J. Math. Biol. (2008) 57:811–840
DOI 10.1007/s00285-008-0190-2

Mathematical Biology

A branching process for the early spread of a transposable element in a diploid population

John M. Marshall

Received: 18 May 2007 / Revised: 1 May 2008 / Published online: 29 May 2008

Abstract Transposable elements (TEs) face significant challenges upon transfer into a new host population, invariably beginning their invasion with only a single element. The fate of this element is a product of its internal properties, the population dynamics of the host species, and genetic drift. We present a continuous-time multi-type branching process to model the early stages of TE spread. The model incorporates seasonal population size changes and is applicable to diploid hosts for prevalences up to 10%. We reproduce standard results of TE population dynamics and show that population growth may have a greater influence on reducing TE loss probability than a transpositional burst. These results are applied to the planned use of a TE to drive an antimalarial gene into an *Anopheles gambiae* population. The model favors a transgenic release immediately following the dry season when the *An. gambiae* population begins to grow. Increasing the number of transgenic hosts released has the greatest influence on reducing the probability of TE loss. Following release, the rate at which the TE increases its proportion in the population is most sensitive to its replicative transposition rate. The model recommends a replicative transposition rate greater than 0.1 per TE per generation to satisfy public health goals.

Keywords Transposable element · Branching process · Population growth · Anopheles gambiae · Malaria

This research was supported by grant numbers R01AI51633 and 1R56AI072549 from the National Institutes of Health. A portion of this work was originally presented at the 55th annual meeting of the American Society of Tropical Medicine and Hygiene.

J. M. Marshall (✉)
Department of Biomathematics, UCLA School of Medicine,
Los Angeles, CA 90095-1766, USA
e-mail: johnmm@ucla.edu

Mathematics Subject Classification (2000) 60J85 · 92D25

Abbreviations

TE Transposable element

1 Introduction

Transposable elements (TEs) are particularly interesting genomic components due to their ability to transpose replicatively within a genome and hence spread throughout a population despite a fitness cost. The replicative ability of TEs has led to their widespread prevalence in the genomes of many taxa to the extent that various families of TEs account for ∼90% of the Salamander genome [1] and ∼45% of the human genome [2]. In some cases, studies have shown TE spread to be very rapid. The *P* element is perhaps the best example of this, having spread through most of the wild-type *Drosophila melanogaster* population in the span of a few decades [3].

Observations like these have inspired the idea of using TEs as drive mechanisms for spreading disease resistance genes into vector populations [4,5]. In recent years, advances in molecular biology [6,7] and ecology [8] have allowed this idea to become a feasible control strategy for vector-borne disease. Hence it is of great epidemiological interest to have some idea of the molecular and ecological conditions under which a TE will spread through a host population.

Models of TE dynamics have tended to focus on factors affecting element spread at the level of the genome and the host individual. Here, the replicative ability of the TE is weighed against its fitness cost to the host, its excision rate, and the amount by which transposition is suppressed with increasing copy number. Accounting for these factors, a number of models have been proposed to study the evolution and distribution of TEs in general [9–13]. These analyses have tended to focus on equilibrium distributions, comparing them against the distribution of TE copy number in nature.

The conditions necessary for TE spread have been investigated in a number of studies that have used a meiotic drive parameter to characterize the departure from Hardy-Weinberg equilibrium due to the presence of a TE [14–16]. Other properties of TEs, such as the maximum transposition rate and the various mechanisms of transpositional regulation, have been investigated by Le Rouzic and Capy [17] and Struchiner et al. [18]. Additionally, the tendency for TEs to transpose locally rather than distally has been addressed in simulations by Rasgon and Gould [19]. At the level of the population, Deceliere et al. [20] have modeled the effects of migration and demographic history on TE dynamics; and a simulation environment has been developed to integrate the demographic and molecular considerations [21].

In this paper, we apply a branching process model to examine the early stages of a TE invasion. Although our model has similar structural details to previously-published branching process models [11,12,22], it differs in that it is applicable to diploid hosts. One of the places where haploid models of TE spread break down for diploid organisms is in the rate at which uninfected hosts become infected with a single TE. For haploid hosts, de novo TE acquisition may occur by transduction, transformation or conjugation, and is assumed to occur at a constant rate [11,12,22]

Uninfected diploid hosts, on the other hand, give birth to TE-infected offspring by mating with a TE-infected host. This rate increases as the number of infected hosts and their mean TE copy number increase.

While haploid branching process models have focused on equilibrium distributions of TE copy number [11,12,22]; the diploid branching process proposed here is particularly well-suited to calculating TE loss probabilities and explicitly examining the effects of population size changes on TE spread. The motivation for these extensions is the planned use of a TE to drive an antimalarial gene into an *Anopheles gambiae* mosquito population in Africa. Despite this focus, the predictions of the model apply generally and similar strategies are currently being considered for the control of dengue in *Aedes aegypti* [8].

An. gambiae is the main vector of malaria in tropical Africa, and consequently it is the primary species being considered to host a TE for the purpose of malaria control. A major feature of the demography of *An. gambiae* is the existence of population size changes within and between years [23,24]. The most dramatic of these occur between the dry season and the peak of the rainy season when the population size may change by several orders of magnitude within six months [24]. The population is also structured chromosomally by the existence of up to five chromosomal forms that may be partially or totally reproductively isolated [25,26]; and geographically by its concentration in discrete patches corresponding to villages [27].

We will restrict our attention to the spread of a TE through a single chromosomal form of *An. gambiae* in a single village. This will allow us to focus on the effects of temporal population structure on the release strategy. A successful transgenic release on the village scale must have two basic properties. Firstly, it must be efficient in establishing the effector gene in the host population [28]; and secondly, the drive system must work within a time frame acceptable to public health goals [29]. We will address these requirements by investigating the conditions that minimize the probability of TE loss and maximize the rate of TE spread. Within this context, we will make recommendations regarding the parameters that will be required in order to effectively control vector-borne disease.

2 Model formulation

We use a continuous-time multi-type branching process to model the early stages of TE spread through a randomly mating host population. Particles in the model are of T types corresponding to hosts infected with i copies of the TE, where $i \in \{1, 2, \ldots, T\}$ and $T \geq 1$. Here, T can be approximated as the number of sites that will be occupied in the early stages of TE spread. A host having i TE copies may also be referred to as a "type-i" host. Uninfected hosts are not kept track of in this model because the majority of individuals belong to what may be thought of as a reservoir of uninfected hosts.

2.1 Reproduction

Mating between organisms cannot be modeled explicitly within the confines of a branching process model due to the requirement that the particles in a branching

process must be independent [30]. However, since the vast majority of individuals are uninfected in the early stages of spread, we can imagine that all mating events involving infected hosts will be with individuals from the reservoir of uninfected hosts. This approximation is valid up to a prevalence of about 10% since, in a randomly mating population, less than 1% of all matings will involve two infected hosts at these prevalences. Mating between two infected hosts is modeled in Sect. 4.5 for prevalences greater than 10%.

We consider a budding model in which hosts do not die when they have offspring. This enables us to separate the birth and death rates for each host type. Here, hosts of type-i are assumed to mate with uninfected host organisms from the reservoir to give rise to offspring at a constant rate θ. The number of TEs in the offspring's genome is then determined by: (a) whether a replicative transposition or element deletion event occurred in the cell that gave rise to the gamete contributed by the infected host; and (b) the number of TEs in this diploid cell that are passed on to the haploid gamete during meiosis.

2.2 Transposition and deletion

Transposition and deletion are modeled by assuming that a proportion α_i of gametes are derived from cells in which a replicative transposition event has occurred, while a proportion β_i of gametes are derived from cells in which an element deletion event has occurred. The replicative transposition rate for a type-i host, α_i, is equal to the replicative transposition rate per TE in a type-i host, u_i, multiplied by the number of TEs in the host genome, i (i.e. $\alpha_i = iu_i$). Here, u_i is generally a decreasing function of i to account for suppression of transposition with increasing copy number [31–33]. Similarly, the deletion rate for a type-i host, β_i, is equal to the deletion rate per element, v, multiplied by the number of elements in the host genome, i (i.e. $\beta_i = iv$), where v is generally considered as a constant.

A number of alternative models for the relationship between transposition rate and TE copy number were considered to assess their influence on the early spread of the TE. In each of these models, a represents the transposition rate in a genome containing a single TE (i.e. $u_1 = a$). Our default model describes the replicative transposition rate, u_i, as a linear function of i,

$$u_i = a(1 - b(i - 1)), \tag{1}$$

where b is the fraction by which transposition rate falls off with each additional TE copy. The no transpositional regulation model corresponds to the linear transpositional regulation model with the parameter b set to zero (i.e. $u_i = a$). The threshold regulation model [17] describes the case where no regulation occurs until a certain critical TE copy number is reached, i_t, after which transposition is suppressed by a constant amount, $a - a_r$, i.e.

$$u_i = \begin{cases} a, i \leq i_t \\ a_r, i > i_t \end{cases}, \quad a > a_r. \tag{2}$$

Finally, the continuous regulation model [17] describes the case where each additional TE copy reduces the transposition rate by a smaller increment, reaching a transposition rate of $(a - a_r)/2$ for a TE copy number of i_c, and finally converging to a minimum transposition rate of a_r as copy number becomes very large, i.e.

$$u_i = (a - a_r)2^{-(i-1)/(i_c-1)} + a_r. \tag{3}$$

Equation 3 is qualitatively similar to the hyperbolic equation for transposition rate as a function of TE copy number proposed by Charlesworth and Charleswoth [9]. The continuous regulation model is the commonly used when self-regulation of transposition is considered [34] and could be a consequence of the production of repressors [35] or overproduction inhibition [36]. The threshold regulation model is less commonly used, although a number of molecular mechanisms have been proposed for its existence [17].

2.3 Host fitness

The fitness cost associated with additional TE copies is modeled by varying the death rate, μ_i, according to the number of TE copies, i, that the host genome contains. Here, μ_i is an arbitrary increasing function of i. A number of alternative models for the relationship between death rate and TE copy number were considered to assess their influence on the early spread of the TE. In each of these models, the death rate of a host containing no copies of the TE is set to 1, and d represents the increase in death rate for a host containing a single TE in its genome (i.e. $\mu_1 = 1 + d$). Our default model describes the death rate, μ_i, as a linear function of i,

$$\mu = 1 + di. \tag{4}$$

This model suggests that each additional genomic insertion has an equal and additive effect on host fitness during the early stages of TE spread, possibly due to the effects of insertional mutagenesis [37]. The neutral insertion model is a special case of this model with the parameter d set to zero, i.e. $\mu_i = 1$. The independent fitness cost model [9] describes the case where each additional TE copy increases the host death rate by an identical factor, $1 + d$, i.e.

$$\mu_i = (1 + d)^i. \tag{5}$$

Finally, the log-concave fitness cost model [9] is similar to the linear fitness cost model with the exception that TE copy number is raised to the power p, where $1 < p < 2$, i.e.

$$\mu_i = 1 + di^p. \tag{6}$$

In both the independent fitness cost model and the log-concave fitness cost model, the increase in host death rate rises with each additional TE copy. This is desirable since additional fitness costs arise at higher copy numbers, for example Montgomery et al. [38]

have proposed that ectopic recombination can occur between TEs of the same family present at different sites, and Brookfield [39,40] has proposed that fitness costs arise primarily through the act of transposition at higher copy numbers. The log-concave fitness cost model is favored by Charlesworth and Charlesworth [9] since it allows an equilibrium distribution of copy number to be obtained in their model of the population dynamics of TEs.

