

# Consequences of recurrent gene flow from crops to wild relatives

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

*Proceedings of the Royal Society of London Series B, Biological Sciences* **270**:1879–1886 (2003)

## **Abstract**

Concern about gene flow from crops to wild relatives has become widespread with the increasing cultivation of transgenic crops. Possible consequences of such gene flow include genetic assimilation, wherein crop genes replace wild ones, and demographic swamping, wherein hybrids are less fertile than their wild parents, and wild populations shrink. Using mathematical models of a wild population recurrently receiving pollen from a genetically fixed crop, we find that the conditions for genetic assimilation are not stringent, and progress toward replacement can be fast, even for disfavored crop genes. Demographic swamping and genetic drift relax the conditions for genetic assimilation and speed progress toward replacement. Genetic assimilation can involve thresholds and hysteresis, such that a small increase in immigration can lead to fixation of a disfavored crop gene that had been maintained at a moderate frequency, even if the increase in immigration is cancelled before the gene fixes. Demographic swamping can give rise to “migrational meltdown,” such that a small increase in immigration can lead to not only fixation of a disfavored crop gene but also drastic shrinkage of the wild population. These findings suggest that the spread of crop genes in wild populations should be monitored more closely.

## Introduction

Pollen from agricultural crops often reaches wild plants growing nearby, and when the wild species are closely related to the crops, hybridization often ensues (Hancock et al., 1996; Ellstrand et al., 1999). Twelve of the world's thirteen most important crops are known to hybridize with wild relatives somewhere in their agricultural ranges (Ellstrand et al., 1999), and many other cases of crop–wild hybridization are known (e.g., Klinger et al., 1991; Love, 1994; Luby and McNichol, 1995). In most cases, hybrids are fertile to some extent (Hancock et al., 1996; Ellstrand et al., 1999), and in many cases, they are nearly as fertile as their wild parents or even more so (e.g., Langevin et al., 1990; Hauser et al., 1998a, 1998b; Snow et al., 1998). Fertile hybrids mediate gene flow, “the incorporation of genes into the gene pool of one population from one or more other populations” (Futuyma, 1998), and many studies have detected crop genes in wild populations (e.g., Xu et al., 1996; Linder et al., 1998; Bartsch et al., 1999). Gene flow from crops to wild relatives has been happening for millennia (Hancock et al., 1996; Ellstrand et al., 1999), but concern about it has become widespread with the increasing cultivation of transgenic crops (Snow and Morán Palma, 1997; Hails, 2000; Ellstrand, 2001), especially in view of reports of escaped transgenes (e.g., Hall et al., 2000; Quist and Chapela, 2001; Brasher, 2002).

Regardless of whether transgenes are involved, the consequences of crop–wild gene flow can be problematic. One possible consequence is that crop genes may replace wild ones; this is known as genetic assimilation (Levin et al., 1996; Wolf et al., 2001). It is a conservation problem, because it threatens the genetic diversity of wild populations. Another possible consequence is that if hybrids are less fertile than their wild parents, then wild populations may shrink; this is known as demographic swamping (Levin et al., 1996; Wolf et al., 2001). It is a conservation problem, because smaller populations are more vulnerable to habitat disruption, inbreeding depression, and other risks (Ellstrand and Elam, 1993; Levin et al., 1996). Demographic swamping can aggravate genetic assimilation, in that as a wild population shrinks, it is likely to produce less pollen, so any pollen it receives from a crop becomes a larger fraction of the total (Ellstrand and Elam,

1993; Levin et al., 1996). Alternatively, if hybrids are more fertile than their wild parents, then hybrids may be invasive. Invasiveness is a conservation problem when natural areas are invaded and an agricultural one when farmland is invaded. Gene flow to wild relatives is implicated in the evolution of more aggressive weeds for seven of the world's thirteen most important crops (Ellstrand et al., 1999).

These possibilities have been recognized for some time, but there have been few theoretical studies of them; we know of two. Huxel (1999) simulated models of one locus with two alleles, one allele initially fixed in a native species and the other fixed in an invading species, making various assumptions about the fitnesses and interfertilities of natives, invaders, and hybrids. He found that genetic assimilation can be fast if the invading allele is favored by selection and is possible even if the invading allele is disfavored by selection. Wolf et al. (2001) simulated ecologically detailed models of a hybridizing plant species, omitting explicit genetics but classifying each individual as native, invader, or hybrid. Like Huxel, Wolf et al. found that genetic assimilation can be fast for favored invaders and is possible even for disfavored ones. They also found that the speed of genetic assimilation is sensitive to the immigration rate, native fitness, and hybrid fitness.

Here, we study genetic assimilation and demographic swamping using mathematical models of a wild population recurrently receiving pollen from a genetically fixed crop. We are interested in the conditions for crop genes to replace wild ones, in the timing of replacement, and in the relationship between the frequencies of crop genes and the population size. More specifically, we want to understand how the conditions, timing, and relationship are likely to depend on the immigration rate, selection strengths, dominance levels, and population size. We begin with a basic model for genetic assimilation at one locus in a population of large and constant size. The most striking implication of this model is the possibility of thresholds and hysteresis, such that a small increase in immigration can lead to fixation of a disfavored crop allele that had been maintained at a moderate frequency, even if the increase in immigration is cancelled before the allele fixes. We continue with an extended model for genetic assimilation at one locus in a population of large but variable size, so disfavored crop alleles lead to demographic swamping. The most striking

implication of this model is the possibility of “migrational meltdown” (Ronce and Kirkpatrick, 2001), such that a small increase in initial immigration can lead to not only fixation of a disfavored crop allele but also drastic shrinkage of the wild population. We then briefly consider genetic drift. In smaller populations, more-strongly disfavored crop alleles fix, and fixation times tend to be shorter. Finally, we briefly consider two loci, emphasizing whether the evolution of the loci together differs qualitatively from what might be expected on the basis of considering each locus separately. When epistasis is weak and linkage is loose, two-locus outcomes usually fulfill one-locus expectations. Our models pertain to crop–wild gene flow through pollen, but many of our conclusions probably also apply to crop–wild gene flow through seed, crop–crop gene flow through pollen or seed, and invader–native gene flow in plants or animals.

## **Genetic assimilation**

### **Model**

Consider a wild species initially fixed for an allele  $A_1$  and a closely related crop fixed for an allele  $A_2$  at a diploid locus  $A$ . (If  $A_2$  represents a transgene, then  $A_1$  represents its absence.) Suppose a wild population recurrently receives crop pollen, and offspring are selected on genotype at  $A$ . We are interested in the conditions for  $A_2$  to replace  $A_1$  and in the timing of replacement.

For simplicity, assume the wild species is annual with no seed bank, so that generations are nonoverlapping. Assume the crop is permanently fixed for  $A_2$ , so that it suffices to model the wild population. This assumption is appropriate for crops raised from commercially supplied seed. (It would be a simple change to assume the crop is heterozygous at  $A$ .) Assume pollen flow is followed by random union of gametes, so that it suffices to follow allele frequencies rather than genotype frequencies. In particular, assume wild and hybrid plants flower synchronously, do not preferentially self-fertilize, and produce pollen that functions normally on each other’s stigmas. Asynchronous flowering, preferential self-fertilization, and hybrid pollen that malfunctions on wild stigmas impede genetic assimilation. For plants with these traits, our model yields, for example, a lower bound on the immigration rate needed for a given disfavored  $A_2$  to replace  $A_1$ .

