

# Population genetics of transgene containment

Ralph Haygood<sup>1,\*</sup>, Anthony R. Ives<sup>1</sup>, and David A. Andow<sup>2</sup>

<sup>1</sup>Department of Zoology, University of Wisconsin, Madison, WI 53706

<sup>2</sup>Department of Entomology and Center for Community Genetics, University of Minnesota, St. Paul, MN 55108

\*rhaygood@wisc.edu

*Ecology Letters* 7:213–220 (2004)

## **Abstract**

Several strategies have been proposed for creating transgenic cultivars from which transgene escape to wild relatives would seem unlikely; for example, to impede escape through pollen, a transgene could be inserted into chloroplast DNA (cpDNA), which in many crops is rarely transmitted through pollen. None of these strategies would be failsafe; for example, the rate of cpDNA transmission through pollen may be low but nonzero in many crops. Here, we study how the probability distribution of escape time depends on the rates of pollen and seed flow from the crop to wild populations, the number and sizes of the wild populations, the selection coefficient for the transgene, and a leakage parameter characteristic of the strategy, for example, the rate of cpDNA transmission through pollen. We find that that even with a leakage parameter as small as  $10^{-3}$ , the probability of escape within as few as ten generations could be appreciable. This finding challenges technologists to develop strategies with very small leakage parameters and regulatory officials to develop appropriate procedures for risk assessment.

## Introduction

Concern about gene flow from crops to wild relatives has become widespread with the increasing cultivation of transgenic crops (Snow and Morán Palma, 1997; Hails, 2000; Ellstrand, 2001). Much of this concern is due to the expectation that upon acquiring transgenes promoting resistance to pests, diseases, or other stresses, some wild species would become aggressive invaders of natural areas or farmland (e.g., Snow et al., 2003). There is also concern about deleterious effects of pesticides encoded by transgenes on nontarget organisms (e.g., Ponsard et al., 2002) and about accelerated evolution of resistance to such pesticides in target organisms (e.g., Gould, 1998), problems that would be aggravated if wild species acquired such transgenes. These concerns are among the motivations for transgene containment strategies, which aim to allow the cultivation of transgenic crops with lower risk of transgene escape.

Problematic consequences of crop–wild gene flow are not limited to transgenes or to crop genes that would be selectively favored in wild populations (Haygood et al., 2003). However, some proposed containment strategies are oriented toward transgenes, in that they require control of where the crop gene to be contained is located (e.g., in chloroplast rather than nuclear DNA (Daniell et al., 1998), tightly linked to a so-called mitigation gene (Gressel, 1999), or flanked by motifs recognized by a site-specific recombinase (Keenan and Stemmer, 2002)). Moreover, the crop genes most likely to escape and hence most in need of containment are those that would be selectively favored in wild populations. Several containment strategies suitable for transgenes that would be selectively favored in wild populations have been proposed (Daniell, 2002). None of them would be failsafe, so transgene escape would be inevitable if pollen or seed flow from the crop to wild populations continued indefinitely. However, if the escape time would most likely exceed, say, a thousand years, then this inevitability might be practically unimportant. Thus the probability distribution of escape time calls for investigation.

Here, we present the first quantitative analyses of transgene containment strategies. Specifically, using mathematical models of wild populations recurrently receiving pollen and seeds from

a genetically fixed crop, we study how the probability distribution of escape time depends on the rates of pollen and seed flow, the number and sizes of the wild populations, the selection coefficient for the transgene under wild conditions, and a leakage parameter characteristic of the strategy. We suppose the transgene is selectively favored under wild conditions, and we define escape as initial establishment leading to eventual fixation of the transgene in a wild population. First, we analyze escape through pollen of a transgene inserted into chloroplast DNA, a strategy that has been greeted with enthusiasm (Bilang and Potrykus, 1998; Gray and Raybould, 1998; Sansom, 2003) and seems likely to be commercialized quickly (e.g., <http://www.chlorogen.com>). This detailed analysis of chloroplast transformation introduces ideas relevant to many other strategies, and next, we briefly indicate how similar reasoning applies to some other strategies. Finally, we consider the impact of seed flow in addition to pollen flow. The most striking implication of our analyses is that even with a leakage parameter as small as  $10^{-3}$ , the probability of escape within as few as ten generations could be appreciable.