2.4 Gamete formation

The number of TEs in the haploid gamete is determined by the number of TEs in the diploid cell that are passed on during meiosis. For a diploid cell with i copies of the TE we assume, to a first approximation, that all of these TEs are far enough apart from each other that they segregate independently. Under this assumption, the probability of having j copies in a gamete is proportional to the number of ways of choosing j elements from a total of i. Similarly, if a replicative transposition event has occurred in the diploid cell, then the probability of having j copies in the gamete is proportional to the number of ways of choosing j elements from $i+1$, or from $i-1$ if a deletion event has occurred. The number of TEs in the offspring's genome is equal to the number of TEs in the haploid gamete from the infected host because the other parent is uninfected by the TE so contributes no elements to the offspring's genome.

Putting this all together within the framework of a multi-type continuous-time branching process [30,41], we have a finite number of independently acting infected hosts of T types that reproduce and die. Each host having i element copies lives an exponentially distributed length of time with death intensity $\lambda_i = \mu_i + \theta$, and at the end of its life produces on average f_{ij} hosts with j element copies according to the equation,

$$\begin{aligned} f_{ij} = &\frac{1}{2^{i-1}}\binom{i-1}{j}\frac{\theta}{\lambda_i}\beta_i + \frac{1}{2^i}\binom{i}{j}\frac{\theta}{\lambda_i}(1-\alpha_i-\beta_i) \\ &+\frac{1}{2^{i+1}}\binom{i+1}{j}\frac{\theta}{\lambda_i}\alpha_i + \frac{\theta}{\lambda_i}1_{\{i=j\}}, \end{aligned} \tag{7}$$

where $i, j \in \{1, 2, \ldots, T\}$.

2.5 Population size changes

Growth and decline in the total population size can be incorporated within the framework of the branching process simply by altering the value of the birth rate parameter, θ. This is related to the population growth rate of the reservoir of uninfected hosts, r, by the following equation,

$$\theta = 2(1+r). \tag{8}$$

The population size of uninfected hosts is then modeled using the exponential growth model,

$$N_t = N_0 e^{rt}, \tag{9}$$

where N_0 is the initial population size, N_t is the population size at time t, and time is measured in generations. We chose an exponential model based on field measurements taken by Touré et al. [42] and modeled by Taylor and Manoukis [43]. For exponential population growth, r should be positive; while for exponential population decline, r should be negative. For a fluctuating population size, the sign of r should be time-dependent and oscillate in sign.

3 Parameter values

Table 1 contains the default parameter values for our model. At present there is little or no data regarding the post-integration behavior of the candidate TEs in *An. gambiae*, so most of these values are taken from measurements in other species, in particular *D. melanogaster* [44–48]. While we cannot assume equivalence between parameter

Table 1 Parameter values, ranges and source references

Parameter	Definition	Estimated value	Range	Main references
a	Replicative transposition rate	0.1 TE^{-1} gen^{-1}	$[10^{-5}, 0.3]$	[49,53]
a_r	Repressed transposition rate	10^{-4} TE^{-1} gen^{-1}	$[10^{-5}, 10^{-3}]$	[17,62]
b	Transpositional regulation parameter	0.1	$[-0.1, 0.1]$	[56,60]
i_t	Threshold TE copy number	5	[2,5]	[17,56]
i_c	TE copy number when transpositional repression is 50%	3	[2,5]	[17,56]
v	TE deletion rate	4×10^{-6} TE^{-1} gen^{-1}	–	[45,46]
d	Fitness cost of TE	0.02 TE^{-1}	$[10^{-5}, 0.4]$	[44,95]
p	Power determining rate of decrease in host fitness	1.5	[1,2]	[9]
r	Population growth rate	0.2 gen^{-1}	$[-0.4, 0.4]$	[23,24]
N_0	Initial population size	2×10^3	$[2 \times 10^3, 2 \times 10^6]$	[23,79]
n_r	Release size	50	$[1, 10^3]$	Charles Taylor (personal communication)
i_r	Copy number at release	1	[1,4]	–
t_r	Time of release	0 gen	[0, 22.8]	Mahamoudou Touré (personal communication)

estimates in *An. gambiae* and *D. melanogaster*; both are insects of the Order Diptera, and it is hoped that data from *D. melanogaster* will provide initial estimates from which to explore a reasonable parameter space for *An. gambaie*.

3.1 Transposition

The replicative transposition rate with only a single TE in the genome, a, has been estimated indirectly by sampling chromosomes from natural populations [45,49,50], and directly by scoring insertion sites in laboratory populations and rescoring after many generations [46]. This has led to estimated replicative transposition rates ranging from $\sim 10^{-5}$ per element per generation for many TEs in *D. melanogaster* [49,50] and *D. simulans* [51,52] up to transposition rates greater than 0.1 per element per generation for very active individual *I* elements [53,54].

The upper bound of the replicative transposition rate, a, was inflated in our analysis to account for the fact that most replicative transposition rates in the literature are estimated from host genomes with copy numbers much greater than one. This underestimates the transposition rate in genomes having a single TE since transposition tends to be suppressed at higher copy numbers and some TEs tend to show bursts of transpositional activity following introduction into a new host population [55].

3.2 Transpositional regulation

For the linear model of transpositional regulation, we consider both positive and negative values of the fractional reduction in transposition rate with increasing copy number, b. While most studies show that replicative transposition rate is negatively correlated with copy number [56–58], studies of *copia* and *Doc* retrotransposons in *D. melanogaster* suggest that replicative transposition rate is positively associated with copy number [59,60].

For the threshold and continuous regulation models of transposition, we need an estimate of the repressed rate of transposition, a_r, as well as the copy number at which this repression occurs, denoted by i_t for the threshold model and i_c for the continuous regulation model. Since we are interested in the early stages of TE spread when copy numbers tend to be very small, we choose values of these parameters that are consistent with the value of parameter b.

3.3 Deletion

Element deletion events are so rare that they are often not observed in laboratory line experiments [61], and rates of element deletion, v, are thought to be at least two orders of magnitude less than the TE's baseline transposition rate [45,62]. Maside et al. [46] calculated a pooled excision rate of 3.95×10^{-6} per element per generation from a laboratory study of 11 families of TEs in *D. melanogaster*. This is in good agreement with the pooled estimate of 4.05×10^{-6} per element per generation calculated by Nuzhdin et al. [45] from other laboratory line experiments [63–65].

3.4 Host fitness

The parameter d represents the increase in host death rate with each additional TE copy in the genome and is approximately equal to the decrease in host fitness due to each additional TE copy. Following Mackay et al. [44], we estimated the fitness cost of a new genomic insertion by the average fitness cost of a spontaneous mutation. This has been estimated as $\sim$0.02 per element [66–68] 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 [69]. For the log-concave model of host fitness, we consider a power of $p = 1.5$ following Charlesworth and Charlesworth [9].

3.5 Population size changes

We are interested in both long-term population growth, and the seasonal population size changes that occur in *An. gambiae* in Banambani, Mali (Fig. 1). We have chosen to focus on the village of Banambani as a case study for the African continent since the ecology of *An. gambiae* has been intensively studied there [23,70–72].

Recent collections taken in Banambani suggest that peak population densities during the wet season are around 100 times those during the dry season [24], while conservative estimates are that seasonal population densities differ by a factor of ten [23]. The generation time of mosquitoes is around 16 days, following observations that mosquito maturation time is 12 days and the first eggs are laid four days following maturation (Mahamoudou Touré, personal communication). Assuming a ten-fold population size increase over six months (11.4 mosquito generations), this yields a population growth rate of $r = 0.2$ per generation, and a growth rate of $r = -0.2$ per generation for the corresponding population decline (Fig. 1).

These seasonal population size changes are superimposed over historical population expansions. Evidence from microsatellite allele size data suggests a recent population expansion in the malaria vectors *An. gambiae* and *An. arabiensis*[73]. Statistical tests

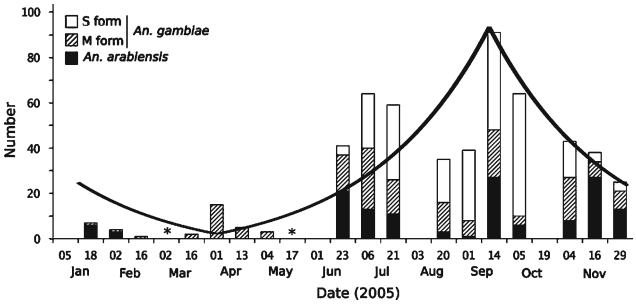

Fig. 1 Graph showing the number of *An. arabiensis* and *An. gambiae* M and S molecular forms typed from collections taken in Banambani, Mali during 2005 (from [24]). The difference in population sizes between wet and dry seasons is clearly evident in both M and S molecular forms and can be modeled approximately by alternating exponential growth and decline (*overlaid curve*)

on this data suggest an expansion that occurred on the order of $10N_e$ generations ago (where N_e is the effective population size prior to the population expansion) and may be contemporaneous with excessive penetration of agriculture into the African forest 4000 years ago [74,75]. Similar population expansions have been inferred for *D. melanogaster* populations. For example, the expansion of *D. melanogaster* into northern Africa and Eurasia $\sim 10^4$ years ago [76,77] is consistent with nucleotide data and coalescent simulations suggesting a 10^2–10^6-fold population size increase occurring around the same time [78].

3.6 Population size

Mark-release-recapture experiments on *An. gambiae* conducted in Banambani over five years found an average peak abundance of $\sim 6 \times 10^4$ [23,43]. Dividing this by three to account for the three chromosomal forms in Banambani yields a peak abundance of $\sim 2 \times 10^4$ per chromosomal form, or a population size prior to growth of $N_0 = 2 \times 10^3$. These population sizes are relatively small compared to the effective ancestral population size of *D. melanogaster*, which has been shown to be $\sim 2.0 \times 10^6$ [79].

