Ecological Applications, 12(6), 2002, pp. 1661–1671

FECUNDITY SELECTION IN A SUNFLOWER CROP–WILD STUDY: CAN ECOLOGICAL DATA PREDICT CROP ALLELE CHANGES?

CHARITY L. CUMMINGS,[1] HELEN M. ALEXANDER,[1,4] ALLISON A. SNOW,[2] LOREN H. RIESEBERG,[3] MIN JU KIM,[3] AND THERESA M. CULLEY[2]

[1]*Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas 66045 USA*
[2]*Department of Evolution, Ecology, and Organismal Biology, Ohio State University, Columbus, Ohio 43210 USA*
[3]*Department of Biology, University of Indiana, Bloomington, Indiana 47405 USA*

Abstract. Genes that spread from transgenic crops to populations of weedy relatives can be a cause of concern if fitness-related, transgenic traits persist and enhance weed invasiveness. Studies of the prevalence of crop-specific genetic markers in wild populations can provide data on such introgression. We conducted a field experiment in eastern Kansas to measure changes in frequencies of crop-specific genetic markers in wild sunflower (*Helianthus annuus*). Three allozyme markers were monitored in three artificial populations that each initially consisted of 100 wild and 100 F_1 hybrid plants. Survival, flowering time, and average seed production per plant were quantified during the first year of the study (1997). Hybrid plants produced only 1–2% as many seeds per plant as wild plants but did not differ in survivorship. Simple selection models incorporating fecundity differences between wild and F_1 hybrids accurately predicted crop allele frequencies in the 1998 seedlings. We predicted that frequencies of crop alleles in 1998 would average ~0.03 for the three populations. Crop-specific allele frequencies dropped from the initial level of 0.25 to a mean of 0.03 in the 1998 seedlings and averaged 0.05 in the next generation of seeds. Accounting for differences in flowering phenology and predispersal seed predation did not improve the accuracy of our predictions for 1998 seedlings. Our results suggest that ecological data can be useful for estimating the frequencies of crop genes following episodes of crop–wild hybridization in sunflowers. This approach can be applied to other study systems in which data on survivorship and fecundity are used to estimate a genotype's evolutionary fitness.

Key words: artificial populations; fitness; genetic model; Helianthus annuus*; hybridization; introgression; phenology, flowering; seed predation; sunflowers; transgenic; weed.*

INTRODUCTION

One of the major concerns with the commercialization of genetically engineered crops is transgene ''escape'' from crops into populations of wild relatives. For example, crop-to-wild hybridization has the potential to introduce beneficial traits such as resistance to herbicides, insect herbivores, disease, and various types of environmental stress, perhaps causing weedy plants to become more vigorous and abundant (e.g., Snow and Morán Palma 1997, Ellstrand et al. 1999, Hails 1999). Weedy relatives are likely to acquire genes from crop cultivars when they co-occur, share pollen vectors, have overlapping flowering periods, and do not have strong reproductive barriers that prevent hybridization and introgression (Ellstrand and Hoffman 1990, Keeler and Turner 1990). Examples of crops that hybridize spontaneously with wild/weedy populations include sunflower (Arias and Rieseberg 1994), squash (Kirkpatrick and Wilson 1988), radish (Klinger et al. 1991, Snow et al. 2001), foxtail millet (Till-Boutrand et al. 1992), sorghum (Arriola and Ellstrand 1996), rice (Oka and Chang 1959, 1961), carrot (Small 1984), and oilseed rape (Jorgensen and Andersen 1994). Little is known about the persistence or ecological effects of crop genes that enter wild populations via pollen movement (Small 1984, Rissler and Mellon 1996, Snow and Morán Palma 1997). Empirical data and models are needed to predict the consequences of commercial-scale cultivation of transgenic crops before these new varieties are evaluated for deregulation.