For now, assume the number of plants surviving to maturity does not vary from one generation to the next (i.e., soft selection), so that it suffices to follow allele frequencies. This assumption is appropriate for some wild populations but not others (Ellstrand and Elam, 1993; Levin et al., 1996; Cummings and Alexander, 2002). For now, assume the number of plants surviving to maturity is large, so that genetic drift is negligible except for rare alleles. This assumption too is appropriate for some wild populations but not others (Ellstrand and Elam, 1993; Levin et al., 1996). For now, assume only  $A$  is under selection, so that it suffices to model one locus. As we will explain, this assumption is inappropriate for most crops and wild relatives. We will extend our model to encompass demographic swamping, genetic drift, and two loci, but understanding the basic model will form a foundation for understanding the extended ones.

Let the frequencies of  $A_2$  and  $A_1$  among gametes before pollen flow be  $q$  and  $p = 1 - q$ . Let the fraction of pollen received from the crop be  $m$ , where  $0 < m \leq 1$ , so the frequencies of  $A_2$  and  $A_1$  among pollen after pollen flow are  $(1 - m)q + m$  and  $(1 - m)p$ . Let the relative fitnesses of the genotypes be  $w_{1/1} = 1$ ,  $w_{1/2} = 1 + hs$ , and  $w_{2/2} = 1 + s$ , where  $-1 < s$  and  $0 \leq h \leq 1$ .  $A_2$  is favored if  $s > 0$ , neutral if  $s = 0$ , or disfavored if  $s < 0$ ; for example, a crop allele promoting pest resistance might be favored in the wild, one conferring herbicide resistance might be neutral, and one inhibiting seed dispersal might be disfavored (Snow and Morán Palma, 1997; Hails, 2000). According to the bounds on  $h$ , the heterozygote fitness is between the homozygote fitnesses; we will not consider overdominance or underdominance. (However, when we model two loci, we will consider the possibility that first-generation hybrids are less fit than either parent, even though neither locus is underdominant.) By standard reasoning (e.g., Roughgarden, 1979, ch. 3), the change in  $q$  from one generation to the next is

$$\Delta q = \frac{qp(q(1 - h) + ph)s + \frac{1}{2}mp(1 + (q - p)^2hs + 2qps)}{1 + 2qphs + q^2s + mp(q(1 - h) + ph)s}. \quad (1)$$

Neglecting fluctuations in  $m$ ,  $s$ , or  $h$ , starting from  $q_0 = 0$ , and applying (1)  $t$  times gives  $q_t$ , the frequency of  $A_2$  in generation  $t$ . As this article went to press, we learned that essentially this model of “continent–island” gene flow has also been considered by Nagylaki (1977, ch. 6).

## Equilibria

It can be shown that  $q$  always monotonically approaches an equilibrium  $\hat{q}$ , which depends on  $m$ ,  $s$ , and  $h$ . If  $s \geq 0$ , then immigration is abetted or at least not opposed by selection, and  $\hat{q} = 1$ . If  $s < 0$ , then immigration is opposed by selection, and  $\hat{q}$  reflects the balance between these forces. When  $m \ll h|s|$ , selection dominates, and by standard reasoning (e.g., Roughgarden, 1979, ch. 3),

$$\hat{q} \approx \frac{m}{2} \left( \frac{1}{h|s|} - 1 \right),$$

which is much less than 1. Alternatively, when  $m \gg |s|$ , immigration dominates, and  $\hat{q} = 1$ . Algebraic analysis of intermediate situations is straightforward and instructive. There is a value  $m^*$ , which depends on  $s$  and  $h$  (see appendix for formula), such that if  $m < m^*$ , then  $\hat{q} < 1$ , whereas if  $m > m^*$ , then  $\hat{q} = 1$ . Figure 1a shows an example. As might be expected (Slatkin, 1987),  $m^*$  and  $s$  are of similar magnitude. Thus an immigration rate of several percent can lead to fixation of a disfavored crop allele with a selection coefficient of several percent.

When selection against  $A_2$  is sufficiently dominant, the transition at  $m = m^*$  from  $\hat{q} < 1$  to  $\hat{q} = 1$  is steep or even discontinuous. There is a value  $h^*$ , which depends on  $s$  (see appendix for formula), such that if  $h \leq h^*$ , then the transition is continuous, whereas if  $h > h^*$ , then it is not. If  $h$  is less than but close to  $h^*$ , then the transition is continuous but steep. Figure 1b shows an example with a discontinuous transition.  $h^*$  exceeds  $2/3$ , but not necessarily by much; for example, if  $|s| < 0.35$ , then  $h^* < 0.7$ . Thus a small increase in immigration can lead to fixation of a disfavored crop allele that had been maintained at a moderate frequency.

When  $h > h^*$ , temporarily increasing  $m$  from less than to greater than  $m^*$  can permanently change the wild population. To understand this possibility, examine Figure 2, which extends Figure 1b by showing all feasible (i.e., real and between 0 and 1) equilibria of (1) for the given  $s$  and  $h$ . When  $m$  is small, there are two feasible equilibria,  $\hat{q}$ , which is attracting, and 1, which is repelling. There is a value  $m^\dagger$ , which depends on  $s$  and  $h$  (see appendix for formula), such that at  $m = m^\dagger$ , a third equilibrium appears between  $\hat{q}$  and 1,  $\check{q}$ , which is repelling, and 1 becomes attracting. As  $m$  increases further,  $\hat{q}$  and  $\check{q}$  approach each other, and at  $m = m^*$ , they collide and vanish, and 1

becomes the only feasible equilibrium. Suppose  $m$  is initially a value  $m^{(1)}$  between  $m^\dagger$  and  $m^*$ , and  $q$  is initially  $\hat{q}$  at  $m^{(1)}$ . While  $m$  is temporarily a value  $m^{(2)}$  greater than  $m^*$ ,  $q$  increases. If  $m$  remains  $m^{(2)}$  long enough that  $q$  becomes greater than  $\check{q}$  at  $m^{(1)}$ , then when  $m$  is again  $m^{(1)}$ ,  $q$  continues increasing rather than returning to  $\hat{q}$  at  $m^{(1)}$ . For example, if  $s = -0.05$  and  $h = 0.9$  (as in Figures 1b and 2), and if  $m^{(1)} = 0.025$  and  $m^{(2)} = 0.05$ , then 42 generations is long enough. Thus a temporary increase in immigration can lead to fixation of a disfavored crop allele, even if the increase in immigration is cancelled before the allele fixes.

## Timing

When  $A_2$  is destined to fix, how long does fixation take? Our model does not yield a meaningful answer to this question. When  $\hat{q} = 1$ ,  $q$  approaches but does not reach 1 in finite time. The problem is that when  $q$  is close to 1, its dynamics in a real population would be dominated by genetic drift, which (1) does not model. We will consider genetic drift shortly. However, the earlier generations of evolution under (1) are instructive.

Figure 3 shows the first 50 generations of evolution under (1) for a given immigration rate and three crop alleles: one is favored, another is neutral, and the third is disfavored.  $\hat{q} = 1$  for all three crop alleles. Increasing or decreasing immigration would move all three curves toward the upper left or lower right. Changing dominance would change the curves little. The curves are concave, because immigration becomes less important as a crop allele becomes more common, so progress toward fixation slows as it proceeds. However, the crop alleles quickly become common; for example, by generation 16, all three are more common than the wild allele.