3.7 Release strategy

In analyses where we considered the natural invasion of a TE following a horizontal transfer event, we initialized the branching process with a single host having a single copy of the TE. In our analyses of the genetic control strategy, we initialized the branching process with n_r transgenic hosts each having i_r TE copies released t_r generations following the beginning of the population growth phase. The dynamics of TE spread were then monitored in continuous time up to a prevalence of 10%.

4 Analysis

4.1 Critical parameter values for TE spread

One of the features of an ideal drive system listed by Braig and Yan [29] is the requirement that the drive must be strong enough to compensate for its inherent fitness cost. If the drive of a TE outweighs its fitness cost, then the TE has a chance to spread through the population and possibly reach fixation; however if the fitness cost outweighs the drive, then the TE will eventually be lost from the population. Within the framework of a branching process model, the possibility of TE spread corresponds to the case where the branching process is supercritical. A supercritical process has a basic reproductive number greater than one [30].

The basic reproductive number is equal to the average number of offspring of the same type that a particle eventually generates [80]. If we restrict ourselves to the very early stages of TE spread when the vast majority of infected hosts have only one or two copies of the TE, then the reproductive number of a type-1 host, $R_{1,1}$, can be

defined recursively as,

$$R_{1,1} = f_{1,1} + f_{1,2}R_{2,1}. \tag{10}$$

Here, $f_{1,1}$ is the average number of offspring of a type-1 host that are of type-1, $f_{1,2}$ is the average number of offspring of a type-1 host that are of type-2, and $R_{2,1}$ is the expected number of type-1 hosts that a type-2 host eventually generates. Following similar reasoning, $R_{2,1}$ can be defined recursively as,

$$R_{2,1} = f_{2,1} + f_{2,2}R_{2,1}, \tag{11}$$

where $f_{2,1}$ and $f_{2,2}$ are similarly defined. Equation 11 can then be rearranged and substituted into Eq. 10 to obtain the basic reproductive number for a type-1 host,

$$R_{1,1} = f_{1,1} + \frac{f_{1,2}f_{2,1}}{1 - f_{2,2}}. \tag{12}$$

Assuming linear models of transpositional regulation and host fitness, we can calculate the basic reproductive number by substituting Eqs. 1, 4, 7 and 8 into Eq. 12. Straightforward algebra then shows that for a type-1 host we have the basic reproductive number,

$$R_{1,1} = \frac{(r+1)\begin{pmatrix} a^2(b-1)(r+1) - 2a(r+1)(b(v-3) - v + 2) \\ -2(v-3)(4d - r + 2v(r+1) + 1 \end{pmatrix}}{2(d + 2r + 3)(4d - r + a(b-1)(r+1) + 2(r+1) + 1)}. \tag{13}$$

The drive of the TE outweighs its fitness cost when $R_{1,1} > 1$.

Before analyzing the full implications of this equation, first let us consider a simplified scenario in which transpositional regulation and element deletions are negligible ($b = v = 0$) and the population size is constant ($r = 0$). This leads to the basic reproductive number,

$$R_{1,1} = \frac{4a + a^2 - 24d - 6}{2(d+3)(a - 4d - 1)}, \tag{14}$$

which is greater than one when,

$$a > d + 1 - \sqrt{1 - 7d^2}. \tag{15}$$

For small fitness costs ($d < 0.01$ per element), Eq. 15 leads to the well known result that TE drive outweighs fitness cost when,

$$a > d, \tag{16}$$

and hence selection and transposition have an equal and opposite influence on TE copy number at equilibrium [45]. For larger fitness costs ($d > 0.01$ per element), the

magnitude of the transposition rate must be slightly greater than the magnitude of the fitness cost in order for TE spread to be critical.

When transpositional regulation and element deletions are not negligible and the population size is not constant, we must use Eq. 13 to determine the conditions for TE spread to be critical (Fig. 2). The critical balance between transposition rate and fitness cost is highly sensitive to changes in the population growth rate. This sensitivity is most visible when the transposition rate and fitness cost are small. For moderate population growth ($r = 0.005$ per generation), TE spread is always supercritical for fitness costs less than 0.005 per element. For moderate population decline ($r = -0.005$ per generation), TE spread is always subcritical for transposition rates less than 0.005 per element per generation. The criticality of TE spread is relatively insensitive to the degree of transpositional regulation, the model of transpositional regulation, and the model of host fitness with increasing copy number (supplemental Appendices at http://johnmm.bol.ucla.edu/te/).

The case where $r = 0.2$ per generation corresponds to the seasonal population growth phase of *An. gambiae*. During this phase, TE spread is supercritical for all transposition rates when the fitness cost is less than 0.2 per element. The implication of this result is that the conditions for supercritical TE spread are less restrictive during periods of population growth. This may enable the TE to become established in its host population prior to a subsequent population decline.

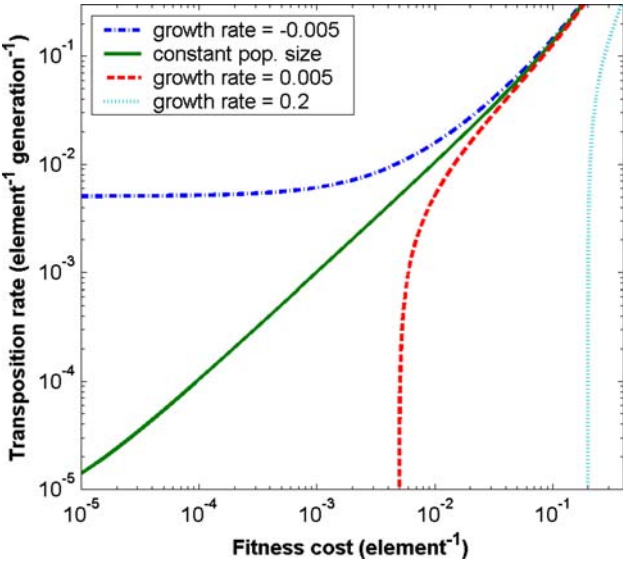

Fig. 2 Critical parameter values for TE spread. The default parameters from Table 1 are used. *Curves* correspond to values of transposition rate and fitness cost for which TE spread is critical. For the region of parameter space to the *left of each curve* TE spread is supercritical; while to the *right of each curve* TE spread is subcritical. The case of a constant population size confirms the well-known result that TE drive outweighs fitness cost when $a > d$. The critical balance between a and d is highly sensitive to changes in population growth rate, especially when the fitness cost and transposition rate are small. The case where $r = 0.2$ per generation corresponds to the seasonal population growth phase of *An. gambiae*

It should be noted that, in the case of *An. gambiae*, population growth is seasonal; while the above calculation applies to a period of continuous population growth. The case of seasonal population growth will be treated in Sects. 4.3 through 4.5. Additionally, knowledge of the conditions necessary for TE spread does not, on its own, confer knowledge of the rate of TE spread and the probability of TE loss. These quantities will be calculated in Sects. 4.2 through 4.4.

4.2 TE loss in a natural population

When TE spread is supercritical the eventual extinction of the TE is uncertain, however the probability that the TE is lost from the population may still be very high. Following a horizontal transfer a TE will be present in its new host population as a single copy in only a single host. At such a low prevalence, the probability of a recently transferred TE being lost through genetic drift is considerable, even if it has a very high transposition rate.

Considering the simplified scenario in which transpositional regulation and element deletion are negligible and population size is constant, Kaplan et al. [81] calculated the probability that a single TE is lost from its host population, e_1, to be the smallest solution of the equation,

$$e_1 = \exp(-(1 - e_1 \exp(-(1 - e_1)u_1))). \tag{17}$$

Here, u_1 is the transposition rate in a host carrying only a single element. This is in good agreement with simulation results of Le Rouzic and Capy [17] for transposition rates of $u_1 < 1$ per element per generation.

To determine the probability that the TE is lost from the host population when the population size is not constant, we use the branching process model described by Eqs. 1–9. We begin by defining the probability generating function for the process [41]. This is a function of the vector $\mathbf{s} = (s_1, s_2, \ldots, s_T)$ and is defined as,

$$P_i(\mathbf{s}) = \sum_j p_{ij} s^j = \sum_j p_{ij} s_1^{j_1} s_2^{j_2}, \ldots, s_T^{j_T}, \tag{18}$$

where p_{ij} is the probability that a type-i particle gives rise to j_1 type-1 particles, j_2 type-2 particles, and so on, and j is defined as the vector $j = (j_1, j_2, \ldots, j_T)$. Substituting the p_{ij} terms from the branching process into Eq. 18, we have the probability generating function,

$$P_i(\mathbf{s}) = \frac{\mu_i}{\lambda_i} + (f_{i,0} + f_{i,T+1})s_i + \sum_{j=1}^{T} f_{ij} s_i s_j - \frac{2(1+r)}{\lambda_i} s_i^2, \tag{19}$$

where $i \in \{1, 2, \ldots, T\}$. The probability that the TE is eventually lost from the population beginning with a single type-1 host, e_1, is then given by the smallest solution of the system of T simultaneous equations [64],

$$P_i(\mathbf{e}) = e_i \forall i, \tag{20}$$

where $\mathbf{e}$ is defined as the vector $\mathbf{e} = (e_1, e_2, \ldots, e_T)$, and e_i is the probability that a type-i host is eventually lost from the population. Note that we have calculated the asymptotic loss probabilities here, although the model is only valid up to a prevalence of 10%. This paradox is averted by considering that beyond a prevalence of 10% the TE has a high enough presence in the population that it is very unlikely to go extinct.