## **Chloroplast transformation**

Transgenes have been experimentally inserted into chloroplast DNA (cpDNA) in several crops, including tobacco (Svab et al., 1990), potato (Sidorov et al., 1999), and tomato (Ruf et al., 2001). From a technological perspective, so-called transplastomic plants are attractive in several ways. For example, each cell contains many copies of the chloroplast genome, which might aid in obtaining high concentrations of transgene products. The most important attraction is that, assuming cpDNA is rarely transmitted through pollen, the risk of transgene escape through pollen would seem low. As noted above, this prospect has been greeted with enthusiasm; for example, one scientist has asserted that “chloroplast transformation...will completely control gene flow.” (M. McGloughlin in Lambrecht, 2001, p. 83) However, as others have pointed out (Cummins, 1998; Stewart and Prakash, 1998), this enthusiasm has been overstated. In many crops, the rate of cpDNA transmission through pollen is not high, but exceptions, such as alfalfa (Schumann and Hancock, 1989) and kiwifruit (Chat et al., 1999), are known. Moreover, low but nonzero transmission rates have

been measured in several crops, including tobacco (Medgyesy et al., 1986) and potato (Simmonds, 1969). Low but nonzero transmission rates may be common; measuring low transmission rates requires large samples, and for most crops, large samples have not been analyzed yet (Milligan, 1992). Assuming cpDNA is ever transmitted through pollen, transgene escape would be inevitable if pollen flow continued indefinitely, so the probability distribution of escape time calls for investigation.

Consider a wild population that recurrently receives a small amount of pollen from a closely related crop fixed for a chloroplast transgene. Specifically, let  $m$  be the fraction of the pollen in the population that flows from the crop per generation ( $0 < m \ll 1$ ). For now, assume there is no seed flow; later, we will drop this assumption. For simplicity, assume there is random union of gametes. In particular, assume there is no preferential self-fertilization and no incompatibility between crop pollen and wild stigmas, styles, or ovules. (Alternatively, assume  $m$  is an effective rate accounting for these complications.) It follows that in the first generation after pollen flow begins, an expected fraction  $m$  of the zygotes are hybrid. (Here and elsewhere in this analysis, we ignore a variance, which has a relatively small effect, assuming pollen is abundant.)

Hybrid plants might be inferior or superior to wild ones in survival or fecundity. Assuming there is little cpDNA transmission through pollen, such fitness differences would be due not to the chloroplast transgene but to other genetic differences between the crop and the wild species. For now, assume there are no such genetic differences, or their effects on fitness are negligible; later, we will drop this assumption, although it is a good approximation for some crops and wild relatives (e.g., sorghum and johnsongrass under agricultural conditions, Arriola and Ellstrand, 1997). For simplicity, assume the wild species is annual with no seed bank. It follows that in the first generation after pollen flow begins, an expected fraction  $m$  of the mature plants are hybrid, and they produce an expected fraction  $m$  of the ovules.

Assume a hybrid zygote can inherit from its father a small number of transgenic proplastids (precursors of chloroplasts and other plastids). Presumably, most hybrid zygotes inherit no trans-

genic proplastids, a few inherit one, fewer inherit two, etc. (Birky, 2001) As a plant develops from a heteroplasmic zygote, its cells tend to become individually homoplasmic (Birky, 2001). (The process resembles the fixation of an allele at an initially polymorphic locus through genetic drift in a finite population, although it tends to proceed more quickly (Birky, 2001).) Thus most cells of the mature plant are individually homoplasmic. Which plastid type, paternal or maternal, fixes in a cell lineage is a matter of chance, assuming neither plastid type has a replicative advantage (Birky, 2001). Thus the mature plant is a mosaic of the plastid types. In particular, a fraction of its gametes are fixed for the paternal plastid type. Let  $\ell$  be the expected fraction ( $0 < \ell \ll 1$ ), which is closely related to the rate of cpDNA transmission through pollen. It follows that in the first generation after pollen flow begins, an expected fraction  $m\ell$  of the ovules are transgenic.

If a transgenic ovule is fertilized by a transgenic pollen grain, then the zygote is fully transgenic, and if a transgenic ovule is fertilized by a nontransgenic pollen grain, then the zygote is either fully transgenic or nearly so, assuming there is little cpDNA transmission through pollen. Henceforth, we ignore the distinction. Thus if a transgenic ovule is fertilized, and the transgenic zygote is recruited, then the second generation after pollen flow begins contains a transgenic plant. Let  $N$  be the number of mature plants in the wild population. It follows that the probability that the second generation contains at least one transgenic plant is approximately  $1 - (1 - m\ell)^N$ . (Here we ignore several variances and the selective advantage of transgenic plants, which have relatively small effects.) Assuming  $m\ell N \ll 1$ , the probability that the second generation contains at least one transgenic plant reduces to approximately  $m\ell N$ , and the probability that the second generation contains more than one transgenic plant is negligible in comparison.

