will become more highly represented at neutral or weakly selected sites, and that the mean fitness cost of a TE will decline over a time-scale "roughly equal to the reciprocal of the rate of transposition." For a transposition rate of $u_1 = 0.1$ TE^{-1}gen^{-1}, this corresponds to a fitness decline over a scale of ten generations. We therefore consider a halftime of $t_{1/2} = 5$ mosquito generations for the mean fitness cost of the TE to fall from d_0 to $(d_0 + d_\infty)/2$. Following Charlesworth and Charlesworth (1983), we also consider a power of

$p = 1.5$ to account for additional fitness costs at higher copy numbers, such as ectopic

recombination (Langley *et al.* 1988).

A realistic range for the rate at which exogenous DNA becomes dissociated from an

engineered TE construct can be deduced from experiments in which *P* elements were screened

for loss of a visible marker gene (Daniels *et al.* 1985; Levis *et al.* 1985; Robertson *et al.* 1988).

Here, expression of the *rosy* and *white* marker genes is disrupted at a rate of $w^* \approx 0.05$ TE^{-1}gen^{-1}. Data on the fractional decrease in replicative transposition rate due to the presence of

exogenous DNA (Spradling 1986; Lampe *et al.* 1998), the fitness cost of mounting an immune

response (Kraaijeveld and Godfray 1997; Moreira *et al.* 2004) and the selective advantage of

being disease-resistant (Beier *et al.* 1999; Marrelli *et al.* 2007) can be used to form realistic

ranges for the parameters δ, σ_- and σ_+ respectively.

Limiting our considerations to a single village where a disease vector is present, we

estimate the initial proportion of disease vectors having one copy of the intact TE construct. The

village of Banambani in Mali, West Africa serves as a well-studied example for the malaria

vector *An. gambiae* (Taylor *et al.* 2001). Here, there is an estimated abundance of 2,000

mosquitoes of any given chromosomal form at the peak of the dry season (Taylor *et al.* 2001).

Considering a release of 50 transgenic mosquitoes, each having a single copy of the intact TE

construct, leads to an initial release proportion of $x_{(1,0)}(0) = 0.025$.

Numerical analysis: As a measure of the spread of disease-resistance, we calculate the

proportion of vectors having one of more copies of the intact construct over time. This quantity

can be calculated numerically using the fourth order Runge-Kutta method. For the

parameterization described above, the maximum number of TEs in the genome is $T = 16$. This

can be used to reduce the number of equations being solved simultaneously.

RESULTS

Dissociation rate: Using this system of differential equations, we consider the impact of varying the dissociation rate on the spread of the disease-resistance gene through a randomly mating vector population. In the simplest case (Figure 3), the resistance gene has no fitness consequences for the host and confers no transpositional handicap on the TE. The total genomic copy number reaches an equilibrium of four within 20 years and is very close to equilibrium after 10 years (Figure 3A-C). Dissociation is hypothesized to occur during the act of transposition, and hence the proportion of dissociated TEs increases most rapidly during the early stages of element spread. Once equilibrium has been reached, transposition still occurs at a slower rate to counteract the forces of selection, deletion and the random effects of gamete sampling. Consequently, the proportion of intact constructs decreases slowly until the resistance gene is lost from the population.

The proportion of element copies that have an intact copy of the disease-resistance gene is largely dependent on the dissociation rate. For a dissociation rate of $0.001 \text{ TE}^{-1}\text{gen}^{-1}$, almost all element copies contain an intact resistance gene, and the average number of intact gene copies per individual is significantly greater than three even after 40 years of spread (Figure 3A). This suggests that, for default parameters, a dissociation rate of $0.001 \text{ TE}^{-1}\text{gen}^{-1}$ is acceptable for public health goals. A dissociation rate of $0.01 \text{ TE}^{-1}\text{gen}^{-1}$ may also be tolerable since, in this case, intact and dissociated TE constructs are present at fairly equal numbers up to 40 years following release (Figure 3B). However, for a dissociation rate of $0.05 \text{ TE}^{-1}\text{gen}^{-1}$, the vast majority of elements are devoid of an intact resistance gene well before the element construct has had a chance to reach equilibrium in the vector population (Figure 3C).

One indicator of an effective disease control strategy is the presence of a high number of disease-resistant vectors over a significant time span. Assuming that only a single copy of the resistance gene is necessary for its expression, the prevalence of disease-resistant vectors is equal to the proportion of vectors having at least one copy of the resistance gene. This quantity is plotted over time in Figure 3D for four different dissociation rates. For default parameters, disease-resistant vectors reach a maximum prevalence greater than 80% for dissociation rates less than 0.013 $TE^{-1}gen^{-1}$ and always reach a maximum prevalence within the first decade of spread.

To put this into perspective, dissociation rates have been measured for *P* elements by screening for the loss of a visible marker gene within the element (Daniels *et al.* 1985; Levis *et al.* 1985; Robertson *et al.* 1988) and are typically on the order of 0.05 $TE^{-1}gen^{-1}$. According to our model, disease-resistant vectors reach a maximum prevalence of 18% at this rate, again suggesting that a dissociation rate of 0.05 $TE^{-1}gen^{-1}$ is unacceptable for public health goals. On the other hand, for a dissociation rate of 0.01 $TE^{-1}gen^{-1}$, disease-resistant vectors reach a maximum prevalence close to 87% and remain at a high prevalence for several decades following release. This confirms that a dissociation rate of 0.01 $TE^{-1}gen^{-1}$ is potentially acceptable for public health goals.