For a population of constant size, TE loss is uncertain when the transposition rate for a single TE copy is greater than the fitness cost of a single genomic insertion (Fig. 3a). This is in agreement with the prediction of Eq. 16. Loss probability falls as transposition rate increases and the fitness cost decreases, but even in the generous

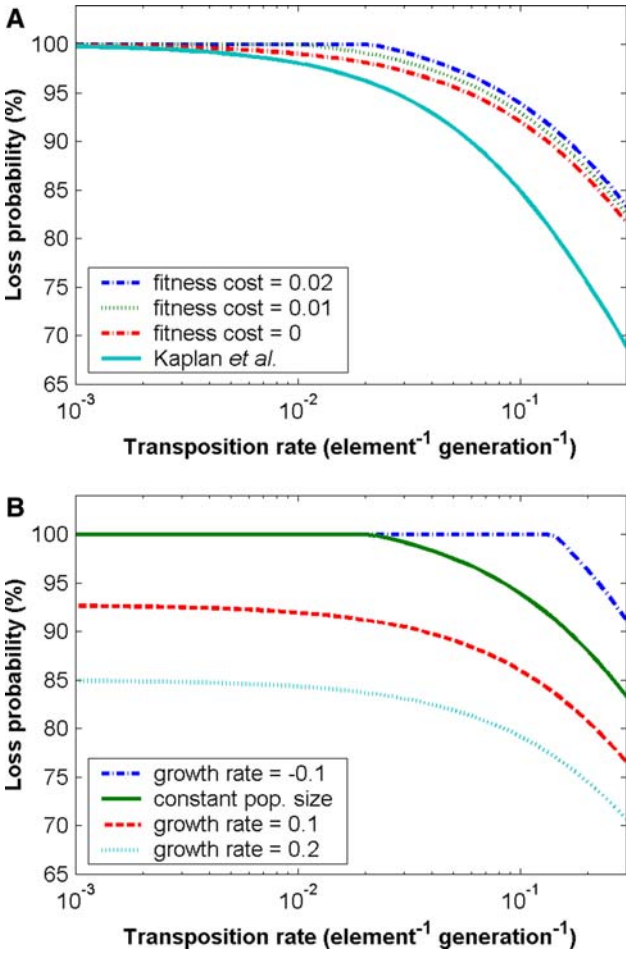

Fig. 3 Asymptotic probabilities of TE loss as a function of replicative transposition rate. The default parameters from Table 1 are used. **a** The fitness cost of a genomic insertion is varied. The case where $d = 0$ per element is comparable to the solution of Kaplan et al. [81]. **b** Population growth rate is varied. The case where $r = 0.2$ per generation corresponds to the seasonal population growth phase of *An. gambiae*

case of a neutral insertion with a transposition rate of 0.1 per element per generation, the loss probability is still as high as 92% (Fig. 3a). This suggests that the majority of TEs will become extinct following their introduction and many introductions are likely to occur before one of these TEs colonizes a new species.

To test for consistency with the framework of Kaplan et al. [81], the smallest solution to Eq. 17 is shown (Fig. 3a). Since Kaplan et al. [81] considered a neutral insertion, the case where $d = 0$ per element provides the best comparison. These curves are in good qualitative agreement with each other, although the solution of Kaplan et al. [81] produces slightly smaller loss probabilities. Incidentally, the loss probabilities of Kaplan et al. [81] are also slightly smaller than the simulated results of Le Rouzic and Capy [17], suggesting that the solution of Kaplan et al. [81] may slightly underestimate the probability of TE loss.

The prospects for a TE to colonize a new host species are dramatically improved during conditions of population growth (Fig. 3b). The impact of population growth on reducing TE loss probability is most visible for population growth on the order of $r = 0.1$ per generation. At this growth rate, TE loss probability is $\sim$7.5% lower than for the case of a constant population size over the entire range of transposition rates inferred from nature. When the rate of population growth is $r = 0.2$ per generation, as is the case during the seasonal population growth of *An. gambiae*, TE loss probability is lowered by an additional 7.5%. Combining a high transposition rate with conditions of population growth will further improve the odds of the TE colonizing a new host species. For a comparable rate of population decline ($r = -0.1$ per generation) TE loss probability is $\sim$7.5% higher than for the case of a constant population size, and TE loss is certain for transposition rates less than 0.11 per element per generation.

The sensitivity of TE loss probability to the model of transpositional regulation and host fitness (Eqs. 1–6) was also tested under a variety of parameterizations (supplemental Appendices at http://johnmm.bol.ucla.edu/te/). The implication of these results is that whether the TE is lost from the population or increases in prevalence exponentially is relatively independent of the way in which transposition rate and fitness cost change with increasing copy number. What happens at higher copy numbers will no doubt affect the distribution of element copy number in the later stages of spread, but whether the TE spreads or not seems to be primarily determined by what happens when the genomic copy number is one.

4.3 TE loss following a transgenic release

Given the impact of population size changes on the eventual loss of a newly transferred TE, the implications this has for genetic control strategies in *An. gambiae* populations is of great interest. Indeed, one of the features of an ideal drive system listed by James [28] is the requirement that the drive system be efficient in establishing an effector gene in the population. Therefore the TE loss probability following a transgenic release should be exceptionally small. To account for the seasonal population size changes of *An. gambiae*, we calculate the probability that the TE is lost from the *An. gambiae* population one year following its release. This allows us to account for a full cycle of population growth and decline.

We consider the release of a single transgenic mosquito t_r generations following the beginning of the population growth phase having i copies of the TE. To calculate the probability that this TE is lost from the population one year following release, we derive a differential equation for the loss probability as a function of time, $e_i(t)$, and solve this at a time one year following release subject to the initial condition $e_i(0) = 0 \forall i$. During the first season, the loss probabilities are characterized by the system of nonlinear ordinary differential equations [30],

$$\frac{de_i(t)}{dt} = -\lambda_i e_i(t) + \lambda_i P_i(\mathbf{e}(t)), \tag{21}$$

where $\mathbf{e}(t)$ is defined as the vector $(e_1(t), e_2(t), \ldots, e_T(t))$.

During the second season, the loss probabilities can be calculated by characterizing the distribution of copy numbers at the end of the first season at time t_1, and calculating the probability that each of these lineages of infected individuals becomes extinct during a time $t_2 = t - t_1$ into the second season. This is encapsulated by the multivariate generating function,

$$Q_i(t_1, \mathbf{e}(t_2)) = \sum_{\mathbf{k}} \Pr(Z_{t_1} = \mathbf{k} | Z_0 = u_i) \mathbf{e}(t_2)^{\mathbf{k}}, \tag{22}$$

where $Z_t = (Z_{t,1}, Z_{t,2}, \ldots, Z_{t,T})$ is the vector of particle counts at time $t \geq 0$, u_i is the standard unit vector whose ith entry is equal to one, and $\mathbf{e}(t_2)$ is the vector of loss probabilities whose entries are calculated using the population growth rate parameter for the second season. This generating function is characterized by the system of nonlinear ordinary differential equations [30],

$$\frac{dQ_i(t, \mathbf{e}(t_2))}{dt} = -\lambda_i Q_i(t, \mathbf{e}(t_2)) + \lambda_i P_i(\mathbf{Q}(t, \mathbf{e}(t_2))), \tag{23}$$

where $\mathbf{Q}(t, \mathbf{e}(t_2))$ is defined as the vector $(Q_1(t,\mathbf{e}(t_2)), Q_2(t,\mathbf{e}(t_2)), \ldots, Q_T(t,\mathbf{e}(t_2)))$. Solving this at a time t_2 into the second season subject to the initial condition $Q_i(0, \mathbf{e}(t_2)) = e_i(t_2) \forall i$ gives the probability of TE loss at time t, where $t = t_1 + t_2$. This procedure can be iterated to calculate the TE loss probabilities during subsequent seasons.

Next, having calculated the TE loss probability for a single transgenic mosquito one year following release, this result can be extended to a transgenic release consisting of $\mathbf{n}$ mosquitoes, where $\mathbf{n}$ is defined as the vector $(n_1, n_2, \ldots, n_T)$ and n_i represents the number of mosquitoes having i copies of the TE at the time of release. This follows simply from the independence of particles so that the collective extinction probability is $\mathbf{e}(t)^{\mathbf{n}}$. For a release consisting of n_r transgenic hosts each having i_r copies of the TE, the extinction probability is $e_{i_r}(t)^{n_r}$.

One of the main results of this calculation concerning the release strategy is that the model favors a transgenic release immediately following the dry season when the *An. gambiae* population begins to grow (Fig. 4a). A release at this time reduces the probability that the TE is lost from the population in the first year following its

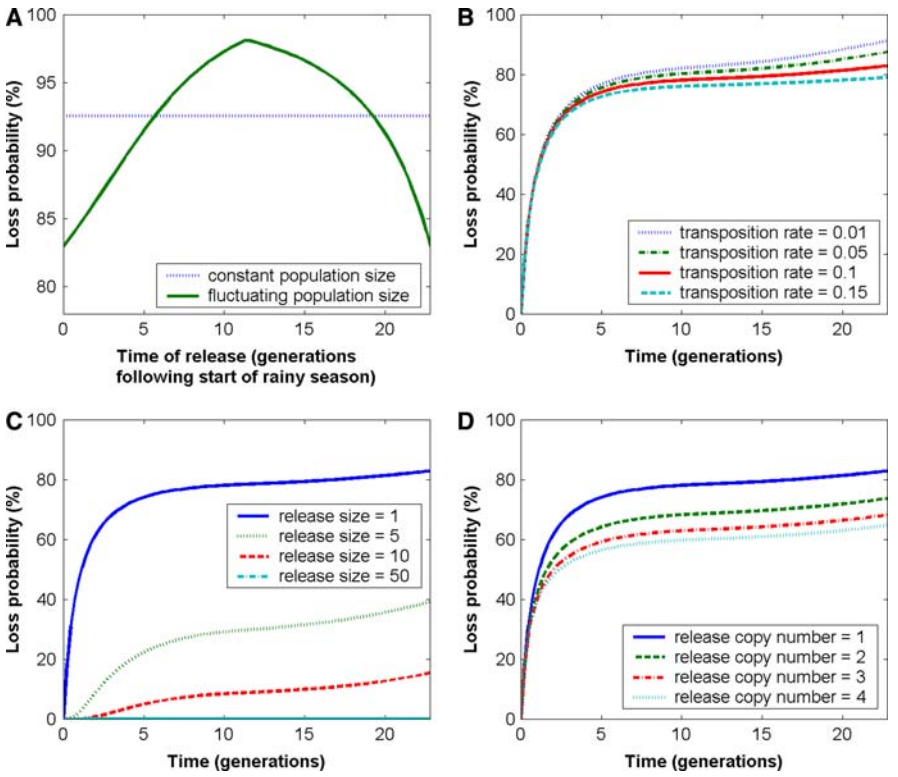

Fig. 4 Probabilities of TE loss 1 year following release. The default parameters from Table 1 are used, with the exception that $n_r = 1$. **a** The *An. gambiae* host population is assumed to grow for 6 months, and then decline for 6 months. Release time is measured in generations following the beginning of the population growth phase. **b** The replicative transposition rate is varied. **c** Release size is increased to 50. **d** Copy number per individual at release is increased to four

release. The model also suggests that, in the event that knowledge of the population growth rate is imprecise, the control strategy should bias towards a release after the beginning of the population growth phase. This follows from observations that TE loss probabilities are smaller for a transgenic release at the beginning of the rainy season than at the end of the dry season (Fig. 4a). A release under these conditions in a population of fluctuating size has a better chance of persisting than an identical release in a population of constant size.