Even if the second generation contains a transgenic plant, subsequent generations may not, due to demographic stochasticity; the selective advantage of transgenic plants is an average effect, and even if it is rather large, until the transgene becomes fairly common, it is vulnerable to loss from the wild population through random reproductive failures of transgenic plants. Let  $s$  be the selection coefficient for the transgene under wild conditions ( $0 < s$ ), that is, let  $1 + s$  be the ratio

of the expected number of offspring of a transgenic plant to the expected number of offspring of a nontransgenic plant. By standard reasoning (e.g., Gale, 1990, ch. 4; Gillespie, 1998, sec. 3.7), the probability that a newly founded transgenic lineage is destined not for extinction but for fixation is approximately  $1 - e^{-2s}$ , assuming  $Ns \gg 1$ . (If also  $s \ll 1$ , then this expression reduces to approximately  $2s$ .) It follows that the probability that the second generation after pollen flow begins contains a transgenic plant whose descendants take over the population is approximately

$$p = m\ell N(1 - e^{-2s}). \quad (1)$$

In this case, we say that the transgene escapes in the second generation. This definition of transgene escape requires eventual fixation, which may seem stringent. However, once a favored gene reaches a moderate frequency, eventual fixation is practically certain (Gale, 1990). Of course, the initial frequency of the escaped transgene is  $1/N$ , and the frequency may stay low for many generations. The mean fixation time is approximately  $2 \ln[N]/s$  generations (Gale, 1990).

If the transgene does not escape in the second generation after pollen flow begins, then the situation is essentially the same in subsequent generations until the transgene escapes or pollen flow ends. In these generations, there may be a small amount of transgenic pollen produced in the wild population and a small number of transgenic plants in lineages destined for extinction, but the effects of these differences from the second generation are small. For simplicity, assume  $m$ ,  $\ell$ ,  $N$ , and  $s$  are the same in each generation. To a good approximation, the process is one of repeated independent trials (generations) with probability of success (transgene escape)  $p$  in each trial. It follows that the probability distribution of escape time is approximately geometric with mean  $1/p$  generations (Ross, 2002).

We have made several approximations, such as neglecting the differences between the second generation after pollen flow begins and subsequent generations until the transgene escapes. It seems prudent to put these approximations to the test of simulation. Figure 1 presents the frequency distribution of escape time over many individual-based simulations (see appendix for details) and the geometric distribution with mean  $1/p$  generations for one set of parameter values. The simu-

lation distribution is well approximated by the geometric one. The same holds for many other sets of parameter values.

The mean escape time is approximately

$$\mu = \frac{1}{m\ell N(1 - e^{-2s})} \quad (2)$$

generations. For example, suppose  $m = 0.1$  and  $N = 100$ , which are plausible for a wild population on the margin of a cultivated field, suppose  $\ell = 0.025$ , which is suggested by an experiment on cpDNA transmission through pollen in tobacco (Medgyesy et al., 1986), and suppose  $s = 0.1$ , which is plausible for a transgene promoting resistance to a major pest. Then  $\mu = 22$  generations. This example may be atypical, but these parameter values are not preposterous.

More importantly, rather than rising to a peak, the probability density of escape time decreases monotonically, so there is substantial probability of escape well before the mean escape time. Indeed, the probability of escape within  $\alpha\mu$  generations is approximately  $1 - e^{-\alpha}$  ( $0 \leq \alpha$ ); for example, if  $\mu = 100$  generations, then the probability of escape within 10 generations is approximately  $1 - e^{-0.1} \approx 10\%$ .

So far, we have considered a single wild population, but if the crop is widely cultivated, then many wild populations may receive pollen from it. Let  $n$  be the number of populations, and for simplicity, assume  $p$  is the same in each population. The probability per generation that the transgene escapes in at least one population is approximately  $1 - (1 - p)^n$ , which reduces to approximately  $pn$ , assuming  $pn \ll 1$ . The probability distribution of time to escape in at least one population is approximately geometric with mean  $1/(pn)$  generations; for example, if the mean escape time in each population is approximately  $1/p = 1000$  generations, but there are  $n = 100$  populations, then the mean escape time over all populations is approximately  $1/(pn) = 10$  generations.

The least satisfactory aspect of this analysis is the assumption that apart from the chloroplast transgene, there are no genetic differences between the crop and the wild species, or their effects on fitness are negligible. A crop and a closely related wild species are likely to differ at several