An interesting conclusion that follows from Figure 3D is that, if the prevalence of disease-resistant vectors reaches a significantly high proportion in the vector population, then this prevalence will remain high for several decades following release. This implies that a high maximum prevalence of disease-resistant vectors will also tend to satisfy the requirement of a long duration of prevalence. For the remainder of this paper, we will use the maximum prevalence of disease-resistant vectors as an indicator of an effective disease-control strategy.

The qualitative predictions of this model are relatively insensitive to changes in lesser-known parameters such as the fitness cost of a new TE insert, d_0, the halftime for TE fitness cost to fall, $t_{1/2}$, and the proportion of transgenic mosquitoes released, $x_{(1,0)}(0)$. The release proportion of 2.5%, although motivated by the use of 50 transgenic mosquitoes in a small village, also applies, for example, to the release of 5,000 transgenic mosquitoes in an endemic region with 200,000 wild-type mosquitoes. Increasing the release proportion to 10% decreases the time taken for the resistance trait to reach near-fixation and moderately increases the maximum prevalence that the resistance trait reaches in the population. Decreasing the release proportion to 1% has the converse effect (results not shown).

Transpositional handicap: We extend this simple model of dissociation to the case where the effector gene compromises the rate of replicative transposition (Figure 4A). Using a parameter, δ, to account for the fractional reduction in transposition rate due to the presence of an effector gene, we model the impact of a transpositional handicap on the prevalence of disease-resistant vectors. Even a modest transpositional handicap can significantly reduce the maximum prevalence of disease-resistance. As seen in Figure 4A for a transpositional handicap of $\delta = 0.25$, the dissociation rate must be less than 0.005 TE^{-1}gen^{-1} for the maximum prevalence of disease-resistant vectors to exceed 80%. When the transpositional handicap is $\delta = 0.5$ then the dissociation rate must be less than 0.001 TE^{-1}gen^{-1} to achieve the same maximum prevalence; and when the transpositional handicap is $\delta = 0.95$ then dissociation can essentially not be tolerated.

To put these results into perspective, the RNAi-based strategy for resistance to multiple dengue serotypes currently consists of an effector gene, marker gene and regulatory elements amounting to ~6.5 kb (Kenneth Olson, personal communication). A *Himar1 mariner* element

containing a construct of this size would suffer a reduction in transposition rate of $\delta = 0.97$ (Lampe *et al.* 1998). Typical constructs now used in laboratory experiments are on the order of ~5-6 kb (Marcelo Jacobs-Lorena, personal communication), and would suffer a transpositional handicap of $\delta = 0.95$ if attached to a *Himar1 mariner* transposase gene (Lampe *et al.* 1998). Any construct acceptable for release in the field will need to have at least two or three effector genes in order to counteract evolution of the disease agent and to increase effectiveness of the intervention (Marcelo Jacobs-Lorena, personal communication). These field release constructs will be on the order of ~8-12 kb and, contained within a *Himar1 mariner* transposon, would cause a transpositional handicap of $\delta = 0.99$ or more (Lampe *et al.* 1998).

Although these transposition rate calculations specifically apply to the *Himar1 mariner* element, they clearly make the point that the transpositional handicap of an effector gene is of great concern to any transposon-mediated control strategy. The transpositional properties of any candidate element must be measured in both the presence and absence of a transgene construct, and the size of the construct should be minimized wherever possible, while still satisfying the requirements for effective intervention. As seen in Figure 4A, a transpositional handicap of $\delta = 0.1$ is tolerable for dissociation rates less than 0.01 TE^{-1}gen^{-1}; and higher transpositional handicaps are tolerable for smaller dissociation rates. An additional concern, however, is that gap repair becomes significantly less efficient for segments larger than 11 kb (Johnson-Schlitz and Engels 2006). This provides an additional incentive for minimizing the size of the transgene construct.

Fitness consequences: We consider the impact of a fitness cost associated with the effector gene using a parameter, σ_-, to account for the increase in vector death rate due to the effector gene. As with the case of a transpositional handicap, even a small fitness cost to the host

can significantly reduce the maximum prevalence of disease-resistant vectors. As seen in Figure 4B, when the fitness cost of the effector gene is $\sigma_- = 0.005\ \mathrm{TE^{-1}}$ then the dissociation rate must be less than $0.005\ \mathrm{TE^{-1}gen^{-1}}$ for the maximum prevalence of disease-resistant vectors to exceed 80%. When the fitness cost of the effector gene is $\sigma_- = 0.01\ \mathrm{TE^{-1}}$, the maximum tolerable dissociation rate to reach the same prevalence is $0.001\ \mathrm{TE^{-1}gen^{-1}}$; and when the fitness cost is $\sigma_- = 0.05\ \mathrm{TE^{-1}}$ then effective disease control is virtually impossible.

To account for a selective advantage associated with the effector gene, we introduce a parameter, σ_+, to model the mean increase in mosquito longevity due to the resistance trait. According to our model results (Figure 4C), when the fitness benefit of the effector gene is small ($\sigma_+ = 0.005 - 0.01$) then the dissociation rate required for effective disease control is moderately relaxed. However, when the fitness benefit is large ($\sigma_+ = 0.05$) then disease control becomes feasible for a large range of dissociation rates – the maximum prevalence of disease-resistant vectors exceeds 80% for dissociation rates less than $0.05\ \mathrm{TE^{-1}gen^{-1}}$ and exceeds 75% for dissociation rates less than $0.1\ \mathrm{TE^{-1}gen^{-1}}$.