Assuming a release at the beginning of the rainy season, Fig. 4b then depicts the TE loss probability as a function of time for a variety of transposition rates. As expected, a TE with a high transposition rate is less likely to be lost from the host population. At the end of the year, a TE with a transposition rate of 0.01 per element per generation has a loss probability of 91% while a TE with a transposition rate of 0.1 per element per generation has a loss probability of 83%. TE loss probabilities 1 year following release are relatively insensitive to the model of transpositional regulation and host fitness (supplemental Appendices at http://johnmm.bol.ucla.edu/te/).

The simplest and most effective way to improve the efficiency of a TE drive strategy is to release more transgenic mosquitoes (Fig. 4c). According to model predictions, increasing the release size to ten reduces the loss probability to 15% one year following release. For a release size of 25 the loss probability falls below 1%. Another way to increase the number of TEs released into the population is to increase the copy number of the introduced mosquitoes (Fig. 4d). This reduces the loss probability of the TE, but to a lesser extent than increasing the release size by the same amount. Therefore we favor the release of a greater number of transgenic mosquitoes each having a single copy of the TE.

4.4 Rate of TE spread following a transgenic release

To satisfy the second requirement of Braig and Yan [29] that the drive system must work within a public health time frame, we consider the influence of population growth and various model parameters on reducing the time taken for the TE to reach a proportion of 10% in the *An. gambiae* population.

The most intuitive and complete information derivable from the branching process is the mean and variance in the number of disease vectors having different TE copy numbers over time. For an initial release consisting of $\mathbf{n}$ mosquitoes, this is given by the matrix exponential [41],

$$E(Z_t) = \mathbf{n}e^{t\Omega}, \tag{24}$$

where Z_t is the random vector of particle counts whose entries $Z_{t,i}$ represent the number of disease vectors having i copies of the TE at time t, and Ω is the branching process matrix whose entries are given by $\lambda_i(f_{ij} - 1_{\{i=j\}})$. The above equation is valid for the first season, however during the second season the particle counts are given by,

$$E(Z_t) = \mathbf{n}e^{t_1\Omega_1}e^{t_2\Omega_2}, \tag{25}$$

where Ω_1 and Ω_2 are the branching process matrices calculated using the first and second season population growth rates respectively. This procedure can be iterated to obtain the vector of particle counts during subsequent seasons.

The variance of the particle count vector during the first season can be calculated following the methodology in the Appendix of Dorman et al. [41]. This applies to the variance matrix $V_i(Z_t)$ for a transgenic release involving a single transgenic mosquito having i copies of the TE. For a release consisting of $\mathbf{n}$ transgenic mosquitoes, the variance matrix follows from the independence of clans [30], and is given by,

$$\mathrm{Var}(Z_t) = \sum_{i=1}^{T} n_i V_i(Z_t). \tag{26}$$

During the second season, the variance matrix can be calculated by conditioning on the state of the particle count vector at the end of the first season. Simple algebraic

manipulation then gives,

$$\mathrm{Var}(Z_t) = \sum_{i=1}^{T} E(Z_{t_1,i}) V_i(Z_{t_2}) + \sum_{i=1}^{T} \mathrm{Var}(Z_{t_1,i}) E_i(Z_{t_2}) E_i(Z_{t_2})^*, \qquad (27)$$

where $E_i(Z_t)$ represents the mean particle count vector at time t given a single transgenic mosquito having i copies of the TE at time 0. This approach can be iterated to calculate the variance in particle counts over subsequent seasons.

Applying this approach to a release of 50 transgenic mosquitoes, we calculated the mean and variance in the number of each host type over time. The mean number of transgenic hosts increases exponentially during the rainy season and then decreases exponentially during the subsequent dry season, reflecting the dynamics of the uninfected *An. gambiae* population (Fig. 5a). The replicative behavior of the TE also leads to transgenic *An. gambiae* being over five times more prevalent 1 year following release. Of the 260 expected transgenic *An. gambiae* at the end of the dry season, 236 have a single copy of the TE, 23 have two TE copies, and 1 has three TE copies (Fig. 5a). This Poisson-like distribution of copy number is largely a consequence of the increase in copy number due to replicative transposition being counteracted by the reduction of copy number due to gamete formation and mating with uninfected hosts.

In contrast to the results for TE loss probabilities, the rate of population growth has no influence on the proportion of hosts having the TE 1 year following its release (Fig. 5b). For higher population growth rates, the proportion of individuals having the TE increases more quickly during the rainy season and more slowly during the dry season, eventually ending at the same proportion regardless of population growth rate over the period of a year. Despite this independence, a release at the beginning of the population growth phase is recommended since it results in the TE having the highest initial proportion in the population.

Whether the TE increases in proportion in the first year following its release is largely determined by the difference between its replicative transposition rate and fitness cost (Fig. 5c). The TE increases its proportion in the host population when the magnitude of its transposition rate exceeds its fitness cost, which is an interesting parallel to Eq. 16. The requirement of a rapid spread [30] is most effectively achieved by ensuring that the transposition rate is sufficiently high. For example, the TE reaches a proportion of 10% within a year of its release when its transposition rate is greater than 0.087 per element per generation (Fig. 5c). The rate of TE spread 1 year following release is relatively independent of the model of transpositional regulation and host fitness (supplemental Appendices at http://johnmm.bol.ucla.edu/te/).

Finally, the waiting time for the mean number of transgenic *An. gambiae*to reach a proportion of 10% in the population is predominantly a function of replicative transposition rate and release size (Fig. 5d). While the waiting times starting with release sizes of 50 and 100 are relatively similar, the waiting times beginning with a single transgenic host are significantly longer. Similar comparisons over a smaller range of release sizes are shown in the supplemental Appendices (http://johnmm.bol.ucla.edu/te/). The relevance of these comparisons is that a more effective strategy than releasing hundreds of mosquitoes into one population might be to release fewer mosquitoes into multiple villages or into multiple chromosomal forms

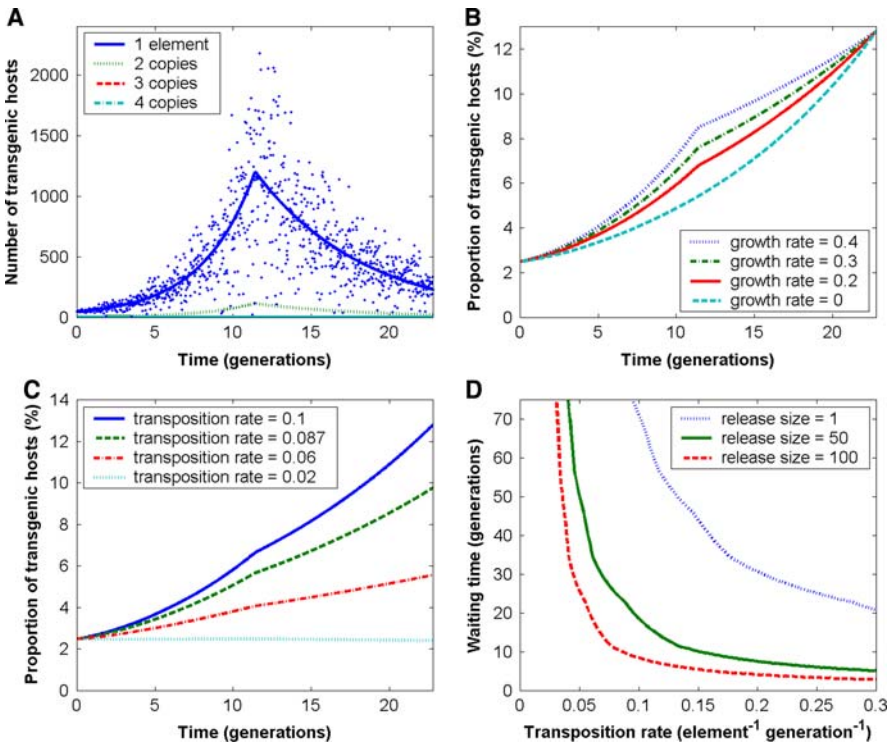

Fig. 5 Rate of TE spread for the first year following a transgenic release. The default parameters from Table 1 are used. **a** The number of transgenic *An. gambiae* with different TE copy numbers are tracked over time. Mean numbers of hosts are shown as *bold lines*. The variance of host numbers having a single TE is illustrated by distributing points about the mean according to a normal distribution with variance equal to the variance in host numbers (note that the distribution of host numbers does not necessarily obey a normal distribution). **b** The proportion of transgenic *An. gambiae* is tracked over time. Population growth rate is varied. **c** Replicative transposition rate is varied. **d** The waiting time for the TE to reach a proportion of 10% in the host population is calculated as a function of replicative transposition rate. Release size is varied

in the same village. This should reduce the fixation time across a larger geographical area.

4.5 Rate of TE spread beyond a prevalence of 10%

One major restriction of the branching process model is the assumption that host organisms do not interact during mating. This is not a good approximation for prevalences of transgenic hosts above 10%; hence another modeling framework is required to make inferences regarding the ultimate fate of the TE in the host population.

We consider a system of ordinary differential equations in which the proportion of hosts having k copies of the TE is kept track of by its own differential equation,

$$\frac{dx_k(t)}{dt} = \theta(t) \sum_{i=0}^{T} \sum_{j=0}^{T} p_{ijk} x_i(t) x_j(t) - \mu_k x_k(t). \tag{28}$$

Here, $x_k(t)$ represents the proportion of hosts having k TE copies, where $k \in \{0, 1, \ldots, T\}$ and T can be approximated as the number of genomic sites that will be occupied when the TE reaches equilibrium in the host population. This equation simply describes the difference in birth and death rates for hosts having k TE copies—$\theta(t)$ is the overall birth rate parameter; p_{ijk} is the probability that, when a host having i TE copies mates with a host having j TE copies, the offspring has k TE copies; and μ_k is the death rate of a host having k TE copies, as previously described.

The form of the birth rate term $\theta(t)$ is given by setting the total birth rate equal to the total death rate. Then, to calculate the probability that an offspring has k TE copies, we first consider the probability, p_{im}, that a diploid host having i TE copies produces a haploid gamete having m TE copies. This is almost identical to the quantity f_{im} for the branching process and is given by,

$$p_{im} = \frac{1}{2^{i-1}} \binom{i-1}{m} \beta_i + \frac{1}{2^i} \binom{i}{m} (1 - \alpha_i - \beta_i) + \frac{1}{2^{i+1}} \binom{i+1}{m} \alpha_i. \quad (29)$$

The probability that, when a type-i host mates with a type-j host, the offspring is of type-k is then given by,

$$p_{ijk} = \sum_{m=0}^{k} p_{im} p_{j,(k-m)}, \quad (30)$$

where $i, j, k \in \{0, 1, \ldots, T\}$, and both p_{im} and $p_{j,(k-m)}$ are given by Eq. 29.