loci (e.g., maize and teosinte differ at five loci of large effect, Lynch and Walsh, 1998, ch. 15), and hybrid plants are often inferior to wild ones (e.g., sunflower, Snow et al., 1998). Hybrid inferiority slows transgene escape by lowering the rate at which transgenic lineages are founded and raising the rate at which they go extinct. However, the slowing is unlikely to be great unless the inferiority is severe. Presumably, loci underlying hybrid inferiority are more likely to be in the nuclear genome than in the chloroplast genome, the simplest possibility being a single nuclear locus. For simplicity, assume the chloroplast transgene and the nuclear locus determine fitness multiplicatively (or nearly so), which is likely when they affect different life stages (or mainly so). Let  $s'$  be the selection coefficient against heterozygotes under wild conditions ( $0 < s' \leq 1$ ), that is, let  $1 - s'$  be the ratio of the expected number of offspring of a heterozygous plant to the expected number of offspring of a wild-type homozygous plant, where both plants are transgenic or both plants are nontransgenic. (Assuming  $m \ll 1$ , the selection coefficient against crop-type homozygotes is relatively unimportant.) It can be shown (see appendix for details) that the probability distribution of escape time remains approximately geometric, and provided  $s'$  is not too small, the mean escape time increases by a factor of about  $(1 + s')/(1 - s')$ ; for example, for the same parameter values as in Figure 1 and a single nuclear locus with  $s' = 0.5$ , the mean escape time increases by a factor of about 3 (the mean over  $10^4$  simulations is 339 generations). When not one locus but several loci underlie hybrid inferiority, transgene escape may be slowed further, but the effect of spreading the same inferiority over more loci is unlikely to be large; for example, when the same selection coefficient against heterozygotes (0.5) is spread over not one nuclear locus but ten nuclear loci with free recombination, multiplicative fitness determination, and the same selection coefficient per locus ( $1 - 0.5^{1/10}$ ), the mean escape time increases further by about 5% (the mean over  $10^4$  simulations is 357 generations).

## **Other strategies**

Although escape of a chloroplast transgene is unlikely to be delayed much by moderately disfavored nuclear genes, transgene escape can be greatly slowed when a transgene is tightly linked to

a moderately disfavored gene in the same genome. This possibility is the essence of another proposed strategy, known as transgenic mitigation (Daniell, 2002). Given a gene that is desirable or neutral under agricultural conditions but disfavored under wild ones, a transgene could be inserted along with this so-called mitigation gene in a tandem construct (Gressel, 1999). This proposal has not been experimentally demonstrated yet, but candidate mitigation genes may become available soon. Assuming hybrids are not sterile and linkage is not complete, transgene escape would be inevitable if pollen flow continued indefinitely. Consider a transgene–mitigation gene construct in the nuclear genome. By reasoning similar to that for chloroplast transformation, the probability distribution of escape time is approximately geometric with mean

$$\mu = \frac{s'}{m(1-s')rN(1-e^{-2s})} \quad (3)$$

generations, where  $m$  and  $N$  are as before,  $s'$  is the selection coefficient against first-generation hybrids ( $0 \ll s' < 1$ ),  $r$  is the frequency of recombination between the genes or within the mitigation gene ( $0 < r \ll 1/2$ ), and  $s$  is the selection coefficient for transgenic recombinant backcrosses ( $0 < s$ );  $m(1-s')/s'$  is approximately the immigration–selection balance frequency of mature plants having one copy of the construct,  $r/2$  is the expected fraction of the gametes from these plants having the transgene but lacking the mitigation gene, multiplying by  $2N$  gives approximately the probability that one of these gametes contributes to founding a transgenic lineage, and  $1 - e^{-2s}$  is approximately the probability that this lineage does not go extinct. Equation (3) is the same as (2), except that  $(1-s')r/s'$  replaces  $\ell$ . The most important attraction of this strategy is that  $r$  could be smaller than  $10^{-5}$ .

Similar reasoning applies to many other strategies. Typically, the probability distribution of escape time is approximately geometric with mean approximately inversely proportional to the rate of pollen flow, the size of the wild population, the selection coefficient for the transgene under wild conditions, and a leakage parameter characteristic of the strategy. For example, a transgene could be inserted flanked by motifs recognized by a site-specific recombinase, the gene for the recombinase could be inserted under the control of a chemically inducible promotor, and

the inducer could be applied to activate the recombinase and hence excise the transgene before the crop flowers (Keenan and Stemmer, 2002). Assuming activation and excision are not perfect, their failure rate plays the same role for this strategy as  $\ell$  for chloroplast transformation or  $(1 - s')r/s'$  for transgenic mitigation.

## Seed flow

Crop pollen may easily spread, via wind, pollinating animals, and other mechanisms, so it is reasonable for concern about transgene escape to focus on pollen flow. However, crop seeds may also spread, via spillage in handling, removal by animals, and other mechanisms, and seed flow contributes appreciably to gene flow between some crops and wild relatives (e.g., beet, Arnaud et al., 2003). In principle, seed flow might be more problematic than pollen flow, in that once crop seeds reach an area, crop plants may be able to invade without hybridizing with wild plants. We will assume crop plants do hybridize with wild plants, but the possibility of invasion without hybridization should be kept in mind.