It is difficult to put these results into perspective due to the lack of reliable comparative fitness measurements in vector populations; however fitness costs have been documented in several insect species due to both mounting an immune response (Moret and Schmid-Hempel 2000; Ahmed *et al.* 2002) and maintaining the physiological machinery necessary to do so (Kraaijeveld and Godfray 1997; Koella and Boete 2002). Ahmed *et al.* (2002) found that egg production was reduced by 18.6% in *An. gambiae* mosquitoes whose immune system was artificially stimulated with lipopolysaccharides. Given that the proportion of mosquitoes infected with malaria parasites varies between 1-20% in Africa (Beier *et al.* 1999), this corresponds to a mean fitness cost of $\sigma_- = 0.002 - 0.04\ \mathrm{TE^{-1}}$, depending on the region of Africa. Measurements

of the fitness cost of immune surveillance are not as well-documented as they are for immune deployment.

These considerations are meaningless if the corresponding selective advantage of the resistance trait is not also considered. Indeed, Hurd *et al.* (2005) found that the fitness cost of resistance to malaria is equal to the selective advantage of avoiding infection, and hence the two factors effectively cancel each other out. If this were the case, then fitness effects could be completely ignored.

A recent encouraging result is that transgenic mosquitoes have been engineered that exhibit no measurable fitness cost when fed on *Plasmodium*-free blood (Moreira *et al.* 2004) and exhibit a fitness benefit when fed on *Plasmodium*-infected blood (Marrelli *et al.* 2007). Although these results were obtained from cage experiments using the rodent malaria parasite *Plasmodium berghei*; it is conceivable that similar results could be obtained for the human malaria parasite *Plasmodium falciparum* and vector *An. gambiae*. Marelli *et al.* (2007) measured a 50% fitness benefit of being transgenic and a 35% fitness cost of being homozygous for the transgene; however this fitness measurement was on a population of mosquitoes that were all exposed to *Plasmodium*. In western Kenya, ~6.3% of *An. gambiae* mosquitoes are infected with *Plasmodium* (Shililu *et al.* 1998); while in northern Ghana, ~11% of *An. gambiae* mosquitoes are infected (Appawu *et al.* 2003). Assuming that a 50% fitness benefit is possible without overdominance, then these mosquito infection rates correspond to a mean selective advantage of $\sigma_+ = 0.03$ and $\sigma_+ = 0.05$ respectively.

Thus far, the effects of transpositional handicaps and fitness costs and benefits have been considered in isolation. However, of considerable interest is whether the selective advantage of resistance could counteract the detrimental effects of a transpositional handicap. In Figure 4D,

we consider the impact of a fitness benefit on the spread of refractoriness when there is a transpositional handicap of $\delta = 0.95$. Interestingly, the maximum prevalence of disease-resistant vectors depends more on the fitness benefit than on the dissociation rate. For a fitness benefit of $\sigma_+ = 0.05$, the maximum prevalence of disease-resistant vectors varies between 76% for a dissociation rate of 0.001 TE^{-1}gen^{-1} and 67% for a dissociation rate of 0.1 TE^{-1}gen^{-1}. For fitness benefits of $\sigma_+ = 0.01$ or less, disease control is not feasible. The implication of this result is that a highly advantageous resistance gene ($\sigma_+ > 0.05$) can spread to a high prevalence in the vector population irrespective of a daunting dissociation rate or transpositional handicap.

DISCUSSION

Several commentaries on the use of TEs to drive resistance genes into vector populations have expressed concern that dissociation will ultimately lead to the resistance gene being lost from the population (Curtis 2003; Hahn and Nuzhdin 2004). One of the key conclusions of this modelling effort is that the epidemiological outcome of the gene drive strategy critically depends on the rate at which the resistance gene dissociates from the drive system. For slow rates of dissociation (0.01 TE^{-1}gen^{-1}) disease control is expected to occur within decades, while for intermediate rates of dissociation (0.05 TE^{-1}gen^{-1}) disease-resistant vectors are predicted to reach a much smaller maximum prevalence (less than 20%) and to be lost within decades. These contrasting scenarios highlight the importance of having good estimates of dissociation rates for the TEs and host species being considered for transgenic release. They also lead to the interesting prediction that, if dissociation inhibits the success of the project, then the resistance gene could be lost from the vector population within a human time frame.

19

Another recommendation of the model is that the size of the effector gene and associated regulatory elements should be restricted in order to limit the transpositional handicap of a TE attached to its transgenic load. While small transpositional handicaps may be tolerable, experiments on the *Himar1 mariner* element indicate that the effector gene constructs currently being developed in the laboratory would suffer a reduction in transposition rate of at least 95% if attached to this element. It is therefore important that, in addition to measuring the dissociation rate of the candidate elements, we also measure their transposition rates both in the presence and absence of the transgene construct. It is not sufficient that a transposable element replicates quickly within a genome – its transposition rate must also be relatively unaffected by exogenous DNA.

The fitness impact of the refractory gene is also of great importance in determining its fate in the vector population. Although a selective advantage may not be sufficient on its own to achieve fixation (Boëte and Koella 2003), when combined with a drive system it greatly improves the chances of success. A large selective advantage of refractoriness is a crucial factor in outweighing the daunting effects of dissociation and reduction in transposition rate. It is therefore essential to have comparative measurements of the fitness of both transgenic and wild-type mosquitoes both exposed and unexposed to the disease agent. If a transgenic release becomes more feasible, it will also be essential to make these measurements under taxing field conditions using the human vector and disease agent, and to develop innovative ways for differentiating these effects from inbreeding depression and the process of transgenesis. Such measurements are not currently available.