In order to test the consistency of the differential equation model with the branching process model of TE spread, we calculated the proportion of transgenic hosts over time for a variety of transposition rates (Fig. 6). For both modeling frameworks, we used the continuous model of transpositional regulation, the linear model of host fitness, and the default parameters from Table 1 with the exception that $r = 0$. The results suggest that the two modeling frameworks are consistent up to a prevalence of 10%, and that the major assumption of the branching process – that each mating only involves a single transgenic host – is indeed satisfied when the prevalence of transgenic hosts is low.

Figure 6 also implies that the early stages of TE spread provide a good indication of the ultimate fate of the TE in the population. When the transposition rate is 0.02 per element per generation, the TE declines in prevalence in the early stages and then continues to decline. When the transposition rate is 0.1 per element per generation, the TE quickly reaches a prevalence of 10% and then continues quickly to near-fixation. When the transposition rate is 0.06 per element per generation, the TE reaches 10% prevalence slowly and approaches near-fixation slowly.

The TE reaches a high maximum prevalence in the host population (Fig. 6), but does not reach complete fixation because there will always be a proportion of offspring that do not inherit the TE. Fixation is preferable to near-fixation; however a high prevalence of the effector gene may be sufficient to decrease the reproductive number of the pathogen to below one. A parallel epidemiological model is required for an accurate assessment of disease control.

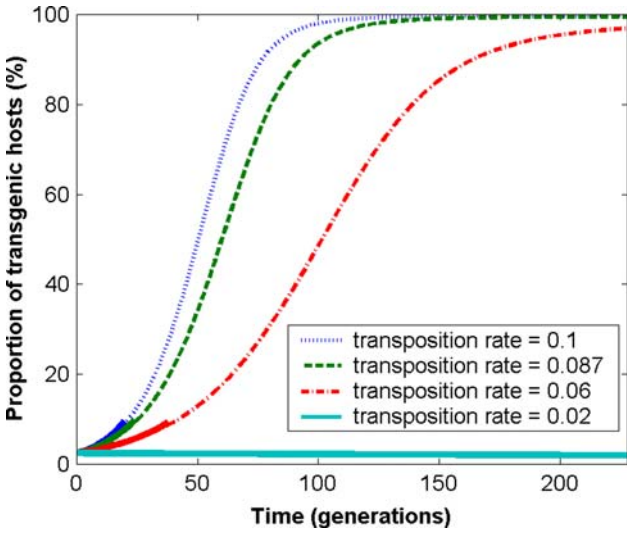

Fig. 6 Rate of TE spread for the first 10 years following a transgenic release. Predictions of the branching process model are shown as *thick*, *solid curves* for prevalences less than 10%. Predictions of the ODE model are shown as *narrow*, *patterned curves* for prevalences up to 100%. The default parameters from Table 1 are used, with the exception that $r = 0$. Transposition rate is varied

4.6 Sensitivity of model parameters

The sensitivity of the TE loss probability to any continuous model parameter γ can be calculated by taking its partial derivative, $\partial \mathbf{e}(t)/\partial\gamma$, and scaling this by the range of values observed in nature, $\Delta\gamma$. During the first season, this can be calculated by taking the partial derivative of Eq. 21 with respect to γ and integrating up to time t_1 subject to the initial condition that $\partial e_i(0)/\partial\gamma = 0 \forall i$. During the second season, the sensitivity of the TE loss probability can be calculated by taking the partial derivative of Eq. 22 with respect to γ and integrating up to time t_1 subject to the initial condition that $\partial Q_i(0, \mathbf{e}(t_2))/\partial\gamma = \partial e_i(t_2)/\partial\gamma \forall i$. Here, the second season population growth rate is used in the calculation of the initial condition, and the first season population growth rate is used in the integration.

The sensitivity of the rate of TE spread can be inferred from the sensitivity of the dominant eigenvalue, $(\partial\rho/\partial\gamma)\Delta\gamma$, when the population growth rate is set to 0. This follows from the observation that, on the scale of years, the rate at which the proportion of TEs in the population increases is independent of r (Fig. 5c). The sensitivity of the dominant eigenvalue can be calculated following the methodology in Sect. 11 of Dorman et al. [41].

Table 2 shows the sensitivities of the TE loss probability and rate of TE spread to the continuous parameters of our model. Sensitivities for the alternative models of transpositional regulation and host fitness are shown in the supplemental Appendices (http://johnmm.bol.ucla.edu/te/). These sensitivity analyses do not account for discrete parameters such as the release size and copy number at release; however a number of important implications can be deduced for prioritizing parameters that should be accurately known for a transgenic release. Since we have more control over and better

Table 2 Sensitivity of loss probability and rate of TE spread to model parameters

Model parameter γ	Parameter range $\Delta\gamma$	Sensitivity of asymptotic loss probability $(\partial e_1/\partial\gamma)\Delta\gamma$	Sensitivity of one year loss probability $(\partial e_1(t)/\partial\gamma)\Delta\gamma$	Sensitivity of rate of element spread $(\partial\rho/\partial\gamma)\Delta\gamma$
a	0.3 TE^{-1} gen^{-1}	−0.20	−0.19	0.28
b	0.2	0.0016	0.00094	−0.0035
d	0.4 TE^{-1}	0.37	0.37	−0.47
r	0.8 gen^{-1}	−0.68	−0.56	–

estimates of transposition rate and release size, then we should focus on these parameters in engineering efforts. However, it is also important that we have good estimates of less easily manipulated parameters such as the population growth rate and fitness cost of a genomic insertion. These are some of the most influential factors in the spread of the TE and are essential to ensure that our model predictions are accurate.

Since the sensitivities have been calculated using partial derivatives, their signs tell us whether the TE loss probability and rate of spread are increasing or decreasing functions of each model parameter. According to Table 2, loss probability is a decreasing function of transposition rate and population growth rate, and an increasing function of fitness cost and transpositional regulation parameter b. The rate of TE spread has the opposite dependencies.

5 Discussion

The proposed model provides us with a succinct framework to explore the early stages of TE spread through a diploid population. We can also make some crude recommendations regarding the implications this has for the use of TEs as drive mechanisms in the control of vector-borne disease. We chose a branching process to model this phenomenon since it has the benefit of being analytically tractable while still having the stochasticity of a Markov chain [30]. This has enabled us to investigate the probability that a TE is lost from its host population as a result of genetic drift, and how this probability is affected during periods of population growth and decline.

Transposable elements face significant challenges upon being transferred into a new host population, thus making the consideration of genetic drift particularly relevant [17]. Whether a TE begins its invasion through horizontal transfer mediated by a vector [82,83], introgression [84] or the de novo appearance of a new TE within the host species, the process invariably begins with only a single element. The fate of the element in its new host population is then a product of its internal properties, the population dynamics of the host species, and chance.

5.1 Transpositional bursts

There is a large body of evidence supporting the existence of heightened transposition rates at low copy numbers, which may enable TEs to prosper in the early stages of

spread. Several mechanisms may contribute to this, including hybrid dysgenesis [85, 86], overproduction inhibition [36], and gene silencing [87]. Furthermore, evidence for a recent transpositional burst in the *DINE*-1 element of *Drosophila yakuba* has been presented by Yang et al. [55]. Predictions of this and other models [17,81] suggest that a heightened transposition rate following introduction into a novel species will reduce the loss probability and speed up the rate of TE spread in the new population. However, even under generous conditions (a neutral TE and heightened transposition rate of 0.1 per element per generation), the loss probability can still be as high as 92% (Fig. 3a).

5.2 Population growth

The prospects for a TE to colonize a new host species are dramatically improved during conditions of population growth. According to model predictions, when the rate of population growth is $r = 0.2$ per generation, as is the case during the seasonal population growth phase of *An. gambiae*, the probability of TE loss is lowered by ~15% compared to the case of a constant population size (Fig. 3b). Although seasonal population growth occurs on a time scale of six months, growth can also occur over a much longer time period when a species colonizes a new geographical location. The recent population expansion of the malaria vectors *An. gambiae* and *An. arabiensis* [73] and the expansion of *D. melanogaster* into northern Africa and Eurasia [76,77] may be examples of this. The prediction of this model is that such a period of growth is a ripe time for a TE to colonize a new species.

5.3 Implications for a transgenic release

Any plan to release a TE into a novel host population for the purpose of disease control should aim to exploit the existence of both heightened transposition rates and population size changes to improve its likelihood of success. In the case of *An. gambiae*, dramatic population size changes between the dry season and the peak of the rainy season (Fig. 1) should be used to the advantage of the release strategy. To this end, a transgenic release is recommended immediately following the dry season when the *An. gambiae* population begins to grow (Fig. 4a). This allows the TE the best chance to establish itself in the population prior to the subsequent population decline. A less intuitive prediction is that a release slightly after the beginning of the growth phase is more efficient than a release prior to the growth phase. This suggests that, in the event that knowledge of population size changes is imprecise, the release strategy should bias towards a late release.

In addition to manipulating the release time, the model makes a number of other suggestions to lower the loss probability and increase the rate of TE spread through the *An. gambiae* population. TE loss probability is most effectively reduced by increasing the release size of transgenic mosquitoes to more than 25, while increasing the TE copy number of the released mosquitoes is less effective since the TEs are more likely to be lost at once (Fig. 4c, d). This is a larger release size than that suggested by Struchiner et al. [18] who suggested a release of more than eight transgenic hosts;

however Struchiner et al. [18] also found that the TE copy number of the released hosts is relatively inconsequential.

The most effective way to ensure that the TE spreads at a rate acceptable to public health goals is by ensuring that the introduced TE has a high replicative transposition rate upon release. Here, the model recommends that a TE with a replicative transposition rate of 0.1 per element per generation is realistic [53,54] and will reach a proportion of 10% in a village population in an acceptable time frame (Fig. 4C). This is in agreement with Rasgon and Gould [19] for the case where transposition is biased towards unlinked sites, and Le Rouzic and Capy [88] who suggest that realistically useful TEs will have a transposition rate as high as 0.1 per element per generation.

A number of parameters that are less easily engineered also have important implications for the success of the disease control strategy. The fitness cost of a genomic insertion is the most influential parameter on the rate of TE spread, and the rate of population growth is one of the most influential parameters on TE loss probability (Table 2). It should be noted that most of the parameter values used to reach these conclusions were measured in *D. melanogaster*; hence it is important that we seek accurate estimates of these parameters in species, such as *An. gambiae*, being considered for transgenic disease control.

