Mathematical models as aids for design and development of experiments: The case of transgenic mosquitoes

Michael A. Robert, Mathieu Legros, Luca Facchinelli, Laura Valerio,
Janine M. Ramsey, Thomas W. Scott, Fred Gould, Alun L. Lloyd

Supplementary Material

§1. Mathematical Model

We present a discrete time, discrete state age-structured stochastic model that links population genetics and population dynamics to simulate the experimental design described in the main text. Our cohort-based model tracks, on a day-by-day basis, the (integer) numbers of adult mosquitoes, classified by age, sex, and genotype. We let $M_{g,a,d}$ be the number of adult males of genotype g and age a on day d, $F_{a,d}$ be the number of females of age a on day d, and $E_{g,d}$ be the number of eggs of genotype g that are laid on day d. We consider three possible genotypes: wild-type ($g = 1$), heterozygous FK ($g = 2$), and homozygous FK ($g = 3$). As discussed in the main text, eggs can be either wild-type or heterozygous, adult males can be of any genotype, although the only source of homozygotes are from FK releases, and the only reproductively-viable adult female genotype is wild-type (this assumes, as we do throughout this study, that the female killing mechanism is 100% effective).

An individual model simulation represents one experiment with one treatment and one control cage. For the duration of the experiment, N wild-type larvae are placed into the control cage on the first day of each week. The number of larvae placed in the treatment cage varies over time, reflecting the numbers of eggs laid in the treatment cage relative to the control cage. The same baseline number, N, is used throughout the stabilization period, but beginning with the week of the initial release, the number is determined by finding the input necessary to make the ratio of larvae input to eggs laid in the previous week equal in both the treatment and control cages. That is, the number of larvae placed in the treatment cage each week is N_w, where, at the beginning of week w,

$$N_w = N \cdot \frac{E_{w-1}^T}{E_{w-1}^C}$$

1

where, $E_{w-1}^T = \sum_{g=1}^{2} \sum_{d=1}^{7} E_{g,d}^T$ and $E_{w-1}^C = \sum_{d=1}^{7} E_{1,d}^C$ represent the total number of eggs laid in the treatment (T) and control (C) cages in the previous week. For the remainder of this section, we will discuss only the treatment cage unless otherwise noted, so we drop the superscripts that indicate the cage type.

As discussed in the main text, the genotype distribution of the N_w larvae introduced into the treatment cage in a given week reflects the genotype distribution of eggs laid in the cage in the previous week. The numbers of wild-type and heterozygous larvae, which we write as n_1 and n_2 respectively, are chosen from a multinomial distribution: $n = (n_1, n_2) \sim Multinomial(N_w, H_w)$, where $H_w = (h_{1,w}, h_{2,w})$, with

$$h_{g,w} = \frac{E_{g,w-1}}{E_{1,w-1} + E_{2,w-1}}.$$

Here, $h_{g,w}$, represents the frequency of genotype g in the eggs laid during week w.

Releases of homozygous FK pupae begin at week 14; these releases occur in addition to the N_w larvae returned from eggs laid in the cage. We take the release number to be a constant multiple of the baseline input. That is, the number of homozygous FK mosquitoes placed into the cage each week is rN. The release ratio, r, describes the ratio of FK pupae to wild-type larvae input. For example, for an input of 10:1 (FK : wild-type), $r = 10$.

A larval cohort comprises the individuals placed into a cage on a given day. The model simulates the emergence of these larvae based on an emergence time distribution, i.e. day-by-day probabilities that an individual will emerge, as shown in Figure 1 of the main text. These probabilities, together with a nonzero probability of juvenile mortality, sum to one. The numbers of individuals that emerge on each day following the placement of the cohort, or that die before emerging, is chosen from a multinomial distribution with these probabilities. (Figure 1 presents the emergence distribution for males. We assume females emerge one day later than males, therefore the distribution of female emergence is simply shifted by one day). The sex ratio of eggs laid is assumed to be 1:1, so the sexes of the emerging adults are determined by sampling from a binomial distribution with probability parameter 0.5. We remark that alternative experimental designs in which other immature age classes, i.e. eggs or pupae, are introduced into the cage could be modeled by an appropriate modification of the emergence time distribution.