Model limitations: Symptomatic of any mathematical analysis, simplifications have been made in this model that may compromise its predictions. Firstly, the choice of *Herves* as a

model element that has an average copy number of four (Subramanian *et al.* 2007) is not representative of the vast number of TEs that are present at much higher copy numbers (Charlesworth *et al.* 1992). This choice was motivated in part by the fact that *Herves* rarely undergoes internal deletion, and also by the fact that model complexity greatly increases with element copy number. To assess the dynamics of TEs with higher copy numbers, another modeling framework should be proposed.

Secondly, in the model formulation all genomic elements are assumed to act independently during gamete formation, implying that all element copies are at least 50 centimorgans apart. This contradicts the tendency for TEs to jump locally rather than distally (Tower *et al.* 1993; Zhang and Spradling 1993; Guimond *et al.* 2003) and to home in on certain genomic regions (Guimond *et al.* 2003). This is an important consideration since clustering of element copies can reduce the rate of element spread (Rasgon and Gould 2005); however in order to focus on the impact of dissociation, this consideration is beyond the scope of the present study.

Future directions: The model presented here is of special relevance to diseases such as malaria and dengue fever that are currently being considered for transposon-mediated control strategies (Scott *et al.* 2002). The present study could be supplemented by imbedding the spread of the resistance gene within a model of the epidemiology of malaria or dengue fever (Boëte and Koella 2002). This would give us a better idea of the exposure rate of vectors to the malaria parasite or dengue virus. It would also allow us to better predict the maximum prevalence of disease-resistant vectors and hence the rate of dissociation that is required for effective disease control.

ACKNOWLEDGEMENTS

LITERATURE CITED

Appawu, M. A., K. M. Bosompem, S. Dadzie, U. S. McKakpo, I. Anim-Baidoo *et al.*, 2003
Detection of malaria sporozoites by standard ELISA and VecTest[TM] dipstick assay in
field-collected anopheline mosquitoes from a malaria endemic site in Ghana. Tropical
Medicine and International Health **8**: 1012-1017.

Ahmed, A. M., S. L. Baggott, R. Maingon, and H. Hurd, 2002 The costs of mounting an immune
response are reflected in the reproductive fitness of the mosquito *Anopheles gambiae*,
Oikos **97**: 371-377.

Beier, J. C., G. F. Killeen, and J. I. Githure, 1999 Short report: entomologic inoculation rates and
Plasmodium falciparum malaria prevalence in Africa. Am. J. Trop. Med. Hyg. **61**: 109-
113.

Boëte, C., and J. C. Koella, 2002 A theoretical approach to predicting the success of genetic
manipulation of malaria mosquitoes in malaria control. Malaria Journal. **1 (3)**.

Boëte, C., and J. C. Koella, 2003 A model for coevolution of immunity and immune evasion in
vector-borne diseases with implications for the epidemiology of malaria. Am. Nat. **161**:
698-707.

Charlesworth, B., 1991 Transposable elements in natural populations with a mixture of selected
and neutral sites. Genet. Res. **57**: 127-134.

Charlesworth, B., and D. Charlesworth, 1983 The population dynamics of transposable elements.
Genet. Res. **42**: 1-27.

Charlesworth, B., A. Lapid, and D. Canada, 1992 The distribution of transposable elements
within and between chromosomes in a population of *Drosophila melanogaster*. I.
Element frequencies and distribution. Genet. Res. **60**: 103-114.

Charlesworth, B., P. Sniegowski, and W. Stephan, 1994 The evolutionary dynamics of repetitive DNA in eukaryotes. Nature **371**: 215-220.

Craig, G. B., 1963 Prospects for vector control through manipulation of populations. Bull. World Health Organization. **29**: 89-97.

Crow, J. F., and M. J. Simmons, 1983 The mutation load in *Drosophila*, pp. 1-35 in *The genetics and biology of Drosophila vol. 3c*, edited by M. Ashburner, H. L. Carson and J. N. Thompson. Academic Press, London.

Curtis, C. F., 1968 Possible use of translocations to fix desirable genes in insect pest populations. Nature **218**: 368-369.

Curtis, C. F., 2003 Measuring public-health outcomes of release of transgenic mosquitoes, pp. 223-234 in *Ecological Aspects for Application of Genetically Modified Mosquitoes*, edited by W. Takken and T. W. Scott. Wageningen, The Netherlands.

Daniels, S. B., M. McCarron, C. Love, and A. Chovnick, 1985 Dysgenesis-induced instability of *rosy* locus transformation in *Drosophila melanogaster*: analysis of excision events and the selective recovery of control element deletions. Genetics **109**: 95-117.

Engels, W. R., 1989 *P* elements in *Drosophila melanogaster*, pp. 439-484 in *Mobile DNA*, edited by D. E. Berg and M. M. How. ASM Press, Washington DC.

Gelbart, W. M., and R. K. Blackman, 1989 The *hobo* element of *Drosophila melanogaster*. Prog. Nucleic Acid Res. Mol. Bio. **36**: 37-46.

Guimond, N., D. K. Bideshi, A. D. Pinkerton, P. W. Atkinson, and D. A. O'Brochta, 2003 Patterns of *Hermes* transposition in *Drosophila melanogaster*. Molec. Gen. Genet. **268**: 779-790.

Hahn, M. W., and S. V. Nuzhdin, 2004 The fixation of malaria refractoriness in mosquitoes. Curr. Biol. **14**: 264-265.