Information regarding the likelihood of crop allele persistence typically comes from two types of studies: those that indirectly estimate the likelihood of gene persistence using ecological data on hybrid survival and fecundity (Langevin et al. 1990, Klinger and Ellstrand 1994, Frello et al. 1995, Linder and Schmitt 1995, Arriola and Ellstrand 1996, Mikkelsen et al. 1996, Snow et al. 1998), and those that directly estimate the frequency of cultivar markers in generations following hybridization (Luby and McNicol 1995, Whitton et al. 1997, Linder et al. 1998). In this study, we integrate both approaches. First, we estimate the frequency of selectively neutral crop genes using ecological data. Second, we evaluate our predictions by comparisons with directly estimated frequencies of three crop-specific allozyme markers.

Manuscript received 3 August 2001; revised 7 January 2002; accepted 31 January 2002; final version received 21 March 2002.
[4] Corresponding author. E-mail: halexander@ku.edu

When these two methods yield similar results, this implies that ecological data can be used to estimate the frequency of fitness-related transgenes in field trials that precede commercialization.

Single locus selection models may adequately predict the frequencies of introgressing transgenes because transgenic traits are typically conferred by single-locus genes with Mendelian inheritance (e.g., Dietz-Pfeilstetter and Kirchner 1998). Although these models are a simplification of actual individual life cycles, they have the advantage of being easy to understand intuitively and require only simple calculations to predict genotype frequencies in the next generation. Estimation of the parameters in a simple one-locus selection model requires empirical data that are relatively straightforward to obtain, such as relative survival and relative seed production for different genotypes. When a simple model can adequately estimate changes in genotype frequencies, it provides a useful tool to synthesize information about the biology of a population and to predict genetic composition in the next generation.

In this study we focus on crop alleles in sunflowers (*Helianthus annuus* L.). Crop and wild sunflower cooccur in many regions of the United States. Wild *H. annuus* is a native, annual weed that is widespread throughout much of the United States, reaching its greatest abundance in the Midwest (Heiser 1954); the domesticated sunflower (also *H. annuus*) is commonly cultivated in the Plains states and California. Wild sunflower, which is usually found in crop margins, roadsides, and unmanaged, disturbed areas, is considered to be a serious weed in corn, soybean, cultivated sunflower, and other crops (e.g., Gillespie and Miller 1984, Geir et al. 1996, Al-Khatib et al. 1998). Although no transgenic sunflowers have been deregulated to date, biotechnology companies are developing transgenic varieties that are resistant to insects (Lepidoptera and Coleoptera), a fungal disease (*Sclerotinia*), and herbicides.[5]

Previous research has shown that transgenes in cultivated sunflower are very likely to spread to nearby populations of wild *H. annuus* via pollen movement (Arias and Rieseberg 1994). In experimental populations, foraging bees carried crop-specific genetic markers to wild plants as far as 1000 m from stands of cultivated sunflower (Arias and Rieseberg 1994). In addition, a 6.4-km isolation zone is recommended to prevent commercial sunflower seed nurseries from being contaminated by unwanted wild sunflower pollen (Smith 1978). Crop-specific genetic markers have been found at frequencies of 25–42% in wild sunflower populations growing adjacent to the crop, providing strong evidence for spontaneous hybridization (Arias and Rieseberg 1994, Whitton et al. 1997). Previous studies show that F_1 wild–crop hybrids usually had lower fecundity than wild plants, but were vigorous and capable of backcrossing with wild plants (Snow et al. 1998). Episodes of crop-to-wild hybridization are likely to lead to the long-term persistence of nondeleterious crop genes in wild populations (Linder et al. 1998).

To investigate this process following a known hybridization event, we established three artificial populations with crop-specific genetic markers. Our questions were:

(1) Will frequencies of neutral crop markers increase, decrease, or remain constant over time in wild populations? (2) Is the lifetime fecundity (survival $\times$ seed production) of F_1 hybrid genotypes relative to wild genotypes a useful predictor of the proportions of crop genes that are transmitted to the next generation?

Methods

Experimental populations

In 1997, we established three experimental populations in Kansas, each consisting of 100 wild seedlings and 100 F_1 wild–crop hybrids (hereafter referred to as hybrids). We used these founder plants to establish semi-natural populations that were followed for two growing seasons. Methods are described in Cummings et al. (1999) and below. In this study, we examined the effects of cross type (wild vs. hybrid) on survival, lifetime seed production, flowering phenology, and predispersal seed predation. We did not include all components of fitness, such as early seedling establishment, male reproductive success, postdispersal seed predation, or the longevity of dormant seeds. We used our data to estimate expected frequencies of three crop-specific genetic markers one generation later (1998 seedlings) to allow comparisons with observed allele frequencies of the 1998 seedlings.