5.4 Model limitations

Symptomatic of any mathematical analysis, simplifications have been made that may compromise the model predictions. Firstly, in the model formulation all TEs 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 [89,90] and to home in on certain genomic regions [47,91]. Clustering of element copies can reduce the rate of TE spread since tightly linked element copies rarely segregate during meiosis and are effectively inherited as a single unit [19]. This also threatens the result that the majority of hosts only have a single element copy in the early stages of TE spread, since the replicative effect of transposition is counteracted less by the dilutive effect of sampling during gamete formation when elements are clustered. These are clearly important considerations and worthy of much further study; however we have neglected them in the present analysis in order to focus on the implications of population size changes.

Secondly, the branching process model is restrictive since it only applies to TE spread up to a prevalence of 10%. The assumption that all matings involve at least one uninfected host is required by the independence of particles in a branching process, but doesn't allow us to ask questions about the waiting time for a TE to reach fixation. Despite this, modeling TE spread up to a prevalence of 10% captures the impact of population size changes fairly well, since the main effects of population growth and decline are stochastic and most relevant when the TE is present in small numbers. Furthermore, the mean number of transgenic hosts for prevalences above 10% can be calculated using the differential equations model (Eqs. 28–30).

Finally, of relevance to a transgenic release, another molecular detail that has not been accounted for in this paper is the potential for internal deletion to lead to loss

of the effector gene from the TE construct [3,92,93]. A related concern is that TEs that have undergone internal deletion may spread through the population more quickly than intact introduced TEs that still carry their transgenic load. Future modeling should assess the severity of this concern and experimental studies should be conducted on candidate TEs in *An. gambiae*.

5.5 Future directions

To gain a deeper insight into the impact of population structure on TE spread, we should consider the spatial as well as temporal population structure of a species, and how these interact together. In the case of *An. gambiae*, this would involve accounting for the existence of multiple chromosomal forms [25,26,94] and how their relative abundances change throughout the year (Fig. 1). This should be superimposed over the geographical distribution of mosquito populations in discrete patches corresponding to villages [25] and the relative abundance of each chromosomal form from village to village [42]. Clearly there is a relationship between the spatial and temporal dimensions of population structure, and an understanding of this will better enable us to assess the potential benefits of TE-mediated strategies for the control of vector-borne disease.

Acknowledgments I am particularly grateful to Prof Ken Lange for guidance on the use of continuous time branching processes and to Prof Charles Taylor for guidance on ecological issues. I am also grateful to Dr Nicholas Manoukis and Dr Mahamoudou Touré for allowing the reproduction of Fig. 1 in this manuscript, and to two anonymous reviewers for constructive comments on an earlier draft.

References

1. Marracci, S., Batistoni, R., Pesole, G., Citti, L., Nardi, I.: *Gypsy*/*Ty3*-like elements in the genome of the terrestrial salamander, *Hydromantes* (*Amphibia Urodela*). J. Mol. Evol. **43**, 584–593 (1996)
2. Biémont, C., Vieira, C.: The influence of transposable elements on genome size. J. Soc. Biol. **198**, 413–417 (2004)
3. Engels, W.R.: *P* elements in *Drosophila melanogaster*. In: Berg, D.E., How, M.M. (eds.) Mobile DNA, pp. 439–484. ASM Press, Washington DC (1989)
4. Craig, G.B.: Prospects for vector control through manipulation of populations. Bull. WHO **29**, 89–97 (1963)
5. Curtis, C.F.: Possible use of translocations to fix desirable genes in insect pest populations. Nature **218**, 368–369 (1968)
6. Atkinson, P.W., James, A.A.: Germline transformants spreading out to many insect species. Adv. Genet. **47**, 49–86 (2002)
7. Marrelli, M.T., Li, C., Rasgon, J.L., Jacobs-Lorena, M.: Transgenic malaria-resistant mosquitoes have a fitness advantage when feeding on Plasmodium-infected blood. Proc. Natl. Acad. Sci. USA **104**, 5580–5583 (2007)
8. Scott, T.W., Takken, W., Knols, B.G., Boëte, C.: The ecology of genetically modified mosquitoes. Science **298**, 117–119 (2002)
9. Charlesworth, B., Charlesworth, D.: The population dynamics of transposable elements. Genet. Res. **42**, 1–27 (1983)
10. Charlesworth, B., Langley, C.H.: The evolution of self-regulated transposition of transposable elements. Genetics **112**, 359–383 (1986)
11. Sawyer, S., Hartl, D.L.: Distribution of transposable elements in prokaryotes. Theor. Popul. Biol. **30**, 1–16 (1986)
12. Moody, M.E.: A branching process model for the evolution of transposable elements. J. Math. Biol. **26**, 347–357 (1988)

13. Brookfield, J.F.Y., Badge, R.M.: Population genetics models of transposable elements. Genetica **100**, 281–294 (1997)
14. Ribeiro, J.M.C., Kidwell, M.G.: Transposable elements as population drive mechanisms: specification of critical parameter values. J. Med. Entomol. **31**, 10–16 (1994)
15. Kiszewski, A.E., Spielman, A.: Spatially explicit model of transposon-based genetic drive mechanism for displacing fluctuating populations of Anopheline vector mosquitoes. J. Med. Entomol. **35**, 584–590 (1998)
16. Boëte, C., Koella, J.C.: A theoretical approach to predicting the success of genetic manipulation of malaria mosquitoes in malaria control. Malaria J. **1**, 3 (2003)
17. Le Rouzic, A., Capy, P.: The first steps of a transposable element invasion: parasitic strategy vs. drift. Genetics **169**, 1033–1043 (2005)
18. Struchiner, C.J., Kidwell, M.G., Ribeiro, J.M.C.: Population dynamics of transposable elements: copy number regulation and species invasion requirements. J. Biol. Syst. **13**, 455–475 (2005)
19. Rasgon, J.L., Gould, F.: Transposable element insertion location bias and the dynamics of gene drive in mosquito populations. Insect Mol. Biol. **14**, 493–500 (2005)
20. Deceliere, G., Charles, S., Biémont, C.: The dynamics of transposable elements in structured populations. Genetics **169**, 467–474 (2005)
21. Deceliere, G., Letrillard, Y., Charles, S., Biémont, C.: TESD: a transposable element dynamics simulation environment. Bioinformatics **22**, 2702–2703 (2006)
22. Basten, C.J., Moody, M.E.: A branching process model for the evolution of transposable elements incorporating selection. J. Math. Biol. **29**, 743–761 (1991)
23. Taylor, C., Touré, Y.T., Carnahan, J., Norris, D.E., Dolo, G., Traore, S.F., Edillo, F.E., Lanzaro, G.C.: Gene flow among populations of the Malaria vector *Anopholes gambiae* in Mali, West Africa. Genetics **157**, 743–750 (2001)
24. Manoukis, N.C.: Studies on the ecology and adaptation of *Anopholes gambiae* in Mali and their impacts on malaria transmission and control. PhD Thesis, University of California, Los Angeles (2006)
25. Touré, Y.T., Petrarca, V., Coluzzi, M.: Nueva entita del compleso *Anopholes gambiae* in Mali. Parrasitologia **25**, 367–370 (1983)
26. Coluzzi, M., Petrarca, V., Di Deco, M.A.: Chromosomal inversion intergradation and incipient speciation in *Anopholes gambiae*. Boll. Zool. **52**, 45–63 (1985)
27. Touré, Y.T., Dolo, G., Petrarca, V., Traore, S.F., Bouare, M., Dao, A., Carnahan, J., Taylor, C.E.: Mark-release-recapture experiments with *Anopholes gambiae s.l.* in Banambani Village, Mali, to determine population size and structure. Med. Vet. Ent. **12**, 74–83 (1998)
28. James, A.A.: Gene drive systems in mosquitoes: rules of the road. Trends Parasitol. **21**, 64–67 (2005)
29. Braig, H.R., Yan, G.: The spread of genetic constructs in natural insect populations. In: Letourneau, D.K., Burrows, B.E. (eds.) Genetically engineered organisms: assessing environmental and human health effects, pp. 251–314. CRC Press, New York (2001)
30. Lange, K.: Applied Probability. Springer, New York (2003)
31. Weinreich, M.D., Gasch, A., Reznikoff, W.S.: Evidence that the *cis* preference of the *Tn*5 transposase is caused by nonproductive multimerization. Genet. Dev. **8**, 2363–2374 (1994)
32. Wu, C.T., Morris, J.R.: Transvection and other homology effects. Curr. Opin. Genet. Dev. **9**, 237–246 (1999)
33. Townsend, J.P., Hartl, D.L.: The kinetics of transposable element autoregulation. Genetica **108**, 229–237 (2000)
34. Labrador, M., Corces, V.G.: Transposable element-host interactions: regulation of insertion and excision.
Annu. Rev. Genet. **31**, 381–404 (1997)
35. Lemaitre, B., Ronsseray, S., Coen, D.: Maternal repression of the *P* element promoter in the germline of *Drosophila melanocaster*: a model for the *P* cytotype. Genetics **135**, 149–160 (1993)
36. Lohe, A.R., Hartl, D.L.: Autoregulation of mariner transposase activity by overproduction and dominant-negative complementation. Mol. Biol. Evol. **13**, 549–555 (1996)
37. Mackay, T.F.: Transposable elements and fitness in *Drosophila melanogaster*. Genome **31**, 284–295 (1989)
38. Montgomery, E., Charlesworth, B., Langley, C.H.: A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*. Genet. Res. **49**, 31–41 (1987)