Hurd, H., P. J. Taylor, D. Adams, A. Underhill, and P. Eggleston, 2005 Evaluating the costs of mosquito resistance to malaria parasites. Evolution **59**: 2560-2572.

James, A. A., 2005 Gene drive systems in mosquitoes: rules of the road. Trends Parasitol. **21**: 64-67.

Johnson-Schlitz, D. M., and W. R. Engels, 2006 The effect of gap length on double-strand break repair in *Drosophila*. Genetics **175**: 2033-2038.

Knols, B. G. J., and W. Scott, 2003 Ecological challenges concerning the use of genetically modified mosquitoes for disease control: synthesis and future perspectives, pp. 235-242 in *Ecological Aspects for Application of Genetically Modified Mosquitoes*, edited by W. Takken and T. W. Scott. Wageningen, The Netherlands.

Koella, J. C., and C. Boëte, 2002 A genetic correlation between age at pupation and melanization immune response of the yellow fever mosquito *Aedes aegypti.* Evolution **56**: 1074-1079.

Kraaijeveld, A. R., and H. C. Godfray, 1997 Trade-off between parasitoid resistance and larval competitive ability in *Drosophila melanogaster*. Nature **389**: 278-280.

Lampe, D. J., T. E. Grant, and H. M. Robertson, 1998 Factors affecting the transposition of the *Himar1 mariner* transposon *in vitro*. Genetics **149**: 179-187.

Langley, C. H., E. A. Montgomery, R. Hudson, N. Kaplan, and B. Charlesworth, 1988 On the role of unequal exchange in the containment of transposable element copy number. Genet. Res. **52**: 223-236.

Le Rouzic, A., and P. Capy, 2005 The first steps of a transposable element invasion: parasitic strategy vs. drift. Genetics **169**: 1033-1043.

Le Rouzic, A., and P. Capy, 2006 Reversible introduction of transgenes in natural populations of insects. Insect Mol. Biol. **15**: 227-234.

Levis, R., T. Hazelrigg, and G. M. Rubin, 1985 Effects of genomic position on the expression of transduced copies of the white gene of *Drosophila*. Science **229**: 558-561.

Lohe, A. R., C. Timmons, I. Beerman, E. R. Lozovskaya, and D. L. Hartl, 2000 Self-inflicted wounds, template-directed gap repair and a recombination hotspot: Effects of *mariner* transposase. Genetics **154**: 647-656.

Mackay, T. F., R. F. Lyman, and M. S. Jackson, 1992 Effects of *P* element insertions on quantitative traits in *Drosophila melanogaster*. Genetics **130**: 315-332.

Marrelli, M. T., C. Li, J. L. Rasgon, and M. Jacobs-Lorena, 2007 Transgenic malaria-resistant mosquitoes have a fitness advantage when feeding on *Plasmodium*-infected blood. Proc. Natl. Acad. Sci. USA **104**: 5580-5583.

Maside, X., S. Assimacopoulos, and B. Charlesworth, 2000 Rates of movement of transposable elements on the second chromosome of *Drosophila melanogaster*. Genet. Res. **75**: 275-284.

Moreira, L. A., J. Wang, F. H. Collins, and M. Jacobs-Lorena, 2004 Fitness of anopheline mosquitoes expressing transgenes that inhibit *Plasmodium* development. Genetics **166**: 1337-1341.

Moret, Y., and P. Schmid-Hempel, 2000 Survival for immunity: The price of immune system activation for bumblebee workers. Science **290**: 1166-1168.

Mukai, T., S. I. Chigusa, L. E. Mettler, and J. F. Crow, 1972 Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. Genetics **72**: 335-355.

Nuzhdin, S. V., 2000 Sure facts, speculations, and open questions about the evolution of

 transposable element copy number. Genetica **107**: 129-137.

Nuzhdin, S. V., and T. F. Mackay, 1994 Direct determination of retrotransposon transposition

 rates in *Drosophila melanogaster*. Genet. Res. **63**: 139-144.

Nuzhdin, S. V., E. G. Pasyukova, and T. F. C. Mackay, 1997 Accumulation of transposable

 elements in laboratory lines of *Drosophila melanogaster*. Genetica. **100**: 167-175.

Ohnishi, O., 1977 Spontaneous and ethyl methane-sulfonate-induced mutations controlling

 variability in *Drosophila melanogaster* I: Recessive lethal mutations. Genetics **87**: 335-

 348.

Pasyukova, E. G., S. V. Nuzhdin, T. V. Morozova, and T. F. Mackay, 2004 Accumulation of

 transposable elements in the genome of *Drosophila melanogaster* is associated with a

 decrease in fitness. J. Hered. **95**: 284-290.

Rasgon, J. L., and F. Gould, 2005 Transposable element insertion location bias and the dynamics

 of gene drive in mosquito populations. Insect Mol. Biol. **14**: 493-500.

Riehle, M. A., P. Srinivasan, C. K. Moreira, and M. Jacobs-Lorena, 2003 Towards genetic

 manipulation of wild mosquito populations to combat malaria: advances and challenges. J

 Exp. Biol. **206**: 3809-3816.

Robertson, H. M., C. R. Preston, R. W. Phillis, D. M. Johnson-Schlitz, W. K. Benz *et al.*, 1988 A

 stable genomic source of *P* element transposase in *Drosophila melanogaster*. Genetics

 118: 461-470.

Rubin, E., and A. Levy, 1997 Abortive gap repair: underlying mechanism for *Ds* element

 formation. Molec. Cell Biol. **17**: 6294-6304.