Hybrids in the experimental populations came from hand pollinations between wild and crop genotypes that differed at three allozyme loci (malate dehydrogenase, MDH; menadione reductase, MNR; and 6-phosphoglucodehydrogenase, 6-PGD), such that F_1 hybrids were heterozygous for a crop-specific allele at each locus (see Rieseberg and Seiler [1990] for allozyme methods). To determine whether these loci are tightly linked, we scored 100 progeny from a cross between a hybrid and a wild plant. Segregation ratios conformed to Mendelian expectations and all of the alleles at each pair of loci segregated independently, based on chi-square tests. MNR and 6-PGD may be loosely linked (chi-square = 3.24, 1 df, $P < 0.1$), but larger sample sizes would be required to determine this.

Hand pollinations were carried out in a pollinator-free greenhouse using 17 wild plants from Kansas (seeds collected from Sedgewick, Greenwood, Johnson, and Kay Counties); our crosses should have led to approximately equal contributions of each wild plant. Flower heads on these wild plants received pollen from either another wild plant or one of 10 cultivated plants (USDA cultivar 894). The wild and cultivated plants differed at

[5] URL: ⟨http://www.nbiap.vt.edu⟩

TABLE 1. Survival and fecundity differences in hybrid and wild sunflower plants at each site in 1997.

A) Survival and total flower head production

Site	Plant survival (%)		No. heads†	
	Hybrid ($N = 100$)	Wild ($N = 100$)	Hybrid	Wild
1	100	98	532 (9.6)	5000 (90.4)
2	100	100	269 (10.0)	2423 (90.0)
3	99	100	469 (10.5)	4000 (89.5)

B) Seed production per head

Site	Hybrid heads		Wild heads		Total no. good seeds	
	All seeds	Good seeds‡	All seeds	Good seeds‡	Hybrid	Wild
1	49.1 ± 6.36	29.0 ± 4.16 (59)	131.3 ± 10.40	127.0 ± 10.20 (97)	15 428	635 000
2	22.0 ± 4.56	11.25 ± 2.80 (51)	120.2 ± 9.38	113.27 ± 8.63 (94)	3026	274 453
3	40.9 ± 4.26	25.5 ± 3.02 (62)	170.3 ± 11.7	167.8 ± 11.70 (99)	11 960	671 200

Notes: Seeds per head is based on averages from 100 flower heads (mean ± 1 SE). ''Good'' seeds are seeds that appear to be filled (hard when squeezed with forceps) with no obvious signs of predispersal insect damage (mean ± 1 SE, $N = 100$ heads). Total seeds per site is based on the total number of heads (see part A of table) multiplied by the average number of good seeds per head.

† Numbers in parentheses indicate the percentage contribution of each cross type to the total number of heads in the population.

‡ Numbers in parentheses indicate the percentage of all seeds that were undamaged.

three allozyme loci. None of the wild plants set seed autonomously, being self-incompatible, and unintentional cross-pollination was not a problem since only hand-crossed flower heads (capitulae) set seed. In early June 1997, seeds (i.e., achenes) were nicked to induce germination and several hundred seedlings per cross type were planted in a greenhouse at the Ohio State University. On 16–17 June, these seedlings were transported to Lawrence, Kansas, where they were transplanted to 10 cm diameter pots and placed in partial shade to acclimate to outdoor conditions.

The experimental populations were established in 1997 at three sites (sites 1, 2, and 3) at the University of Kansas Field Station and Ecological Reserves (12 km northeast of Lawrence, Kansas). We chose the site locations to minimize gene flow among populations; sites were ≥0.5 km apart and were separated by small patches of forest and uneven topography. Each site was a grassy, early successional area lacking trees, 25 × 55 m in area, and surrounded by a 3 m high fence to exclude deer and other large vertebrate herbivores.