## Demographic swamping

### Model

Suppose the number of plants surviving to maturity varies from one generation to the next, and as the population grows or shrinks, it produces more or less pollen, so the pollen it receives from the crop becomes a smaller or larger fraction of the total. Assuming the population is initially

at its carrying capacity, we are interested in the conditions for  $A_2$  to replace  $A_1$ , in the timing of replacement, and in the relationship between the frequency of  $A_2$  and the population size. Our model accommodates favored, neutral, or disfavored crop alleles, but our analysis emphasizes disfavored ones, which lead to demographic swamping.

Let the number of plants surviving to maturity be  $N$ . Let the absolute fitnesses of the genotypes be  $W_{1/1} = a/(1+bN)$ ,  $W_{1/2} = a(1+hs)/(1+bN)$ , and  $W_{2/2} = a(1+s)/(1+bN)$ , where  $1 < a$  and  $0 < b$ . Dividing by  $W_{1/1}$  gives the same relative fitnesses as before. By standard reasoning (e.g., Roughgarden, 1979, ch. 3), the change in  $N$  from one generation to the next is

$$\Delta N = \left( \frac{a\bar{w}}{1+bN} - 1 \right) N, \quad (2)$$

where

$$\bar{w} = 1 + 2qphs + q^2s + mp(q(1-h) + ph)s. \quad (3)$$

For a given  $q$ , (2) is the Beverton–Holt model, a reasonable model of population dynamics in discrete time (Yodzis, 1989, sec. 3.1). The carrying capacity is  $(a\bar{w} - 1)/b$  or 0, whichever is greater. It is convenient to measure the population size in units of the initial carrying capacity  $N_0 = (a - 1)/b$ , that is, to change variables from  $N$  to  $n = N/N_0 = bN/(a - 1)$ . By routine algebra,

$$\Delta n = \left( \frac{a\bar{w}}{1 + (a - 1)n} - 1 \right) n. \quad (4)$$

Let the amounts of pollen produced in the wild population and received from the crop be  $P^{(1)}$  and  $P^{(2)}$ . Thus  $m = P^{(2)}/(P^{(1)} + P^{(2)})$ . Suppose  $P^{(1)} = cN$ , where  $0 < c$ ; simple proportionality is a reasonable approximation. Thus  $m = P^{(2)}/(cN + P^{(2)})$ . In particular,  $m$  is initially  $m_0 = P^{(2)}/(cN_0 + P^{(2)})$ . Solving for  $c$  in terms of  $m_0$ ,  $N_0$ , and  $P^{(2)}$  yields

$$m = \frac{1}{1 + \left( \frac{1}{m_0} - 1 \right) n}. \quad (5)$$

Our model is (1) and (4), linked by (5) and (3), with parameters  $m_0$ ,  $s$ ,  $h$ , and  $a$  and initial values  $q_0 = 0$  and  $n_0 = 1$ .

## Equilibria

Simulations indicate that  $q$  and  $n$  always monotonically approach equilibria  $\hat{q}$  and  $\hat{n}$ , which depend on  $m_0$ ,  $s$ ,  $h$ , and  $a$ . If  $s \geq 0$ , then  $\hat{q} = 1$  and  $\hat{n} = (a(1+s) - 1)/(a - 1)$ , that is,  $A_2$  replaces  $A_1$ , and the population grows or at least does not shrink. If  $s < 0$ , then  $\hat{q} \leq 1$  and  $\hat{n} < 1$ , that is,  $A_2$  may or may not replace  $A_1$ , and the population shrinks. Algebraic analysis is no longer straightforward, but simulations indicate that in continuity with our basic model, there is a value  $m_0^*$ , which depends on  $s$ ,  $h$ , and  $a$ , such that if  $m_0 < m_0^*$ , then  $\hat{q} < 1$ , whereas if  $m_0 > m_0^*$ , then  $\hat{q} = 1$ . Likewise, there is a value  $h^*$ , which depends on  $s$  and  $a$ , such that if  $h \leq h^*$ , then the transition at  $m_0 = m_0^*$  from  $\hat{q} < 1$  to  $\hat{q} = 1$  is continuous, whereas if  $h > h^*$ , then it is not. When  $h > h^*$ , there is a value  $m_0^\dagger$ , which depends on  $s$ ,  $h$ , and  $a$ , such that when  $m_0^\dagger < m_0 < m_0^*$ , there is a repelling equilibrium  $\check{q}$  between the attracting equilibria  $\hat{q}$  and 1. Thus genetic assimilation is qualitatively unchanged by demographic swamping.

However,  $m_0^*$ ,  $h^*$ , and  $m_0^\dagger$  are less than  $m^*$ ,  $h^*$ , and  $m^\dagger$  in our basic model, and the transition at  $m_0^*$  is steeper than the one at  $m^*$ . Figures 4a and b show examples parallel to Figures 1a and b. The reason for these effects is that  $q$  experiences positive feedback through  $\bar{w}$ ,  $n$ , and  $m$ : if  $q$  increases, then  $\bar{w}$  decreases, hence  $n$  decreases, and hence  $m$  increases, which tends to increase  $q$  further. Thus as expected (Ellstrand and Elam, 1993; Levin et al., 1996), genetic assimilation is quantitatively aggravated by demographic swamping.

These effects grow as  $a$  decreases and become large as  $a$  approaches  $1/(1 - |s|)$ . When  $a \leq 1/(1 - |s|)$ , the wild population does not persist if  $A_2$  fixes (i.e., for  $q = 1$ , the carrying capacity is 0). (If  $|s|$  is large, then  $m_0^*$  exceeds 1, and the effects of decreasing  $a$  through  $1/(1 - |s|)$  are small. For simplicity, assume  $|s|$  is small.) When  $a$  is close to or less than  $1/(1 - |s|)$ , a steep transition to  $\hat{q} = 1$  is accompanied by a steep transition to  $\hat{n} \ll 1$ . Figure 5 shows such an extreme case. Ronce and Kirkpatrick (2001) dubbed such a transition a “migrational meltdown.” Note that selection against  $A_2$  need not be dominant. Thus a small increase in initial immigration can lead to not only fixation of a disfavored crop allele but also drastic shrinkage of the wild population.

## Timing

When  $s < 0$ ,  $m$  increases as  $q$  increases, so  $q$  reaches any given value sooner. Like the effects on equilibria, this effect on timing is small unless  $a$  is close to or less than  $1/(1 - |s|)$ . For example, if  $m_0 = 0.1$ ,  $s = -0.05$ , and  $h = 0.5$ , then without demographic swamping,  $q$  reaches 0.9 in 68 generations, whereas with demographic swamping and  $a = 1.06$  (as in Figure 5),  $q$  reaches 0.9 in 43 generations.

## Genetic drift

We now briefly consider genetic drift, for two reasons. First, many plant species at risk of extinction occur in small populations (Ellstrand and Elam, 1993; Levin et al., 1996), often fewer than a hundred individuals, and genetic drift can be strong in small populations. Second, even in large populations, genetic drift can be strong for rare alleles, for example,  $A_1$  when  $A_2$  is nearly fixed. We are interested in how the population size affects which crop alleles fix and how long fixation takes.