39. Brookfield, J.F.Y.: Models of repression of transposition in $P - M$ hybrid dysgenesis by P cytotype and by zygotically encoded repressor proteins. Genetics **128**, 471–486 (1991)
40. Brookfield, J.F.Y.: Models of the spread of non-autonomous selfish transposable elements when transposition and fitness are coupled. Gen. Res. **67**, 199–209 (1996)
41. Dorman, K.S., Sinsheimer, J.S., Lange, K.: In the garden of branching processes. SIAM Rev. **46**, 202–229 (2004)
42. Touré, Y.T., Petrarca, V., Traore, S.F., Coulibaly, A., Maiga, H.M., Sankare, O., Sow, M., Di Deco, M.A., Coluzzi, M.: The distribution and inversion polymorphism of chromosomally recognized taxa of the *An. gambiae* complex in Mali, West Africa. Parassitologica **40**, 477–511 (1998)
43. Taylor, C.E., Manoukis, N.C.: Effective population size in relation to genetic modification of *Anopholes gambiae sensu stricto*. In: Takken, W., Scott, T.W. (eds.) Ecological aspects for application of genetically modified mosquitoes, pp. 133–146. Wageningen, The Netherlands (2003)
44. Mackay, T.F., Lyman, R.F., Jackson, M.S.: Effects of *P* element insertions on quantitative traits in *Drosophila melanogaster*. Genetics **130**, 315–332 (1992)
45. Nuzhdin, S.V., Pasyukova, E.G., Mackay, T.F.C.: Accumulation of transposable elements in laboratory lines of *Drosophila melanogaster*. Genetica **100**, 167–175 (1997)
46. Maside, X., Assimacopoulos, S., Charlesworth, B.: Rates of movement of transposable elements on the second chromosome of *Drosophila melanogaster*. Genet. Res. **75**, 275–284 (2000)
47. Guimond, N., Bideshi, D.K., Pinkerton, A.D., Atkinson, P.W., O'Brochta, D.A.: Patterns of *Hermes* transposition in *Drosophila melanogaster*. Molec. Gen. Genet. **268**, 779–790 (2003)
48. Pasyukova, E.G., Nuzhdin, S.V., Morozova, T.V., Mackay, T.F.: Accumulation of transposable elements in the genome of *Drosophila melanogaster* is associated with a decrease in fitness. J. Hered. **95**, 284–290 (2004)
49. Charlesworth, B., Lapid, A., Canada, D.: 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 (1992)
50. Biémont, C., Vieira, C., Hoogland, C., Cizerion, G., Loevenbruck, C., Arnault, C., Carante, J.P.: Maintenance of transposable element copy number in natural populations of *Drosophila melanogaster* and *D. simulans*. Genetica **100**, 161–166 (1997)
51. Nuzhdin, S.V.: The distribution of transposable elements on *X* chromosomes from a natural population of *Drosophila simulans*. Genet. Res. **66**, 159–166 (1995)
52. Vieira, C., Biémont, C.: Transposition rate of the 412 retrotransposable element is independent of copy number in natural populations of *Drosophila simulans*. Mol. Biol. Evol. **14**, 185–188 (1997)
53. Seleme, M., Busseau, I., Malinsky, S., Bucheton, A., Teninges, D.: 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 (1999)
54. Vasilyeva, L.A., Bubenshchikova, E.V., Ratner, V.A.: Heavy heat shock induced retrotransposon transposition in *Drosophila*. Genet. Res. **74**, 111–119 (1999)
55. Yang, H.P., Hung, T.L., You, T.L., Yang, T.H.: Analysis of the highly-abundant transposable element *DINE*-1 suggests a recent transpositional burst in *Drosophila yakuba*. Genetics **173**, 189–196 (2006)
56. Biémont, C.: Dynamic equilibrium between insertion and excision of *P* elements in highly inbred lines from an M' strain of *Drosophila melanogaster*. J. Mol. Evol. **39**, 466–472 (1994)
57. Jensen, S., Gassama, M.P., Heidmann, T.: Taming of transposable elements by homology-dependent gene silencing. Nat. Genet. **21**, 209–212 (1999)
58. Biémont, C., Vieira, C., Borie, N., Lepetit, D.: Transposable elements and genome evolution: the case of *Drosophila simulans*. Genetica **107**, 113–120 (1999)
59. Nuzhdin, S.V., Pasyukova, E.G., Mackay, T.F.C.: Positive association between *copia* transposition rate and copy number in *Drosophila melanogaster*. Proc. R. Soc. Lond. B **263**, 823–831 (1996)
60. Pasyukova, E.G., Nuzhdin, S.V., Filatov, D.A.: The relationship between the rate of transposition and transposable element copy number for *copia* and *Doc* retrotransposons of *Drosophila melanogaster*. Genet. Res. **72**, 1–11 (1998)
61. Nuzhdin, S.V., Mackay, T.F.: Direct determination of retrotransposon transposition rates in *Drosophila melanogaster*. Genet. Res. **63**, 139–144 (1994)
62. Nuzhdin, S.V.: Sure facts, speculations, and open questions about the evolution of transposable element copy number. Genetica **107**, 129–137 (2000)
63. Eggleston, W.B., Johnson-Schlitz, D.M., Engels, W.R.: P–M hybrid dysgenesis does not mobilize other transposable element families in *D. melanogaster*. Nature **331**, 368–370 (1988)

64. Harris, T.: The Theory of Branching Processes. Springer, Berlin (1963)
65. Nuzhdin, S.V., Mackay, T.F.: The genomic rate of transposable element movement in *Drosophila melanogaster*. Mol. Biol. Evol. **12**, 180–181 (1995)
66. Mukai, T., Chigusa, S.I., Mettler, L.E., Crow, J.F.: Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. Genetics **72**, 335–355 (1972)
67. Ohnishi, O.: Spontaneous and ethyl methane–sulfonate-induced mutations controlling variability in *Drosophila melanogaster* I: recessive lethal mutations. Genetics **87**, 335–348 (1977)
68. Crow, J.F., Simmons, M.J.: The mutation load in *Drosophila*. In: Ashburner, M., Carson, H.L., Thompson, J.N. (eds.) In the genetics and biology of *Drosophila*, vol. 3c, pp. 1–35. Academic Press, London (1983)
69. Charlesworth, B.: Transposable elements in natural populations with a mixture of selected and neutral insertion sites. Genet. Res. **57**, 127–134 (1991)
70. Lanzaro, G., Touré, Y., Carnahan, J., Zheng, L., Dolo, G., Traore, S., Petrarca, V., Vernick, K.D., Taylor, C.E.: Complexities in the genetic structure of *Anopheles gambiae* populations in West Africa as revealed by microsatellite DNA analysis. Proc. Natl. Acad. Sci. USA **95**, 14260–14265 (1998)
71. Tripet, F., Dolo, G., Lanzaro, G.C.: Multilevel analyses of genetic differentiation in *Anopholes gambiae s.s.* reveal patterns of gene flow important for malaria-fighting mosquito projects. Genetics **169**, 313–324 (2005)
72. Touré, Y.T., Petrarca, V., Traore, S.F., Coulibaly, A., Maiga, H.M., Sankare, O., Sow, M., Di Deco, M.A., Coluzzi, M.: Ecological genetic studies in the chromosomal form Mopti of *Anopholes gambiae s. str.* in Mali, West Africa. Genetica **94**, 213–223 (1994)
73. Donnelly, M.J., Licht, M.C., Lehmann, T.: Evidence for recent population expansion in the evolutionary history of the malaria vectors *Anopholes arabiensis* and *Anopholes gambiae*. Mol. Biol. Evol. **18**, 1353–1364 (2001)
74. Coluzzi, M., Sabatini, A., della Torre, A., Di Deco, M., Petrarca, V.: A polytene chromosome analysis of the *Anopholes gambiae* complex. Science **298**, 1415–1418 (2002)
75. Ayala, F.J., Coluzzi, M.: Chromosome speciation: humans, *Drosophila* and mosquitoes. Proc. Natl. Acad. Sci. USA **102**, 6535–6542 (2005)
76. Baudry, E., Viginier, B., Veuille, M.: Non-African populations of *Drosophila melanogaster* have a unique origin. Mol. Biol. Evol. **8**, 1482–1491 (2004)
77. Thornton, K., Andolfatto, P.: Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. Genetics **172**, 1607–1619 (2006)
78. Pool, J.E., Bauer DuMont, V., Mueller, J.L., Aquadro, C.F.: A scan of molecular variation leads to narrow localization of a selective sweep affecting both Afrotropical and cosmopolitan populations of *Drosophila melanogaster*. Genetics **172**, 1093–1105 (2006)
79. Li, Y.J., Satta, Y., Takahata, N.: Paleo-demography of the *Drosophila melanogaster*subgroup: application of the maximum likelihood method. Genes Genet. Syst. **74**, 117–127 (1999)
80. Vanden Driessche, P., Watmough, J.: Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. **180**, 29–48 (2002)
81. Kaplan, N., Darden, T., Langley, C.H.: Evolution and extinction of transposable elements in Mendelian populations. Genetics **109**, 459–480 (1985)
82. Silva, J.C., Kidwell, M.G.: Horizontal transfer and selection in the evolution of *P* elements. Mol. Biol. Evol. **17**, 1542–1557 (2000)
83. Robertson, H.M., Soto-Adames, F.N., Walden, K.O., Avancini, R.M.P., Lampe, D.J.: The mariner transposons of animals: horizontally jumping genes. In: Syvanen, M., Kado, C.I. (eds.) Horizontal gene transfer, pp. 268–284. Chapman and Hall, New York (1998)
84. Labrador, M., Farre, M., Utzet, F., Fontdevila, A.: Interspecific hybridization increases transposition rates of *Osvaldo*. Mol. Biol. Evol. **16**, 931–937 (1999)
85. Bregliano, J.C., Kidwell, M.G.: Hybrid dysgenesis determinants. In: Shapiro, J.A. (ed.) Mobile genetic elements, p. 363. Academic Press, San Diego (1983)
86. Bucheton, A.: *I* transposable elements and $I - R$ hybrid dysgenesis in *Drosophila*. Trends Genet. **6**, 16–21 (1990)
87. Jensen, S., Gassama, M.P., Dramard, X., Heidmann, T.: Regulation of *I*-transposon activity in *Drosophila*: evidence for cosuppression of nonhomologous transgenes and possible role of ancestral *I*-related pericentromeric elements. Genetics **162**, 1197–1209 (2002)
88. Le Rouzic, A., Capy, P.: Reversible introduction of transgenes in natural populations of insects. Insect Mol. Biol. **15**, 227–234 (2006)

89. Tower, J., Karpen, G.H., Craig, N., Spradling, A.C.: Preferential transposition of *Drosophila P* elements to nearby chromosomal sites. Genetics **133**, 347–359 (1993)
90. Newfield, S.J., Takaesu, N.T.: Local transposition of a *hobo* element within the decapentaplegic locus of *Drosophila*. Genetics **151**, 177–187 (1999)
91. Hudson, A., O'Connor, M., McCall, K., Bender, W.: *P*-element homing within the bithorax complex. Ann. Conf. Dros. Res. **36**, 240A (1995)
92. Rubin, E., Levy, A.: Abortive gap repair: underlying mechanism for *Ds* element formation. Molec. Cell Biol. **17**, 6294–6304 (1997)
93. Lohe, A.R., Timmons, C., Beerman, I., Lozovskaya, E.R., Hartl, D.L.: Self-inflicted wounds, template-directed gap repair and a recombination hotspot: Effects of *mariner* transposase. Genetics **154**, 647–656 (2000)
94. Coluzzi, M., Sabatini, A.: Cytogenic observations on species A and B of the *Anopholes gambiae* complex. Parassitologia **9**, 71–88 (1967)
95. Irvin, N., Hoddle, M.S., O'Brochta, D.A., Carey, B., Atkinson, P.W.: Assessing fitness costs for transgenic *Aedes aegypti* expressing the GFP marker and transposase genes. Proc. Natl. Acad. Sci. USA **101**, 891–896 (2004)