Scott, T. W., W. Takken, B. G. Knols, and C. Boëte, 2002 The ecology of genetically modified mosquitoes. Science **298**: 117-119.

Seleme, M., I. Busseau, S. Malinsky, A. Bucheton, and D. Teninges, 1999 High-frequency retrotransposition of a marked *I* factor in *Drosophila melanogaster* correlates with a dynamic expression pattern of the ORF1 protein in the cytoplasm of oocytes. Genetics **151**: 761-771.

Shililu, J. I., W. A. Maier, H. M. Seitz, and A. S. Orago, 1998 Seasonal density, sporozoite rates and entomological inoculation rates of *Anopheles gambiae* and *Anopheles funestus* in a high-altitude sugarcane growing zone in western Kenya. Tropical Medicine and International Health **3**: 706-710.

Spradling, A. C., 1986 *P* element-mediated transformation, pp. 175-198 in *Drosophila: A practical approach*, edited by D. B. Roberts. IRL Press, Oxford.

Subramanian, R. A., P. Arensburger, P. W. Atkinson, and D. A. O'Brochta, 2007 Transposable element dynamics of the *hAT* element *Herves* in the human malaria vector *Anopheles gambiae* s.s. Genetics **176**: 2477-2487.

Taylor, C., Y. T. Touré, J. Carnahan, D. E. Norris, G. Dolo *et al.*, 2001 Gene flow among populations of the Malaria vector *Anopholes gambiae* in Mali, West Africa. Genetics **157**: 743-750.

Tower, J., G. H. Karpen, N. Craig, and A. C. Spradling, 1993 Preferential transposition of *Drosophila P* elements to nearby chromosomal sites. Genetics **133**: 347-359.

Townsend, J. P., and D. L. Hartl, 2000 The kinetics of transposable element autoregulation. Genetica **108**: 229-237.

Vasilyeva, L. A., E. V. Bubenshchikova, and V. A. Ratner, 1999 Heavy heat shock induced

retrotransposon transposition in *Drosophila*. Genet. Res. **74**: 111-119.

Zhang, P., and A. C. Spradling, 1993 Efficient and dispersed local *P* element transposition from

Drosophila females. Genetics **133**: 361-373.

TABLE 1

Parameter values, ranges and source references

Parameter	Definition	Estimated value	Range	References
u_1	Replicative transposition rate	0.1 TE^{-1}gen^{-1}	–	Seleme *et al.* (1999), Vasilyeva *et al.* (1999)
c	Transpositional regulation parameter	2.866 TE^{-1}	–	Subramanian *et al.* (2007), Townsend and Hartl (2000)
v	Element deletion rate	4×10^{-6} TE^{-1}gen^{-1}	–	Nuzhdin *et al.* (1997), Maside *et al.* (2000)
d_0	Initial TE fitness cost	0.02 TE^{-1}	[0.01, 0.1] TE^{-1}	Crow and Simmons (1983), Mackay *et al.* (1992)
d_∞	Asymptotic TE fitness cost	10^{-3} TE^{-1}	–	Nuzhdin *et al.* (1997), Pasyukova *et al.* (2004)
$t_{1/2}$	Halftime of TE fitness cost decline	5 gen	[2, 20] gen	Charlesworth (1991)
p	Rate of decrease in vector fitness	1.5	–	Charlesworth and Charlesworth (1983)
w^*	Dissociation rate	0.05 TE^{-1}gen^{-1}	[0, 0.1] TE^{-1}gen^{-1}	Levis *et al.* (1985), Robertson *et al.* (1988)
δ	Transpositional handicap	0.95	[0, 0.99]	Lampe *et al.* (1998); Spradling (1986)

σ_-	Fitness cost of resistance gene	–	[0, 0.05] TE^{-1}	Ahmed *et al.* (2002), Beier *et al.* (1999)
σ_+	Fitness benefit of resistance gene	–	[0, 0.05]	Marrelli *et al.* (2007), Appawu *et al.* (2003)
$x_{(1,0)}$	Initial proportion of transgenic vectors	0.025	[0.01. 0.1]	Taylor *et al.* (2001)

FIGURE LEGENDS

Figure 1. Abortive gap repair mechanism of internal deletion. (A) An intact TE construct consists, at the very least, of a transposase gene (Tr), a disease resistance gene (R) and a pair of inverted repeats (IR) marking its boundaries. (B) Following transposition or deletion, a double-stranded gap is introduced into the host chromosome DNA. (C) This gap is sometimes filled by copying information from a homologous chromosome, sister chromatid or ectopic chromosomal site containing the intact construct. (D) If this process is interrupted, then the central portions of the TE will not be copied. In some cases, this produces a dissociated construct not containing the resistance gene.

Figure 2. Schematic for the spread of an element construct incorporating dissociation of the resistance gene. An individual having two copies of the intact construct (TER) and two copies of the dissociated construct (TE) can gain a copy of either through transposition and lose a copy of either through deletion. The proportion of individuals having this genotype can also increase or decrease as a result of transposition or deletion. Dissociation is a one-way process leading to a TER becoming a TE. Analogous schematics can be drawn for every other genotype.

Figure 3. Element spread with dissociation. The resistance gene has no impact on host fitness or transposition rate. (A-C) Total element copy number reaches an equilibrium of four within 20 years, while the proportion of intact element copies decrease monotonically over time. Lines correspond to numerical solutions of the differential equation model for three different dissociation rates. (D) The prevalence of the disease-resistant vectors reaches a maximum within ten years and then decreases gradually. The maximum prevalence and duration of presence depend largely on dissociation rate.