For simplicity, assume  $N$  is constant, so that we are again considering genetic assimilation without demographic swamping, and assume the wild species is monoecious. From one generation, construct the next as follows. For each offspring, pick a mother at random with fitness-proportionate weighting, that is, if a plant has genotype  $A_i A_j$ , then the probability of picking it is  $w_{i/j}/(\bar{w}N)$ . Pick one of its two alleles at random, which it contributes to the offspring. With probability  $1 - m$ , pick a father and one of its two alleles in the same way. Otherwise, the father is a crop individual, and it contributes  $A_2$ . This model captures the essentials of immigration, selection, and genetic drift.

Assuming there is pollen flow ( $m > 0$ ), and hybrids are not sterile ( $s > -1$ ),  $A_2$  must fix eventually, even if  $s < 0$ ,  $m < m^*$ , and hence  $\hat{q} < 1$  in our basic model. If  $\hat{q} < 1$  but  $N$  is small, then short fixation times can be appreciably frequent. For example, if  $m = 0.1$ ,  $s = -0.12$ , and  $h = 0.5$ , then  $\hat{q} \approx 0.783$ , well below 1. However, in 429 of 10000 replicate simulations for  $N = 50$ ,  $A_2$  fixed within 50 generations.

To understand how the population size affects fixation times, consider again the three crop alleles featured in Figure 3. Table 1 presents statistics on fixation times in replicate simulations for a given immigration rate and three population sizes. Fixation times tend to be shortest in the smallest population, and the effect of the population size is largest for the disfavored crop allele. In the smallest population, the frequency of fixation within 50 generations is appreciable for all three crop alleles.

## Two loci

A crop and wild relative are likely to differ at several loci, if only due to artificial selection on several traits in the crop. In a wild population receiving crop pollen, these loci are initially in complete linkage disequilibrium, and as long as they remain polymorphic, they also remain in linkage disequilibrium, reflecting the balance among immigration, recombination, and possibly selection. If both of two loci affect a trait, then selection on the trait can affect the linkage disequilibrium between the loci. If only one of two loci affects a trait, but the loci are in linkage disequilibrium, then selection on the trait can affect the evolution of both loci. These interactions can be very complicated, and we will not consider all the plausible possibilities. However, we now briefly consider two loci affecting fitness in a way that may be common. We are interested in whether the evolution of the loci together differs qualitatively from what might be expected on the basis of considering each locus separately.

For simplicity, assume  $N$  is constant and large, so that we are again considering genetic assimilation without demographic swamping or genetic drift. In addition to  $A$ , consider locus  $B$  with wild and crop alleles  $B_1$  and  $B_2$ . Let the frequency of recombination between  $A$  and  $B$  be  $r$ , where  $0 < r \leq 0.5$ . The simplest possibility is that  $A$  and  $B$  affect fitness multiplicatively, as specified in Table 2. This lack of epistasis is likely when the two loci affect different life stages; for example,  $A$  might affect seedling recruitment, and  $B$  might affect seed production.  $A_2$  is favored if  $s_A > 0$  or disfavored if  $s_A < 0$ , regardless of  $B$ , and likewise for  $B_2$ . When  $A_2$  is disfavored, call it weakly disfavored if it would fix in the absence of  $B_2$  or strongly disfavored if not, and likewise

for  $B_2$ . Even though neither locus is underdominant, first-generation hybrids can be less fit than either parent, for example, if  $s_A = 0.2$ ,  $s_B = -0.1$ ,  $h_A = 0$ , and  $h_B = 1$ . Standard reasoning (e.g., Roughgarden, 1979, ch. 8) leads to evolution equations. These equations are algebraically complicated, so we do not present them here, and we have studied them mainly using simulations.

Simulations indicate that the frequencies  $q_A$  and  $q_B$  of  $A_2$  and  $B_2$  always approach equilibria  $\hat{q}_A$  and  $\hat{q}_B$ , although the approach is sometimes nonmonotonic. It might be expected that  $\hat{q}_A = 1$  when  $A_2$  is favored or weakly disfavored and  $\hat{q}_A < 1$  when  $A_2$  is strongly disfavored, and likewise for  $\hat{q}_B$ . According to Table 3, which presents statistics on simulations with randomly generated parameter values, these expectations are usually fulfilled. For example, when  $A_2$  is favored, it fixes, regardless of  $B$ , albeit fixation takes longer when  $B_2$  is disfavored. However, one-locus expectations are sometimes violated. The most intriguing violations are unexpected fixations, for example, when  $A_2$  is favored, and  $B_2$  is strongly disfavored, but  $\hat{q}_B = 1$ . This happens when selection against  $B_2$  is dominant, so there is a repelling equilibrium  $\check{q}_B$  (as in Figures 1b and 2), and linkage is tight, so  $B_2$  “hitchhikes” with  $A_2$  long enough that  $q_B$  becomes greater than  $\check{q}_B$ . This is rare, because it requires a delicate balance among immigration, recombination, and selection. The most common violations of one-locus expectations are unexpected polymorphisms, for example, when  $B_2$  is strongly disfavored, and  $A_2$  is weakly disfavored, but  $\hat{q}_A < 1$ . This happens more often when linkage is tighter, so  $B_2$  “drags down”  $A_2$  more effectively. When  $r$  is fixed at 0.5 rather than uniformly distributed between 0 and 0.5, the frequency of such violations is 22% rather than 49%.

These results would change little given weak epistasis, for example,  $w_{1,2/1,2} = (1 + h_A s_A)(1 + h_B s_B)(1 + e_{1,2/1,2})$ , where  $|e_{1,2/1,2}| \ll h_A |s_A|, h_B |s_B|$ , and  $r$ . They would change considerably given some patterns of strong epistasis.

## Discussion

Our models have at least three important general implications. First, the conditions for genetic assimilation are not stringent. With one locus, a favored or neutral crop allele always fixes, and even a disfavored one fixes when the immigration rate exceeds the selection coefficient. With

demographic swamping or genetic drift, more-strongly disfavored crop alleles fix, and fixation times tend to be shorter. With two loci, the crop allele at one locus usually fixes if it would fix in the absence of the crop allele at the other locus. These findings suggest that some amount of genetic assimilation should be expected whenever there is recurrent gene flow from a crop to a wild relative for more than a few generations. Moreover, crop alleles can quickly become common.

Second, when selection against a crop allele is dominant or the intrinsic growth rate of a wild population is small, genetic assimilation can involve thresholds and hysteresis. These findings suggest that when a crop gene or trait is observed at a moderate frequency in a wild population, it should not be assumed that a small increase in immigration, for example, through an expansion of farmland nearby, would cause a small increase in the frequency of the gene or trait. Moreover, when a small increase in immigration causes a large increase in the frequency of a crop gene or trait, it should not be assumed that restoring immigration to the original rate would restore the gene or trait to the original frequency; a greater decrease in immigration might be needed.

Third, when the intrinsic growth rate of a wild population is small, abrupt genetic assimilation can be accompanied by severe demographic swamping. This finding suggests that wild populations subject to weak density dependence are especially vulnerable to recurrent gene flow from crops. We observed “migrational meltdown” in a particular model, with Beverton–Holt-like population dynamics, but this phenomenon transcends the details of our model; Ronce and Kirkpatrick (2001) observed it in a rather different model, with logistic-like population dynamics.