Transgene containment strategies vary in whether they impede escape through pollen only or through both pollen and seeds. Assuming most cpDNA transmission is through ovules, a chloroplast transgene is protected against escape through pollen only. In a wild population, if a crop seed germinates, and the crop seedling survives, then a transgenic lineage has been founded, and if this lineage is not lost to demographic stochasticity, then the transgene has escaped. Let  $\tilde{m}$  be the fraction of the seeds in the population that flow from the crop per generation ( $0 < \tilde{m} \ll 1$ ). Reasoning much as above, the probability that the first generation after seed flow begins contains a transgenic plant is approximately  $\tilde{m}N$ . It follows that the probability distribution of escape time is approximately geometric with mean

$$\mu = \frac{1}{(m\ell + \tilde{m})N(1 - e^{-2s})} \quad (4)$$

generations. Equation (4) is the same as (2), except that  $m\ell + \tilde{m}$  replaces  $m\ell$ . Assuming  $\ell \ll 1$ , it may be that  $\tilde{m} > m\ell$  even if  $m \gg \tilde{m}$ ; seed flow may be the more likely escape route, even if pollen flow is much more prevalent.

In contrast, a transgene tightly linked to a mitigation gene is protected against escape through both pollen and seeds. If crop plants are not sterile under wild conditions, then they produce crop pollen and ovules, which contribute to hybridization. Let  $s''$  be the selection coefficient against crop plants under wild conditions ( $0 \ll s'' < 1$ ). The immigration–selection balance frequency of crop plants is approximately  $\tilde{m}(1 - s'')$ , and their pollen and ovules alike have the transgene–mitigation gene construct. It follows that the probability distribution of escape time is approximately geometric with mean

$$\mu = \frac{s'}{(m + 2\tilde{m}(1 - s''))(1 - s')rN(1 - e^{-2s})} \quad (5)$$

generations. Equation (5) is the same as (3), except that  $m + 2\tilde{m}(1 - s'')$  replaces  $m$ . If  $m \gg \tilde{m}$ , then pollen flow is the more likely escape route.

## Discussion

Our analyses suggest at least two important general conclusions about transgene containment. First, because the probability distribution of escape time is approximately geometric, there is substantial probability of escape well before the mean escape time. Second, because it can expose many wild populations to crop pollen and seeds, widespread cultivation of the crop can substantially aggravate the problem. More concretely, our analyses suggest that transgene containment strategies with leakage parameters larger than  $10^{-3}$  may fail rather quickly; for example, if  $m = 0.05$ ,  $N = 100$ ,  $s = 0.1$ , and the leakage parameter ( $\ell$ ,  $(1 - s')r/s'$ , etc.) is  $10^{-3}$  in each of  $n = 100$  wild populations, then in the absence of additional impediments such as disfavored crop alleles at other loci and even in the absence of seed flow, the probability of escape in at least one population within 10 generations is approximately 60%.  $10^{-3}$  might seem small, but it might not be small enough. There may be a genuine need for smaller leakage parameters.

A very small leakage parameter might be obtained by using two containment strategies for one transgene. For example, if a mitigation gene were available that would function properly in the chloroplast genome, then the probability distribution of escape time in a single wild population

for a transgene inserted into cpDNA along with this gene would be approximately geometric with mean  $1/(mlrN(1 - e^{-2s}))$  generations. (This expression does not contain  $s'$ , assuming the mitigation gene is sufficiently disfavored that a transgenic lineage having this gene is practically certain to go extinct.) The leakage parameter  $lr$  could be smaller than  $10^{-6}$ . Such two-fold strategies have been advocated before (Johnson and Dallimore, 2002), and our analyses strengthen the case for them.

We have assumed transgene insertions are stable, but they need not be; for example, a chloroplast transgene might be transferred into the nucleus (Huang et al., 2003), and if it happened to become inserted near a nuclear promoter, then it might be expressed from the nucleus. We have also assumed transgene expression is stable, but it need not be; for example, a mitigation gene might be silenced, while a linked transgene might still be expressed. Although such instabilities seem likely to be very rare, they should probably be considered when evaluating transgene containment strategies.

We have considered crop-wild hybridization, but there is also concern about crop-crop hybridization (Ellstrand, 2001). Pollen from a transgenic cultivar might reach another cultivar of the same crop growing nearby, leading to hybrid seeds. If seeds were accidentally left behind when the recipient cultivar was harvested, then a process of transgene escape would ensue, much like the one we have considered. Moreover, the transgene might be expressed in hybrid seeds, raising concern about whether products derived from them were safe to use. Transgene containment strategies vary in whether they address these issues. Transgenic mitigation need not address either issue, because the mitigation gene need not be disfavored under the growing conditions of the recipient cultivar, and it need not prevent expression of the transgene in hybrid seeds. In contrast, chloroplast transformation addresses both issues, provided there is only pollen flow. Chemically induced transgene excision addresses both issues, whether there is only pollen flow or both pollen and seed flow.