Within each site, we established a gridwork of 600 holes that were 1.2 m apart, divided into 10 blocks. Within each block, 10 holes were randomly assigned to wild seedlings, 10 were randomly assigned to hybrid seedlings, and the remaining 40 holes were left empty to be used for the founders of the 1998 populations. Each hole contained a 14 cm diameter plastic pot that was 14.5 cm deep and filled with local field soil (bottoms were removed to allow root growth). We planted the founders of our 1997 populations (100 seedlings per cross type at each site) on 30 June (site 1), 1 July (site 2), and 8 July (site 3). A few seedlings died due to transplant shock and were replaced with extras of the same cross type (early survival did not differ between cross types). Throughout the study, the grass between pots was mowed periodically and all local wild sunflower heads within an ~0.5 km radius of each plot were removed; we also removed the few nonexperimental plants that inadvertently appeared within the soil-filled pots. Peak flowering for local sunflowers occurred during early September.

When seeds produced in 1997 were mature, we collected a random, pooled sample of seeds from the plants at each site to obtain founding populations for the following year (see details below). Over the winter, groups of seeds were placed in replicated mesh bags in the soil at a depth of ~8 cm near site 1; thus, any effects of postdispersal seed predation on subsequent allele frequencies were not examined (see Alexander et al. 2001). Seeds were excavated in April 1998, just as they began to germinate, and planted in flats in an outdoor lathhouse. Young seedlings were transplanted to 10 cm diameter pots. Then, we randomly chose 400 seedlings to become the founders of the 1998 populations. These seedlings were transplanted into sunken 14 cm diameter pots at each of the three field sites, as described above (planting dates: 27 May, site 1; 28 May, site 2; 29 May, site 3). Each plot and the surrounding area was maintained as in 1997.

Fitness components of wild vs. hybrid plants

Survival after transplantation was nearly 100% (Table 1), so the main fitness component we recorded was an estimate of female fecundity. Wild sunflowers produce many flower heads of varying sizes over a period of many weeks, making it difficult to directly count seed production. We quantified the numbers of seeds produced by each cross type at each site by multiplying the average number of heads per plant (total flower head production divided by number of plants) by the average number of seeds per head (averages based on

seed counts from a random sample of 100 heads of each cross type from each site (see Cummings et al. 1999 for details).

Seeds were collected to obtain a random, representative sample of all the seeds produced in the 1997 populations. Throughout the flowering period, we labeled 30% of the wild heads in each plot, selecting these at random during each labeling session (approximately every 2–3 d), and recorded the flowering date of each labeled head to within 1–2 days' accuracy. We then estimated the total number of wild heads per plot by multiplying the total number of labeled heads by 3.333. Hybrid plants produced far fewer flower heads, so every head was labeled and dated on these plants. Labeled heads of both cross types were covered with bridal veil bags (tulle cloth, 25×25 cm^2) after the petals senesced to prevent loss of seeds from bird predation or seed dispersal. Bagging the heads at this stage does not appear to affect damage by predispersal insect seed predators (D. Pilson, *personal communication*).

Rather than use all labeled heads for seed collections (i.e., >1000 heads per site), we obtained a random, representative subset of all seeds produced in each plot. These seeds were used for three purposes: to count the number of seeds per head, score for allozyme frequencies, and serve as founders of the 1998 populations. To get a representative subset of the seeds produced in each plot, bagged heads were clipped and sorted based on their flowering dates. A file was developed that contained all collected heads; using a blocked, computer-generated random number process, we sampled a constant proportion of heads at each date. We then randomly chose heads (many more heads than we needed, to minimize sampling error) from this representative list and put their seeds into a container. Heads that produced more seeds would have a greater representation in this container; this is reasonable because a plant or head that produces more seeds should have greater representation in the next generation. We mixed the seeds from these heads, and then used a subset of this mixed seed pool to establish the next generation. In addition, the seeds from a random sample of 100 flower heads from each site and of each cross type were counted to obtain average numbers of seeds produced per head. These heads were also examined for levels of predispersal insect seed damage (see Cummings et al. 1999 for details).