In examples, we have supposed  $m \leq 0.1$ , which is consistent with measured hybridization rates for crops and wild relatives growing tens to hundreds of meters apart (Hancock et al., 1996; Ellstrand et al., 1999); for example, Arriola and Ellstrand (1996) measured hybridization rates of several percent for the predominantly self-fertilizing crop sorghum and weed johnsongrass growing tens of meters apart. However, immigration can be much higher; for example, Whitton et al. (1997) measured hybridization rates up to 42% for wild sunflowers growing near cultivated ones. If  $m$  is, say, 0.25, then in our basic model, the frequencies of crop alleles with  $h = 0.5$  and  $s = 0.05$ , 0,

and  $-0.05$  (as in Figure 3) reach 0.9 in 16, 18, and 20 generations. It is not surprising that after up to 40 years of gene flow from cultivated sunflowers, Linder et al. (1998) observed high frequencies of crop-specific markers in wild sunflower populations.

It is more difficult to determine what values of  $s$  and  $h$  are appropriate. There have been few measurements of crop–wild hybrid fitness under wild conditions. Moreover, when a crop and wild relative differ at several loci, the fitness effects of individual crop alleles are obscure. In most examples, we have supposed  $|s| \leq 0.05$ , which encompasses strong selection by the usual standards of evolutionary genetics. However, this supposition is consistent with some measurements of crop–wild hybrid fitness (e.g., Klinger and Ellstrand, 1994; Cummings et al., 1999; Spencer and Snow, 2001). Moreover, some crop alleles have dramatic effects; for example, most of the spectacular differentiation between maize and teosinte is associated with five loci (Lynch and Walsh, 1998, ch. 15). The fitness effects of transgenes might be conspicuous, because plants differing only in presence versus absence of transgenes could be compared, and the fitness effects might be large; for example, transgenes promoting pest resistance, such as *Bacillus thuringiensis* toxin genes (Snow and Morán Palma, 1997; Hails, 2000), might be strongly favored in the wild.

We have assumed pollen flow is consistent, that is,  $m$  does not vary from one generation to the next unless  $N$  varies. Changing our models to accommodate crop rotation and other sources of periodic variation in pollen flow would be simple, but merely averaging  $m$  over such variation is a good approximation; for example, if the crop is cultivated in alternate years, then replace  $m$  by  $m/2$ . We have also assumed the crop is permanently fixed for its alleles. This assumption is appropriate for industrial agriculture, where crops are raised from commercially supplied seed, so gene flow from wild relatives is negligible. Crop genomes change as new cultivars are marketed, but the changes are typically incremental, so a given crop allele is present in a succession of cultivars. The assumption that the crop is permanently fixed for its alleles is inappropriate for traditional agriculture, where crops are raised each season from seeds harvested the previous one, so gene flow from wild relatives can be appreciable (Jarvis and Hodgkin, 1999).

We have treated gene flow from a crop to a wild relative, but much of our treatment probably also applies to gene flow from a commercial cultivar to a landrace (i.e., traditional cultivar); for example, there are reports of gene flow from transgenic maize to maize landraces in Mexico (Quist and Chapela, 2001). Little is known about the prevalence of such gene flow or its relationship to traditional cultivation practices. Our findings affirm the importance of learning more, especially in view of the potential value of landraces as sources of genes for improving commercial cultivars.

Indeed, although wild relatives and landraces of crops form a small fraction of plant species and varieties, they are peculiarly important. They are the most likely sources of genes for improving crops, and wild relatives are the most likely sources of more aggressive weeds. If only to protect the potentially valuable genetic diversity, our findings affirm the need for empirical studies of crop–wild gene flow, from measuring the relevant traits of crops and wild relatives to developing strategies for monitoring the spread of crop genes in wild populations.

## Appendix

In our basic model, for  $s < 0$ ,

$$\hat{q} = \begin{cases} \frac{h|s| - m(2h - 1)|s| - \sqrt{h^2s^2 - 2m(1 - m)(2h - 1)|s| - m^2(2h - 1)s^2}}{2(1 - m)(2h - 1)|s|} & \text{if } m \leq m^*, \\ 1 & \text{if } m > m^*, \end{cases}$$

$$m^* = \begin{cases} \frac{2(1 - h)|s|}{1 - h|s|} & \text{if } h \leq h^*, \\ \frac{2h - 1 - \sqrt{(2h - 1)(2h - 1 - h^2|s|(2 - |s|))}}{(2h - 1)(2 - |s|)} & \text{if } h > h^*, \end{cases}$$

and

$$h^* = 2 - \frac{3}{2|s|} + \sqrt{2 - \frac{4}{|s|} + \frac{9}{4s^2}}.$$

(For  $h = 1/2$ ,

$$\hat{q} = \begin{cases} \frac{m}{2} \left( \frac{2}{|s|} - 1 \right), & \text{if } m \leq m^*, \\ 1 & \text{if } m \geq m^*. \end{cases}$$

Note that  $h^* > 2/3$ . For  $h > h^*$ ,

$$\check{q} = \frac{h|s| - m(2h - 1)|s| + \sqrt{h^2s^2 - 2m(1 - m)(2h - 1)|s| - m^2(2h - 1)s^2}}{2(1 - m)(2h - 1)|s|}$$

and

$$m^\dagger = \frac{2(1 - h)|s|}{1 - h|s|}.$$

Note that  $\check{q} \geq 1/2$ .

## Acknowledgments

We thank the referees for their suggestions and M. Turelli for pointing out the work of T. Nagylaki.

Our research was supported by a USDA–IFAFS grant.

## References

- Arriola, P. E., and Ellstrand, N. C., 1996. Crop-to-weed gene flow in the genus *Sorghum* (Poaceae): Spontaneous interspecific hybridization between johnsongrass, *Sorghum halepense*, and crop sorghum, *S. bicolor*. *American Journal of Botany* **83**:1153–1160.
- Bartsch, D., Lehnen, M., Clegg, J., Pohl-Orf, M., Schuphan, I., and Ellstrand, N. C., 1999. Impact of gene flow from cultivated beet on genetic diversity of wild sea beet populations. *Molecular Ecology* **8**:1733–1741.
- Brasher, P., 2002. Biotech firm under fire has link to Iowa. *The Des Moines Register*. November 14, 2002. <http://DesMoinesRegister.com/business/stories/c4789013/19735220.html>.
- Cummings, C. L., and Alexander, H. M., 2002. Population ecology of wild sunflowers: Effects of seed density and post-dispersal vertebrate seed predators. *Oecologia* **130**:274–280.
- Cummings, C. L., Alexander, H. M., and Snow, A. A., 1999. Increased pre-dispersal seed predation in sunflower crop–wild hybrids. *Oecologia* **121**:330–338.
- Ellstrand, N. C., 2001. When transgenes wander, should we worry? *Plant Physiology* **125**:1543–1545.