Our analyses simplify the population genetics of transgene containment to its essentials. We recognize the need for more detailed modeling of particular crops and wild relatives. Practical

risk assessment will require modeling informed by measurements to account for many factors our analyses neglect, perhaps especially causes of nonrandom mating such as preferential self-fertilization, gametic incompatibilities, and phenological differences among crop, wild, and hybrid plants. However, we believe that our analyses identify the most important factors, and we expect that our general conclusions will prove robust.

## Appendix

### Simulations

For a chloroplast transgene, we reasoned that the probability distribution of escape time is approximately geometric with mean given by (2). In the simulations we ran to test this reasoning, each of the  $N$  plants in the wild population is represented individually. Plant  $i$  is characterized by the frequency  $q_i$  of the transgene in its cpDNA. Its relative fitness is  $w_i = (1 + s)^{\chi_i}$ , where  $\chi_i = 0$  if  $q_i < 0.5$  or  $\chi_i = 1$  if  $q_i \geq 0.5$ ; this is a crude model, but other reasonable models (e.g.,  $w_i = 1 + q_i s$ ) yield similar results. Initially, the transgene is absent, that is,  $q_i = 0$  for each  $i$ . For simplicity, the wild species is supposed monoecious, and from one generation, the next is constructed as follows. For each offspring, a mother  $M$  is picked with fitness-proportionate weighting, that is,  $M = i$  with probability  $w_i / \sum_j w_j$ . One of  $M$ 's ovules is picked at random. It is fully transgenic with probability  $q_M$  and fully nontransgenic with probability  $1 - q_M$ . With probability  $1 - m$ , a father  $P$  and one of  $P$ 's pollen grains are picked in the same way. Otherwise, the father is a crop plant, and all its pollen grains are fully transgenic. The offspring inherits 50 proplastids, and the number from the father is picked from a binomial distribution with mean  $50\ell$  and variance  $50\ell(1 - \ell)$ ; this is a crude model, but other reasonable models yield similar results. Simulation continues until the transgene is fixed, that is,  $q_i = 1$  for each  $i$ . The reported escape time is the earliest generation such that from this generation on, there was always at least one transgenic plant, that is, a plant arising from a fully transgenic ovule. Strictly speaking, this time may be an underestimate, because a transgenic lineage that eventually goes extinct may be present when the transgenic lineage that eventually fixes is founded. However, assuming  $m\ell N \ll 1$ , the

rate at which transgenic lineages are founded is low, and the underestimation of escape time is negligible.

In the simulations mentioned later in the text, plant  $i$  is further characterized by its genotype at a set of diploid nuclear loci. Fitness determination is multiplicative, with selection coefficients  $s'$  and  $s'' = 2s'$  against heterozygotes and crop homozygotes at each locus; if plant  $i$  is a heterozygote at  $L'_i$  loci and a crop homozygote at  $L''_i$  loci, then  $w_i = (1 + s)^{x_i} (1 - s')^{L'_i} (1 - s'')^{L''_i}$ .

We ran similar simulations to test (3), and the simulation distributions were well approximated by the geometric ones.

### **Effect of a disfavored nuclear gene on escape of a chloroplast transgene**

The analysis is identical to the one in the main text up to the conclusion that in the first generation after pollen flow begins, an expected fraction  $m$  of the zygotes are hybrid. These zygotes are heterozygous for the nuclear gene, so they give rise to an expected fraction  $m(1 - s')$  of the ovules. Reasoning as before, an expected fraction  $m(1 - s')\ell$  of the ovules are transgenic, and the probability that the second generation after pollen flow begins contains a transgenic plant is approximately  $m(1 - s')\ell N$ . With probability  $1/2$ , the ovule from which this plant arose did not carry the nuclear gene, in which case, assuming  $m \ll 1$ , the plant is most likely a wild-type homozygote at the nuclear locus. In this case, the probability that the newly founded transgenic lineage is destined not for extinction but for fixation is again approximately  $1 - e^{-2s}$ . Alternatively, with probability  $1/2$ , the ovule did carry the nuclear gene, in which case the plant is most likely a heterozygote at the nuclear locus. In this case, the expected number of offspring of the plant in the third generation after pollen flow begins is approximately  $1 - s'$ . Each offspring is a wild-type homozygote with probability approximately  $1/2$  and a heterozygote with probability approximately  $1/2$ . Iterating this reasoning, the probability that the transgene escapes in the second generation is approximately