In 1998, the genotypes of the adult plants were unknown and sampling was thus uniform throughout each site. Flower heads were sampled as described above to obtain a constant proportion of flowering heads across flowering dates for genetic analysis. Total flower head production differed among sites and different percentages were collected in different sites (5% at site 1, 10% at site 2, and 5% at site 3). The seeds from the collected heads (~300 heads per site) from each site were placed into separate containers, mixed, and the desired numbers of seeds were counted out.

Persistence of crop-specific genetic markers

A major goal of this study was to determine whether data on the relative fitness of F_1 hybrids can be used to predict allele frequencies in the next generation (1998 seedlings). Initial 1997 allele frequencies were known to be 0.25 for the crop alleles of MDH, MNR, and 6-PGD, because half of the plants in each population were F_1 wild–crop hybrids. We sampled each experimental population at three later stages: seeds in autumn 1997 (all seeds at a site pooled and 309–452 randomly selected seeds were genotyped per locus per site), seedlings in spring 1998 (361–400 seedlings were genotyped per locus per site), and seeds in autumn 1998 (all seeds at a site pooled and 376–513 randomly selected seeds were genotyped per locus per site). Seedlings were easier to score electrophoretically than seeds, and about half of the seeds were germinated prior to electrophoresis. This was accomplished by nicking the achene, storing the seed in moist, semi-sterile conditions in a refrigerator for 3–4 d, and planting the seeds in moist soil in a greenhouse. Newly emerged seedlings were collected and ground for enzyme extraction. Germination rates were high (>90%), so we presume that using germinated seeds did not bias our results against genotypes that may have been dormant.

Calculating expected allele frequencies for 1998 seedlings

We used a general model of selection (Bodmer 1965) to calculate expected genotype frequencies in the generation following the F_1 generation (1998 seedlings). We set the frequency of wild homozygous genotypes to be P_{11}, crop–wild heterozygous genotypes to be P_{12}, and frequency of crop homozygous genotypes to be P_{22}. Bodmer (1965) examined a very general, one-locus model of fecundity selection:

$$P'_{11} = \frac{w_{11}}{\bar{w}}\left[f_{11}m_{11}P_{11}^2 + \frac{1}{2}(f_{11}m_{12} + f_{12}m_{11})P_{11}P_{12} + \frac{1}{4}f_{12}m_{12}P_{12}^2\right] \quad (1)$$

$$P'_{22} = \frac{w_{22}}{\bar{w}}\left[\frac{1}{4}f_{12}m_{12}P_{12}^2 + \frac{1}{2}(f_{12}m_{22} + f_{22}m_{12})P_{12}P_{22} + f_{22}m_{22}P_{22}^2\right] \quad (2)$$

$$P'_{12} = \frac{w_{12}}{\bar{w}}\left[\frac{1}{2}(f_{11}m_{12} + f_{12}m_{11})P_{11}P_{12} + (f_{11}m_{22} + f_{22}m_{11})P_{11}P_{22} + \frac{1}{2}f_{12}m_{12}P_{12}^2 + \frac{1}{2}(f_{12}m_{22} + f_{22}m_{12})P_{12}P_{22}\right]. \quad (3)$$

In this set of equations the frequencies of the three possible genotypes in the next generation are calculated by knowing: (1) the frequencies of the nine possible mating types, specified by the multiplication of the frequency of the two genotypes in question (e.g., random mating: P_{11} [male] $\times$ P_{11} [female]; P_{11} [male] $\times$ P_{12} [female]; P_{11} [male] $\times$ P_{22} [female]; and similarly for crosses of P_{12} [male] and P_{22} [male]; see Hedrick 1985: Table 5.8); (2) the fecundities of genotypes ij (f_{ij} and m_{ij} in females and males, respectively); (3) the relative viabilities (survival) of the three genotypes (w_{11}, w_{12}, w_{22}); and (4) $\bar{w}$ is defined such that $P'_{11} + P'_{12} + P'_{22} = 1$ (also see Hedrick 1985:180–183).