- Ellstrand, N. C., and Elam, D. R., 1993. Population genetic consequences of small population size: Implications for plant conservation. *Annual Review of Ecology and Systematics* **24**:217–242.
- Ellstrand, N. C., Prentice, H. C., and Hancock, J. F., 1999. Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology and Systematics* **30**:539–563.
- Futuyma, D., 1998. *Evolutionary Biology*, 3rd ed. Sinauer, Sunderland, MA.
- Hails, R. S., 2000. Genetically modified plants: The debate continues. *Trends in Ecology and Evolution* **15**:14–18.
- Hall, L., Topinka, K., Huffman, J., Davis, L., and Good, A., 2000. Pollen flow between herbicide-resistant *Brassica napus* is the cause of multiple-resistant *B. napus* volunteers. *Weed Science* **48**:688–694.
- Hancock, J. F., Grumet, R., and Hokanson, S. C., 1996. The opportunity for escape of engineered genes from transgenic crops. *Hortscience* **31**:1080–1085.
- Hauser, T. P., Jørgensen, R. B., and Østergård, H., 1998a. Fitness of backcross and F2 hybrids between weedy *Brassica rapa* and oilseed rape (*B. napus*). *Heredity* **81**:436–443.
- Hauser, T. P., Shaw, R. G., and Østergård, H., 1998b. Fitness of F1 hybrids between weedy *Brassica rapa* and oilseed rape (*B. napus*). *Heredity* **81**:429–435.
- Huxel, G. R., 1999. Rapid displacement of native species by invasive species: Effects of hybridization. *Biological Conservation* **89**:143–152.
- Jarvis, D. I., and Hodgkin, T., 1999. Wild relatives and crop cultivars: Detecting natural introgression and farmer selection of new genetic combinations in agroecosystems. *Molecular Ecology* **8**:S159–S173.
- Klinger, T., Elam, D. R., and Ellstrand, N. C., 1991. Radish as a model system for the study of engineered gene escape rates via crop–weed mating. *Conservation Biology* **5**:531–535.

- Klinger, T., and Ellstrand, N. C., 1994. Engineered genes in wild populations: Fitness of weed-crop hybrids of radish, *Raphanus sativus* L. *Ecological Applications* **4**:117–120.
- Langevin, S. A., Clay, K., and Grace, J. B., 1990. The incidence and effects of hybridization between cultivated rice and its related weed red rice (*Oryza sativa* L.). *Evolution* **44**:1000–1008.
- Levin, D. A., Francisco-Ortega, J., and Jansen, R. K., 1996. Hybridization and the extinction of rare plant species. *Conservation Biology* **10**:10–16.
- Linder, C. R., Taha, I., Seiler, G. J., Snow, A. A., and Rieseberg, L. H., 1998. Long-term introgression of crop genes into wild sunflower populations. *Theoretical and Applied Genetics* **96**:339–347.
- Love, S. L., 1994. Ecological risk of growing transgenic potatoes in the United States and Canada. *American Potato Journal* **71**:647–658.
- Luby, J. J., and McNichol, R. J., 1995. Gene flow from cultivated to wild raspberries in Scotland: Developing a basis for risk assessment for testing and deployment of transgenic cultivars. *Theoretical and Applied Genetics* **90**:1133–1137.
- Lynch, M., and Walsh, B., 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland, MA.
- Nagylaki, T., 1977. *Selection in One- and Two-Locus Systems*. Springer-Verlag, Berlin.
- Quist, D., and Chapela, I. H., 2001. Transgenic DNA introgressed into traditional maize landraces in Oaxaca, Mexico. *Nature* **414**:541–543.
- Ronce, O., and Kirkpatrick, M., 2001. When sources become sinks: Migrational meltdown in heterogeneous habitats. *Evolution* **55**:1520–1531.
- Roughgarden, J., 1979. *Theory of Population Genetics and Evolutionary Ecology: An Introduction*. Prentice-Hall, Upper Saddle River, NJ.

- Slatkin, M., 1987. Gene flow and the geographic structure of natural populations. *Science* **236**:787–792.
- 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.
- Spencer, L. J., and Snow, A. A., 2001. Fecundity of transgenic wild–crop hybrids of *Cucurbita pepo* (Cucurbitaceae): Implications for crop-to-wild gene flow. *Heredity* **86**:694–702.
- Whitton, J., Wolf, D. E., Arias, D. M., Snow, A. A., and Rieseberg, L. H., 1997. The persistence of cultivar alleles in wild populations of sunflowers five generations after hybridization. *Theoretical and Applied Genetics* **95**:33–40.
- Wolf, D. E., Takebayashi, N., and Rieseberg, L. H., 2001. Predicting the risk of extinction through hybridization. *Conservation Biology* **15**:1039–1053.
- Xu, J., Kerrigan, R. W., Callac, P., Horgen, P. A., and Anderson, J. B., 1996. Genetic structure of natural populations of *Agaricus bisporus*, the commercial mushroom. *Journal of Heredity* **88**:482–488.
- Yodzis, P., 1989. *Introduction to Theoretical Ecology*. Harper and Row, New York.

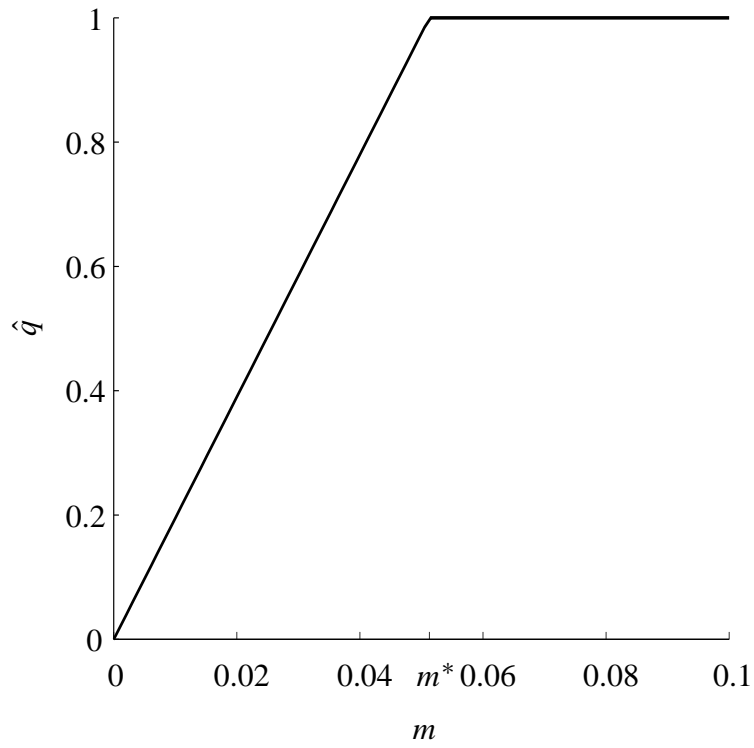


Figure 1a: An example of how  $\hat{q}$  increases with  $m$ .  $s = -0.05$  and  $h = 0.5$ .  $m^* \approx 0.0513$ .

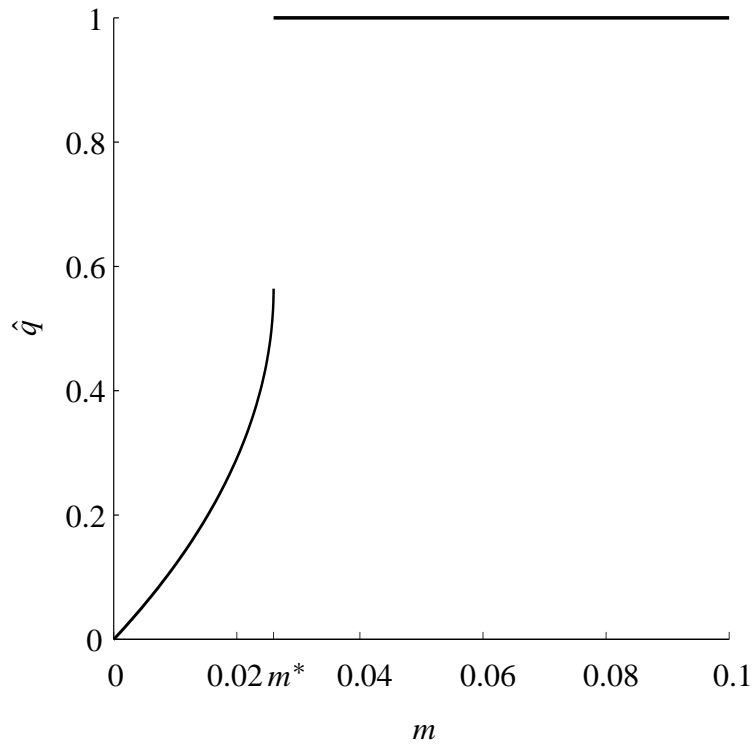


Figure 1b: An example with a discontinuous transition at  $m = m^*$  from  $\hat{q} < 1$  to  $\hat{q} = 1$ .  $s = -0.05$  and  $h = 0.9$ .  $m^* \approx 0.0260$ .