$$\begin{aligned} p &= m(1 - s')\ell N \left( \frac{1}{2} + \frac{1}{2}(1 - s') \left( \frac{1}{2} + \frac{1}{2}(1 - s') \left( \frac{1}{2} + \dots \right) \right) \right) (1 - e^{-2s}) \\ &= m(1 - s')\ell N \frac{1}{2} \left( 1 + \frac{1}{2}(1 - s') + \frac{1}{4}(1 - s')^2 + \dots \right) (1 - e^{-2s}) \end{aligned}$$

$$\begin{aligned}
&= m(1 - s')\ell N \frac{\frac{1}{2}}{1 - \frac{1}{2}(1 - s')} (1 - e^{-2s}) \\
&= m\ell N(1 - e^{-2s}) \frac{1 - s'}{1 + s'},
\end{aligned}$$

which is (1) multiplied by  $(1 - s')/(1 + s')$ . Assuming  $s' \gg 1/N$  and  $s' \gg s$ , we have neglected the effects of genetic drift and the chloroplast transgene on the dynamics of the nuclear gene.

## Acknowledgments

We thank R. Goodman and M. Turelli for reviewing drafts of this article and the referees for their suggestions. Our research was supported by a USDA–IFAFS grant.

## References

- Arnaud, J.-F., Viard, F., Delescluse, M., and Cuguen, J., 2003. Evidence for gene flow via seed dispersal from crop to wild relatives in *Beta vulgaris* (Chenopodiaceae): Consequences for the release of genetically modified crop species with weedy lineages. *Proceedings of the Royal Society of London B, Biological Sciences* **270**:1565–1571.
- Arriola, P. E., and Ellstrand, N. C., 1997. Fitness of interspecific hybrids in the genus *Sorghum*: Persistence of crop genes in wild populations. *Ecological Applications* **7**:512–518.
- Bilang, R., and Potrykus, I., 1998. Containing excitement over transplastomic plants. *Nature Biotechnology* **16**:333–334.
- Birky, Jr., C. W., 2001. The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. *Annual Review of Genetics* **35**:125–148.
- Chat, J., Chalak, L., and Petit, R. J., 1999. Strict paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in intraspecific crosses of kiwifruit. *Theoretical and Applied Genetics* **99**:314–322.

- Cummins, J. E., 1998. Chloroplast transgenic plants are not a gene flow panacea. *Nature Biotechnology* **16**:401.
- Daniell, H., 2002. Molecular strategies for gene containment in transgenic crops. *Nature Biotechnology* **20**:581–586.
- Daniell, H., Datta, R., Varma, S., Gray, S., and Lee, S.-B., 1998. Containment of herbicide resistance through genetic engineering of the chloroplast genome. *Nature Biotechnology* **16**:345–348.
- Ellstrand, N. C., 2001. When transgenes wander, should we worry? *Plant Physiology* **125**:1543–1545.
- Gale, J. S., 1990. *Theoretical Population Genetics*. Unwin Hyman, London.
- Gillespie, J. H., 1998. *Population genetics: A concise guide*. Johns Hopkins University Press, Baltimore, MD.
- Gould, F., 1998. Sustainability of transgenic insecticidal cultivars: Integrating pest genetics and ecology. *Annual Review of Entomology* **43**:701–726.
- Gray, A. J., and Raybould, A. F., 1998. Reducing transgene escape routes. *Nature* **392**:653–654.
- Gressel, J., 1999. Tandem constructs: Preventing the rise of superweeds. *Trends in Biotechnology* **17**:361–366.
- Hails, R. S., 2000. Genetically modified plants: The debate continues. *Trends in Ecology and Evolution* **15**:14–18.
- Haygood, R., Ives, A. R., and Andow, D. A., 2003. Consequences of recurrent gene flow from crops to wild relatives. *Proceedings of the Royal Society of London B, Biological Sciences* **270**:1879–1886.
- Huang, C. Y., Ayliffe, M. A., and Timmis, J. N., 2003. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* **422**:72–76.