For adult cohorts, we assume that there is a day-to-day survival probability for all mosquitoes in an age class. In the simplest model, the number of mosquitoes in an age class that survives from one day to the next is chosen from a binomial distribution with a constant daily survival probability for males (s_m) or females (s_f). More generally, the survival probability can be taken to be age-dependent.

The model keeps track of the genotype distribution of the offspring of a given female cohort. This is determined from the distribution of genotypes of males of mating age on the day (or days) on which the female cohort mated, weighted by the mating fitness cost of each genotype. We call this the **mating pair** distribution. Notice that we treat the offspring genotype distribution as a property of the cohort: this simplifies our model as it means we do not have to track individual females and their offspring genotype distributions, although it means we underestimate the genetic variability of eggs laid by the members of the cohort, particularly when only a few females remain in the cohort.

We first describe how the calculation of the offspring genotype distribution is undertaken when we assume females mate once. Using the mating fitness costs discussed in the main text (additive fitness cost of FK, cost per FK allele equal to $c/2$), the mating fitnesses of each genotype are $\Phi_1 = 1$, $\Phi_2 = 1 - c/2$, and $\Phi_3 = 1 - c$. Hence the probability that a female who mates on day d will mate with a male of genotype k, which we denote by $p_{k,d}$, is given by

$$p_{k,d} = \frac{\sum_a \Phi_k M_{a,k,d}}{\left(\sum_a \sum_{k'} \Phi_{k'} M_{a,k',d} \right)}.$$

Here, the sum is taken over all male age classes of mating age. Writing the Mendelian probability that an offspring of genotype g is born to a wild-type female who mates with a male of genotype k as $P_{g,k}$, the probability that an offspring of a female who mates on day d is of genotype g, denoted by $v_{g,d}$, is given by

$$v_{g,d} = \sum_k P_{g,k} p_{k,d}.$$

We assume that the number of eggs laid by a female mosquito each day is independent of her age and is Poisson distributed with mean λ. The total number of offspring of females of a given age class on a given day, $f_{a,d}$, is therefore Poisson distributed with mean $\lambda \cdot F_{a,d}$. The genotypes of these offspring are determined by sampling from a multinomial distribution: denoting the numbers of eggs of genotype g laid by females of age class a on day d by $e_{g,a,d}$, we have that $(e_{1,a,d}, e_{2,a,d}) \sim Multinomial(f_{a,d}, (v_{1,\hat{d}}, v_{2,\hat{d}}))$, where $\hat{d}$ is the day on which the female cohort mated.

For our simulations involving multiple matings, we consider the offspring distributions that result from matings on both days and form a weighted average of the two, imagining that a fraction z of offspring result from sperm from the first mating and the remaining $1 - z$ of offspring result from sperm from the second mating. We consider $1 - z$ to be the the **degree of polyandry** since the fraction $1 - z$ determines the influence of the second mating on the genotype distribution of the offspring. The numbers and genotypes of the eggs laid by the female cohort of age a on day d are given by $(e_{1,a,d}, e_{2,a,d}) \sim Multinomial(f_{a,d}, (zv_{1,\hat{d}_1} + (1 - z)v_{1,\hat{d}_2}, zv_{2,\hat{d}_1} + (1 - z)v_{2,\hat{d}_2}))$, where $\hat{d}_1$ and $\hat{d}_2$ are the first and second days on which the female cohort mated.

The total number of eggs of a given genotype laid each day is then the sum of the offspring of that genotype from all age classes of females. The distributions of genotypes among eggs laid are then used to determine the distribution of the genotypes of larvae input on a weekly basis as described previously in this section.