We took the above formulae and made some simplifications. First, since we had no ''pure crop'' (i.e., P_{22}) parental plants, the frequency of P_{22} was equal to zero. Also, since survival (viability) from seedling to adult was close to 100% in all of our sites (Table 1), we let $w_{11} = w_{12} = w_{22} = 1$ and thus, did not incorporate viability selection in our model. We estimated genotype frequencies in the parental populations by the number of flower heads produced at a site. This was done because a hybrid plant producing two flower heads does not have the same opportunities for mating as does a wild plant producing 20 flower heads. Thus P_{11} and P_{12} were estimated from the data collected on flower head production at a site for both cross types (as given in Table 1A). We estimated relative female fecundities by the number of seeds produced by an average flower head of a given cross type based on seed counts of a random sample of 100 flower heads of each cross type from each site (Table 1B; see Cummings et al. 1999 for details). We also assumed that there is no selection on pollen donors (i.e., $m_{11} = m_{12} = m_{22}$), thus simplifying the fecundities to the additive model examined by Penrose (1949) (see Hedrick 1985:181 for a discussion). This decision was made based upon the difficulty of obtaining quantitative estimates of male fecundity.

Using the above conceptual framework, we created three models. For model 1, relative female fecundities (f_{11} and f_{12}) were estimated using the total number of seeds. For model 2, we added the effects of seed predation by estimating relative female fecundities (f_{11} and f_{12}) using the total number of seeds escaping predispersal seed predation by insects (= good seeds; see Cummings et al. 1999 for details). Model 3 was developed because of the observation that hybrids flowered much earlier than wild plants (Fig. 1). In this case, we estimated genotype frequencies in the 1998 seedlings assuming pure positive assortative mating (hybrid plants crossing only with hybrid plants and wild plants crossing only with wild plants) (model 3). This was incorporated by recalculating mating type frequencies such that only P_{11}^2 and P_{12}^2 matings occurred, and recalculating mean fitness based on this new mating table. This gave us the following equations:

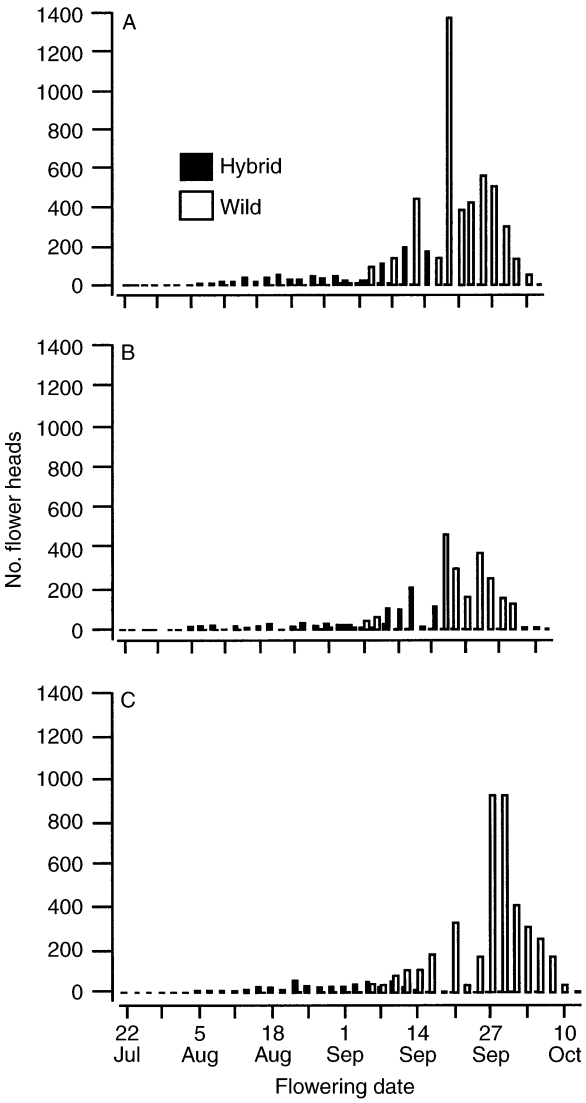

FIG. 1. Phenology of flower head production by wild vs. hybrid plants at each of three experimental populations in 1997.

$$P'_{11} = \frac{P_{11} f_{11} + \frac{1}{4} P_{12} f_{12}}{\bar{w}} \qquad P'_{12} = \frac{\frac{1}{2} P_{12} f_{12}}{\bar{w}}$$

$$P'_{22} = \frac{\frac{1}{4} P_{12} f_{12}}{\bar{w}}.$$

The genotype frequencies and female fecundities used in model 2 (see above) were used in this model.