- Johnson, B., and Dallimore, R., 2002. Beyond gene containment. *Nature Biotechnology* **20**:871.
- Keenan, R. J., and Stemmer, W. P. C., 2002. Nontransgenic crops from transgenic plants. *Nature Biotechnology* **20**:215–216.
- Lambrecht, B., 2001. *Dinner at the New Gene Café: How Genetic Engineering Is Changing What We Eat, How We Live, and the Global Politics of Food*. St. Martin's Press, New York.
- Lynch, M., and Walsh, B., 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland, MA.
- Medgyesy, P., Páy, A., and Márton, L., 1986. Transmission of paternal chloroplasts in *Nicotiana*. *Molecular and General Genetics* **204**:195–198.
- Milligan, B. G., 1992. Is organelle DNA strictly maternally inherited? Power analysis of a binomial distribution. *American Journal of Botany* **79**:1325–1328.
- Ponsard, S., Gutierrez, A. P., and Mills, N. J., 2002. Effect of *Bt*-toxin (Cry1Ac) in transgenic cotton on the adult longevity of four heteropteran predators. *Environmental Entomology* **31**:1197–1205.
- Ross, S. M., 2002. *A First Course in Probability*, 6th ed. Prentice-Hall, Upper Saddle River, NJ.
- Ruf, S., Hermann, M., Berger, I. J., Carrer, H., and Bock, R., 2001. Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. *Nature Biotechnology* **19**:870–875.
- Sansom, C., 2003. Chloroplast research could mean safer GM crops. *Frontiers in Ecology and the Environment* **1**:62.
- Schumann, C. M., and Hancock, J. F., 1989. Paternal inheritance of plastids in *Medicago sativa*. *Theoretical and Applied Genetics* **78**:863–866.
- Sidorov, V. A., Kasten, D., Pang, S.-Z., Hajdukiewicz, P. T. J., Staub, J. M., and Nehra, N. S., 1999. Stable chloroplast transformation in potato: Use of green fluorescent protein as a plastid marker. *The Plant Journal* **19**:209–216.

- Simmonds, N. W., 1969. Variegated mutant plastid chimeras of potatoes. *Heredity* **24**:303–306.
- Snow, A. A., Moran-Palma, P., Rieseberg, L. H., Wszelaki, A., and Seiler, G. J., 1998. Fecundity, phenology, and seed dormancy of F1 wild–crop hybrids in sunflower (*Helianthus annuus*, Asteraceae). *American Journal of Botany* **85**:794–801.
- Snow, A. A., and Morán Palma, P., 1997. Commercialization of transgenic plants: Potential ecological risks. *Bioscience* **47**:86–96.
- Snow, A. A., Pilson, D., Rieseberg, L. H., Paulsen, M. J., Pleskac, N., Reagon, M. R., Wolf, D. E., and Selbo, S. M., 2003. A Bt transgene reduces herbivory and enhances fecundity in wild sunflowers. *Ecological Applications* **13**:279–286.
- Stewart, Jr., C. N., and Prakash, C. S., 1998. Chloroplast transgenic plants are not a gene flow panacea. *Nature Biotechnology* **16**:401.
- Svab, Z., Hajdukiewicz, P., and Maliga, P., 1990. Stable transformation of plastids in higher plants. *Proceedings of the National Academy of Sciences of the United States of America* **87**:8526–8530.

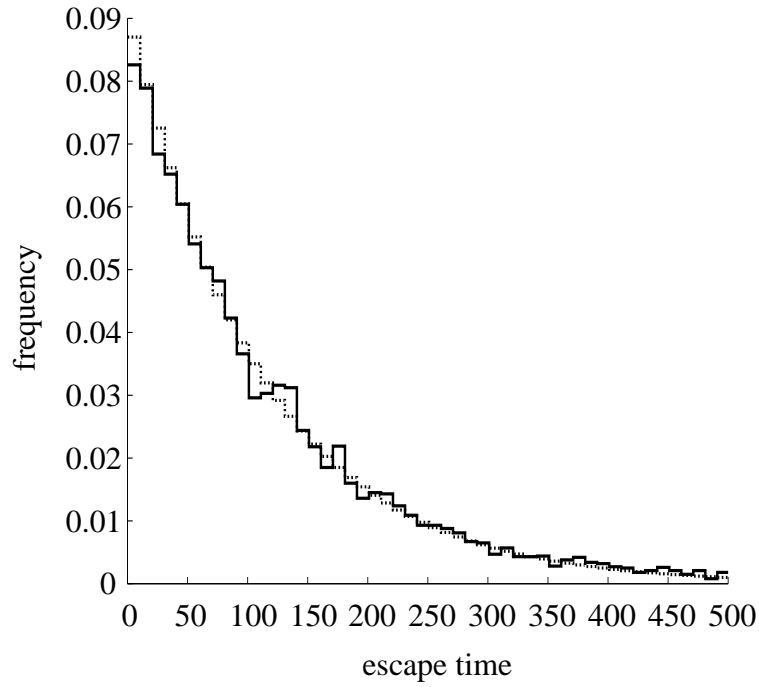


Figure 1: The probability density of escape time for a chloroplast transgene in a wild population where  $m = 0.05$ ,  $\ell = 0.01$ ,  $N = 100$ , and  $s = 0.1$ . The solid histogram represents the frequency distribution of escape time over  $10^4$  individual-based simulations (see appendix for details); the mean is 112 generations. The dotted histogram represents the geometric distribution with mean  $1/p = 110$  generations.