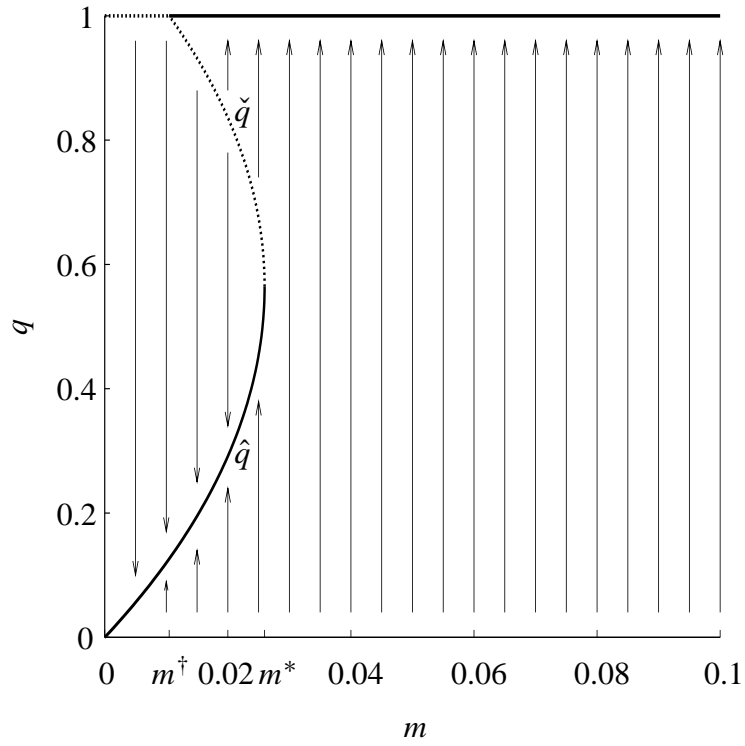


Figure 2: An extension of Figure 1b showing all feasible equilibria of (1) for  $s = -0.05$  and  $h = 0.9$ . Attracting or repelling equilibria are plotted as solid or dotted curves, and arrows indicate how  $q$  evolves for various  $m$ .  $m^\dagger \approx 0.0105$ .

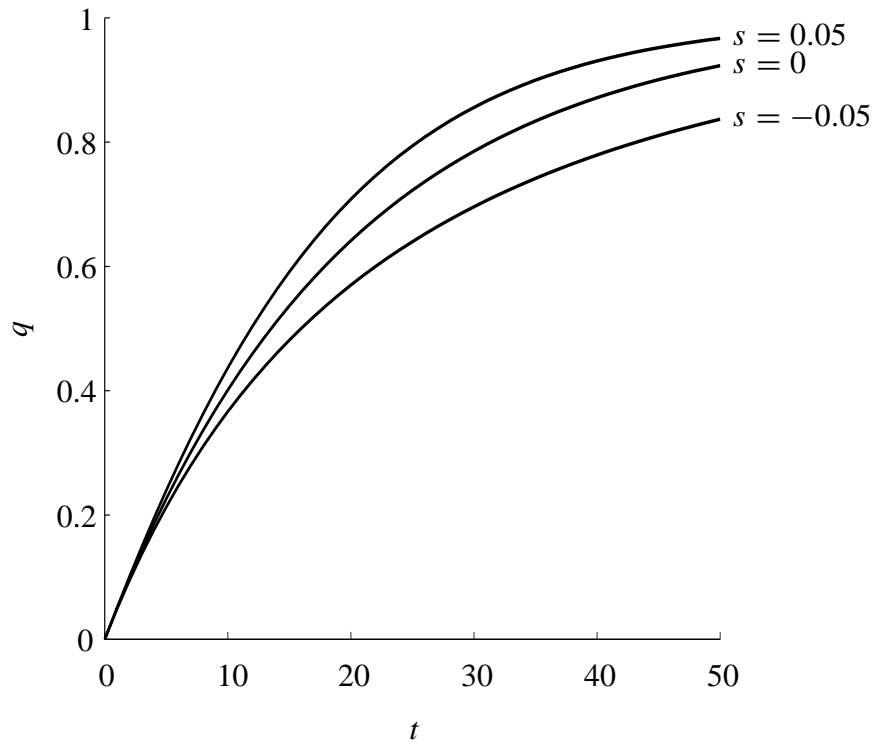


Figure 3: The first 50 generations of evolution under (1) for  $m = 0.1$  and three  $A_2$ 's:  $s = 0.05$  and  $h = 0.5$  (top curve),  $s = 0$  (middle curve), and  $s = -0.05$  and  $h = 0.5$  (bottom curve). (The dynamics are discrete-time, but continuous curves are drawn for ease of reading.)

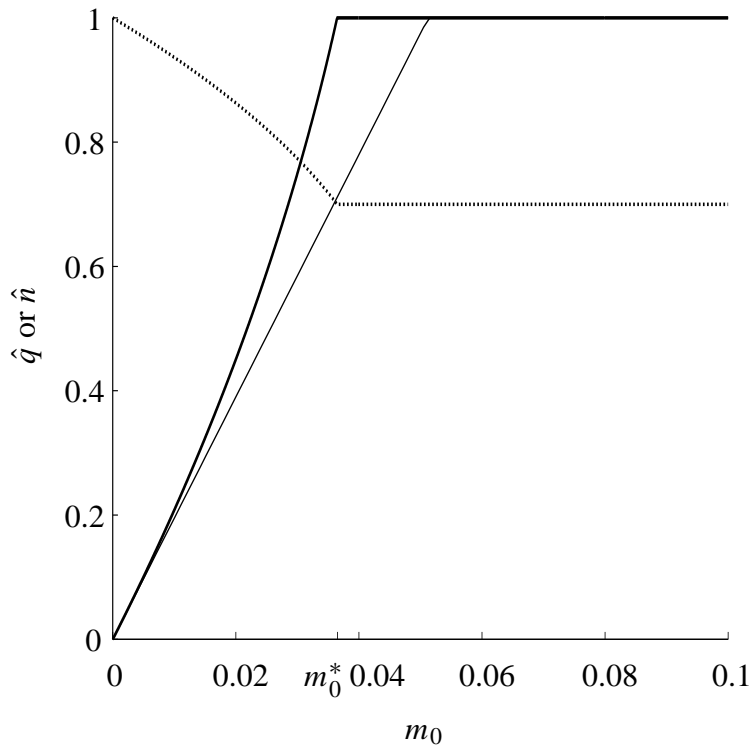


Figure 4a: Like Figure 1a but with demographic swamping.  $s = -0.05$ ,  $h = 0.5$ , and  $a = 1.2$ .  $\hat{q}$  is plotted as a heavy solid curve, and  $\hat{n}$  is plotted as a heavy dotted curve. For comparison,  $\hat{q}$  without demographic swamping is plotted as a light solid curve.  $m_0^* \approx 0.0365$ .