Comparing expected and observed allele frequencies

Observed and expected genotype frequencies under Hardy-Weinberg, and model 1, model 2, and model 3 were compared using a chi-square goodness-of-fit test with the expected and observed counts of genotypes

P'_{12} and P'_{22} lumped so as not to incur unacceptably small counts in these categories. The observed and expected genotype frequencies under all four models were also examined by calculating likelihood estimates for all four models:

$$\text{likelihood}_{\text{model}} = \frac{N!}{X_{11}!X_{12}!X_{22}!} \prod_{i=1}^{3} P_i^{x_i} \quad (4)$$

where X_{11}, X_{12}, and X_{22} are the observed numbers of individuals of each genotype for a site, N = the sum of X_{11}, X_{12}, and X_{22} (this is equivalent to saying that N equals the average sample size for a site across the three loci examined) (values from Appendix), and P_1, P_2, and P_3 are the probabilities of an individual being of a wild homozygote, a hybrid heterozygote, or a crop homozygote genotype, respectively (i.e., the expected genotype frequencies for the three genotypes) as predicted by the model.

The likelihood values were used in a Bayesian approach to calculate posterior probabilities of an individual being each of the three genotypes using the Bayesian formula:

$$\text{posterior probability}_{\text{model}_i} = \frac{\text{prior}_i \ \text{likelihood}_{\text{model}_i}}{\sum_{i=1}^{\text{all}} \text{prior}_i \ \text{likelihood}_{\text{model}_i}}. \quad (5)$$

The prior probabilities for four models (Hardy-Weinberg, model 1, model 2, model 3) for site 1 were assumed to be 0.25 each. The posterior probability for each model was then calculated for site 1 and these posterior probabilities were then used as the prior probabilities for site 2, and the calculated posterior probabilities for each model from site 2 were used as the prior probabilities for site 3. In this way the degree of belief in each model was assessed in three independent situations. The closer the posterior probability for a model is to 1.0, the greater support there is for that model, relative to the other models being considered. This Bayesian approach is a way of distinguishing degree of support for the four models, which can be considered four working hypotheses, a philosophically important method (Hilborn and Mangel 1997). The Bayesian calculations were performed by creating a C++ program (Metrowerks 1999, Cummings 2000).

Results

Survival, reproduction, and predispersal seed damage—1997 growing season

Survival of transplanted seedlings did not differ between cross types and was ~100% at each site (Table 1). Hybrid plants flowered earlier than wild plants, with very little overlap in flowering periods, and produced far fewer seeds per plant (Fig. 1, Table 1). Wild plants had about nine times more heads and three to five times more seed production per head. Predispersal predation by lepidopteran and coleopteran larvae was disproportionately heavy for hybrids. Predispersal predation had little effect on wild plants (97% of seeds/head were undamaged) in comparison to hybrid plants (58% of seeds/head escaped predispersal predation) (Table 1B; Cummings et al. 1999). Site 2 was the least productive in terms of total seed production (Table 1B). After accounting for differences in head number, seeds per head, and predispersal seed predation, we estimate that hybrid plants produced ~2.4% as many total viable seeds as wild plants at site 1, 1.1% as many at site 2, and 1.8% as many at site 3 (Table 1B).

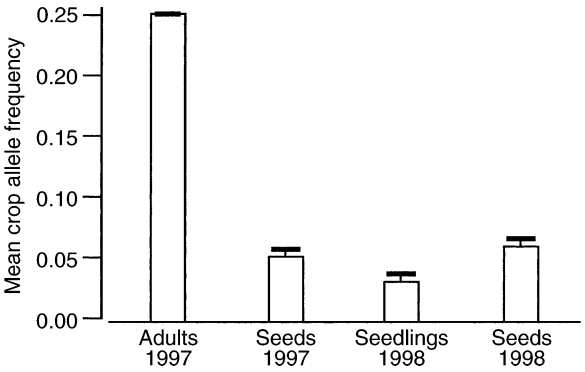

FIG. 2. Observed frequency of crop-specific markers of founders of the 1997 populations (n = 600 plants), seeds from 1997 (n = 1084 seeds), founders of the 1998 populations (n = 1177 plants), and seeds from 1998 (n = 1395 seeds). Values represent means of three sites (populations), calculated by averaging data from three allozyme markers at each site and then averaging across sites. Detailed sample size information is presented in the Appendix.