Figure 4. Maximum prevalence of disease-resistant vectors. The resistance gene influences both transposition rate and host fitness. (A) The replicative transposition rate of an intact element construct is reduced by a fraction, δ. Increasing the value of δ decreases the maximum prevalence of resistant vectors and makes dissociation requirements more demanding. (B) The fitness cost conferred by the resistance gene, σ_-, has a large impact on its maximum prevalence. (C) Dissociation requirements are considerably relaxed when the resistance gene confers a fitness benefit, σ_+. (D) A highly advantageous resistance gene is able to spread irrespective of a daunting dissociation rate or transpositional handicap.
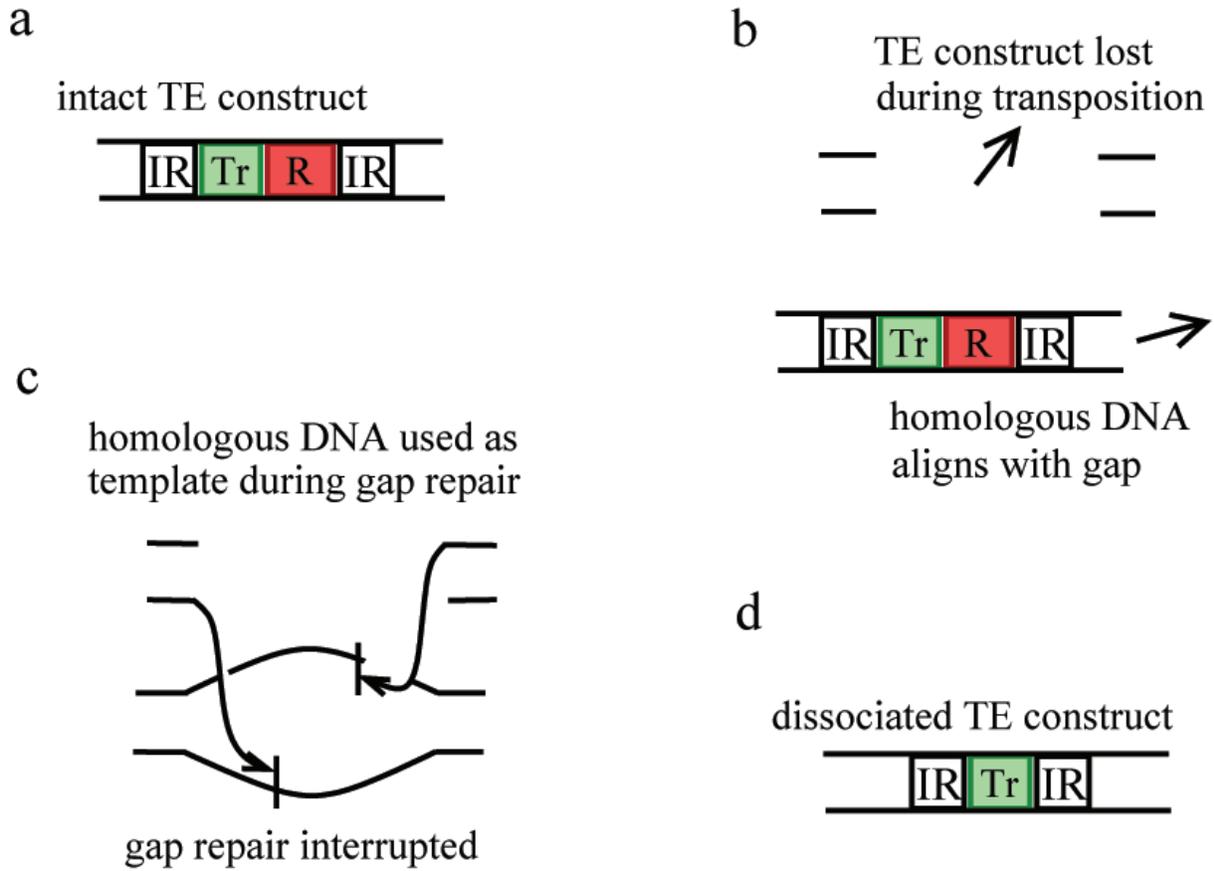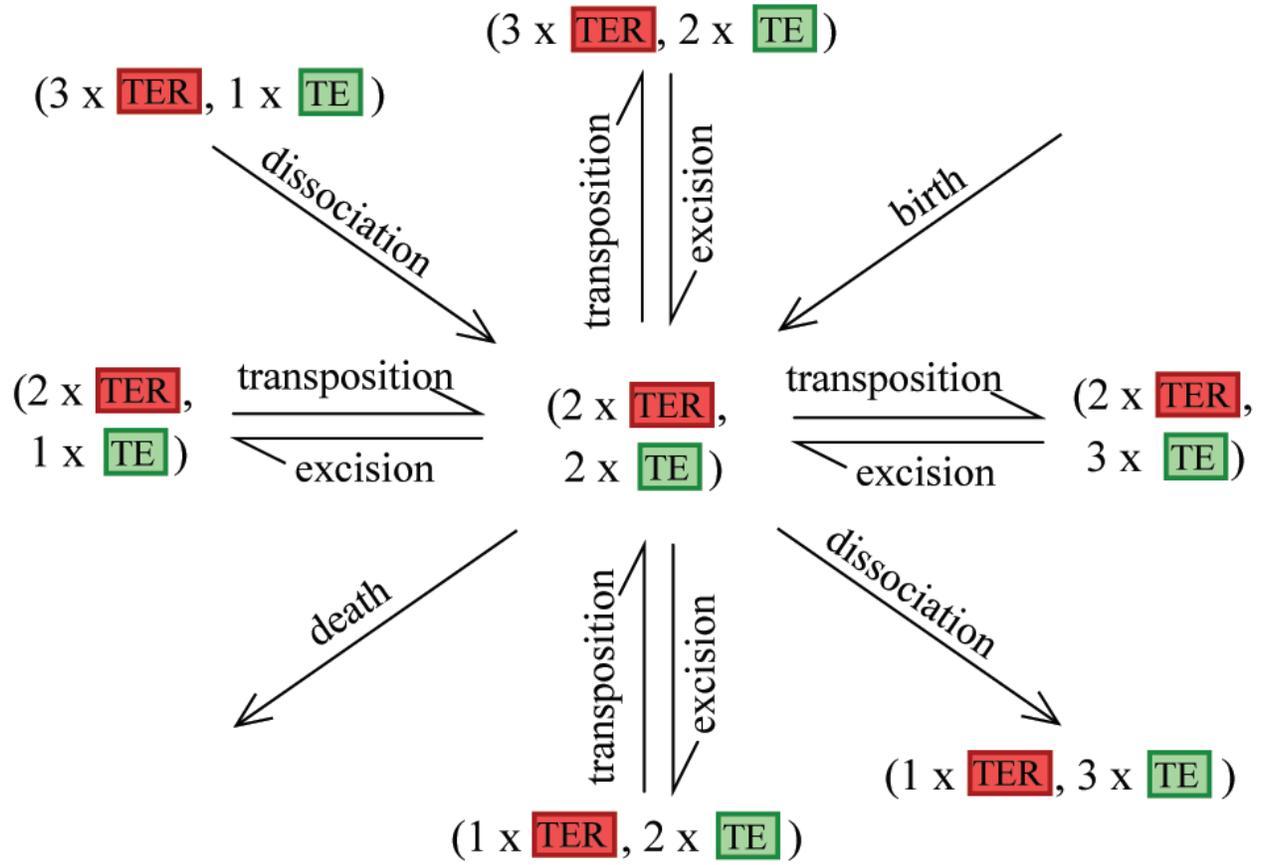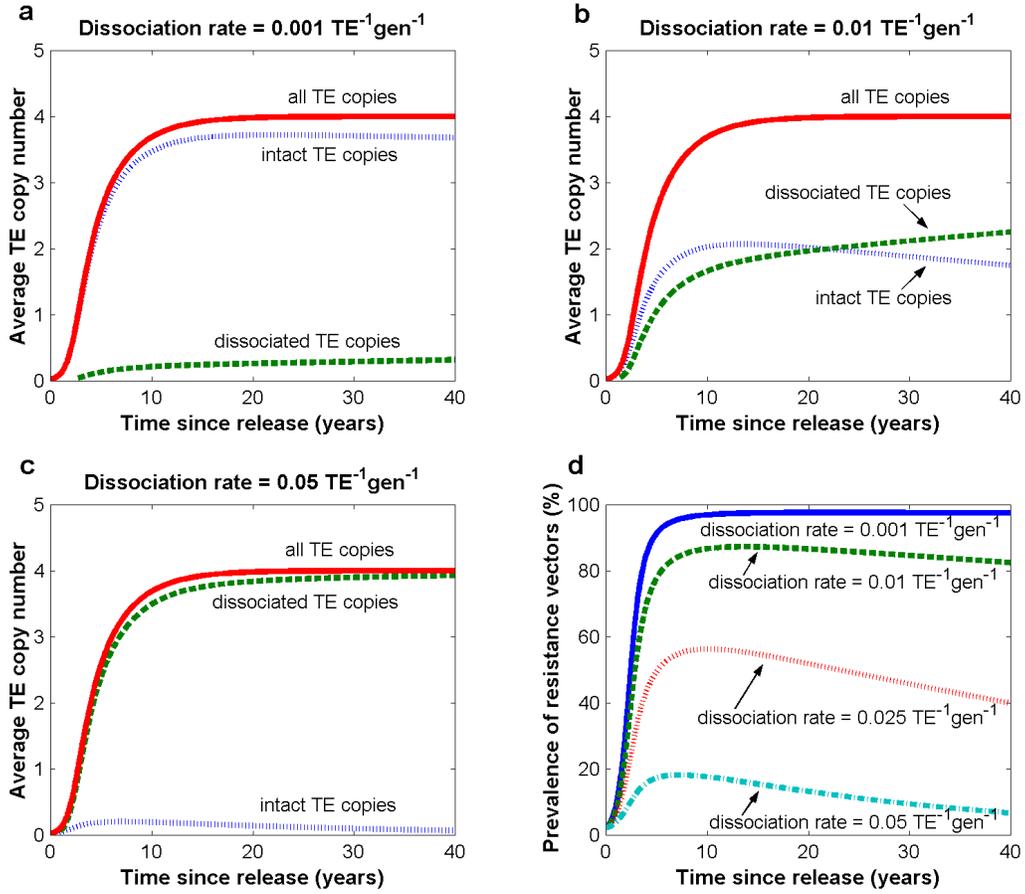
Figure 1

a

intact TE construct

IR | Tr | R | IR

b

TE construct lost
during transposition

IR | Tr | R | IR

homologous DNA
aligns with gap

c

homologous DNA used as
template during gap repair

gap repair interrupted

d

dissociated TE construct

IR | Tr | IR

Figure 2

Figure 3

Figure 4