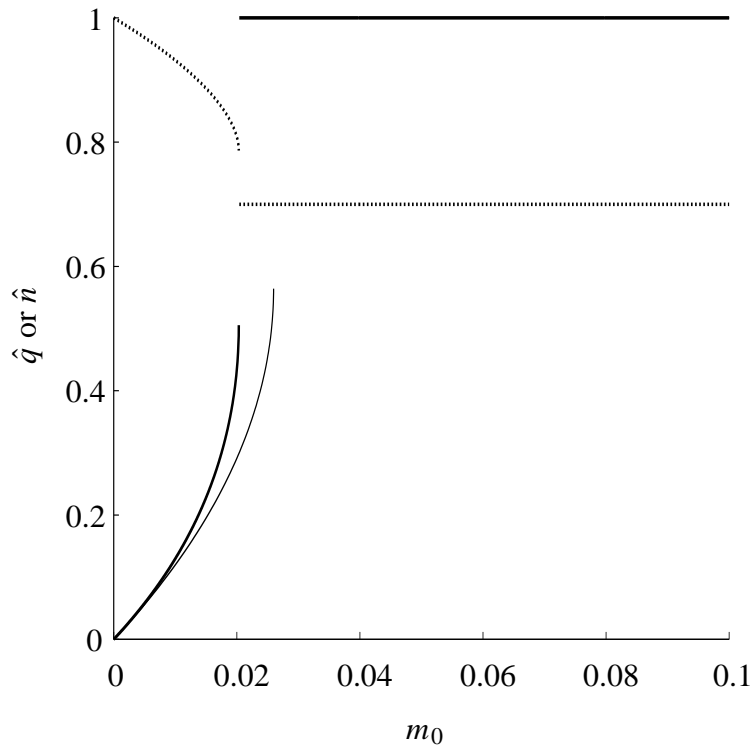


Figure 4b: Like Figure 1b but with demographic swamping.  $s = -0.05$ ,  $h = 0.9$ , and  $a = 1.2$ .  
 $m_0^* \approx 0.0204$ .

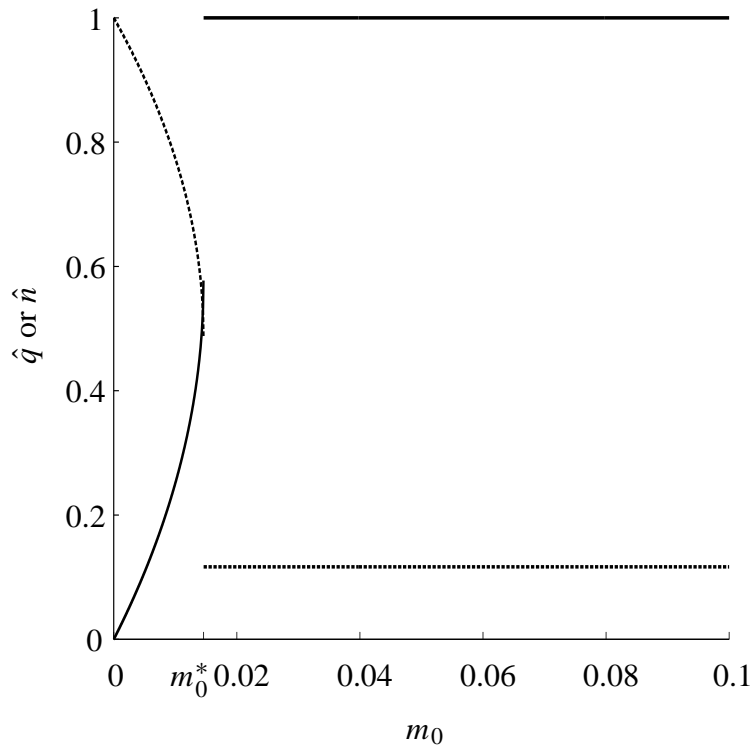


Figure 5: An extreme case of demographic swamping.  $\hat{q}$  is plotted as a solid curve, and  $\hat{n}$  is plotted as a dotted curve.  $s = -0.05$ ,  $h = 0.5$ , and  $a = 1.06$ .  $m_0^* \approx 0.0146$ .

$s$	$N$	$\mu$	$\nu$	$\sigma$
	50	50	47	17
0.05	500	78	75	17
	5000	109	106	17
	50	60	56	23
0	500	101	97	25
	5000	145	141	25
	50	79	71	35
-0.05	500	153	143	47
	5000	239	230	49

Table 1: Statistics on fixation times in replicate simulations for  $m = 0.1$ ,  $N = 50, 500$ , or  $5000$ , and three  $A_2$ 's:  $s = 0.05$  and  $h = 0.5$  (top three rows),  $s = 0$  (middle three rows), and  $s = -0.05$  and  $h = 0.5$  (bottom three rows). Each row presents statistics on 10000 simulations.  $\mu$ ,  $\nu$ , and  $\sigma$  are the mean, median, and standard deviation of the fixation times.

	$B_1B_1$	$B_1B_2$	$B_2B_2$
$A_1A_1$	1	$1 + h_B s_B$	$1 + s_B$
$A_1A_2$	$1 + h_A s_A$	$(1 + h_A s_A)(1 + h_B s_B)$	$(1 + h_A s_A)(1 + s_B)$
$A_2A_2$	$1 + s_A$	$(1 + s_A)(1 + h_B s_B)$	$(1 + s_A)(1 + s_B)$

Table 2: Relative fitnesses when  $A$  and  $B$  affect fitness multiplicatively.  $-1 < s_A$ ,  $-1 < s_B$ ,  $0 \leq h_A \leq 1$ , and  $0 \leq h_B \leq 1$ .

$A_2$	$B_2$	one-locus expectations	frequency of violations	nature of violations
avored	avored	$\hat{q}_A = \hat{q}_B = 1$	0	—
avored	weakly disfavored	$\hat{q}_A = \hat{q}_B = 1$	0	—
avored	strongly disfavored	$\hat{q}_A = 1, \hat{q}_B < 1$	0.09%	$\hat{q}_B = 1$
weakly disfavored	weakly disfavored	$\hat{q}_A = \hat{q}_B = 1$	6.3%	$\hat{q}_A < 1, \hat{q}_B < 1$
weakly disfavored	strongly disfavored	$\hat{q}_A = 1, \hat{q}_B < 1$	49%	$\hat{q}_A < 1$
strongly disfavored	strongly disfavored	$\hat{q}_A < 1, \hat{q}_B < 1$	0	—

Table 3: Statistics on simulations with randomly generated parameter values. In each simulation,  $m$ ,  $h_A$ , and  $h_B$  were uniform deviates between 0 and 1, and  $r$  was a uniform deviate between 0 and 0.5. For favored  $A_2$ 's,  $s_A$  was a uniform deviate between 0 and 1. For weakly disfavored  $A_2$ 's,  $s_A$  was a uniform deviate between  $-1$  and 0, and if the resulting  $A_2$  was strongly rather than weakly disfavored, then the sampled  $s_A$  was discarded, and another one was sampled. Likewise for strongly disfavored  $A_2$ 's. Likewise for  $B_2$ 's. Each row presents statistics on 10000 simulations.