Allozyme frequencies in 1997 and 1998

Initially, each experimental population consisted of 50% F_1 hybrids, so the frequency of crop alleles was 0.25. Crop allele frequencies dropped to ~0.03–0.05 across the three loci examined in the first generation (1997 seeds and 1998 seedlings) and remained at ~0.05 in the next generation of seeds (i.e., seeds from 1998 plants; Fig. 2, Appendix). Crop allele frequencies and genotype frequencies were relatively consistent across the three loci examined and across the three experimental populations, although the frequencies in the 1998 seedlings were slightly lower than the other two groups (Fig. 2, Appendix).

We used four models to calculate the expected allele frequencies for the 1998 seedlings. The null model of Hardy-Weinberg clearly did not predict genotype frequencies well (Tables 2 and 3). The one-locus fecundity selection model (model 1) accurately predicted wild, crop–wild, and crop genotype frequencies in 1998 seedlings (fecundity values were based on abundance of flower heads and total seed production). This model predicted genotype frequencies that were not significantly different from the observed genotype frequencies (Table 2). There was a high calculated Bayesian

TABLE 2. Comparisons between observed and expected frequencies of wild (w) and crop (c) alleles in the 1998 seedlings, based on Hardy-Weinberg and three models described in *Methods*.

Models	Wild allele frequency	Crop allele frequency	Wild (ww)	Hybrid (wc)	Crop (cc)	χ^2	*N*
Hardy-Weinberg (all sites)	0.75	0.25	0.5625	0.3750	0.0625	†	
Site 1							
Observed	0.965	0.035	0.9392	0.05162	0.0091		730
Fecundity selection (model 1)	0.966	0.034	0.9338	0.0653	0.0009	0.05	
+ Seed predation (model 2)	0.970	0.030	0.9407	0.0587	0.0006	0.02	
+ Assortative mating (model 3)	0.999	0.001	0.9981	0.0013	0.0006	671.06†	
Site 2							
Observed	0.976	0.024	0.9546	0.0429	0.0025		396
Fecundity selection (model 1)	0.970	0.030	0.9405	0.0590	0.0005	0.06	
+ Seed predation (model 2)	0.972	0.028	0.9448	0.0549	0.0003	0.03	
+ Assortative mating (model 3)	0.999	0.001	0.9991	0.0006	0.0003	2312.74†	
Site 3							
Observed	0.964	0.036	0.9344	0.0588	0.0067		397
Fecundity selection (model 1)	0.967	0.033	0.9345	0.0647	0.0007	0.00	
+ Seed predation (model 2)	0.969	0.031	0.9392	0.0603	0.0005	0.01	
+ Assortative mating (model 3)	0.999	0.001	0.9984	0.0010	0.0005	1604.22†	

Notes: The observed frequencies are calculated from the averages of three allozyme markers at each population (see the Appendix; sample sizes for chi-square analysis are averages of all three markers). Calculation of expected frequencies is described in *Methods*. The Hardy-Weinberg expectation was always significantly different from observed values.

† $P < 0.0001$ for differences between observed and expected genotype frequencies.

''degree of belief'' in this model for two of the three loci examined, based on calculation of model likelihoods at three sites and analysis by Bayes theorem (Table 3).

We added additional ecological information to two similar models. Model 2 incorporated differences in predispersal seed predation as well as differences in seed production. This model also predicted genotype frequencies that were not significantly different from the observed genotype frequencies (Table 2). There was a high calculated Bayesian degree of belief in the model for one of the three loci examined (Table 3). Model 3 incorporated seed production differences, seed predispersal predation differences, and based predictions of genotype frequencies on the assumption of complete positive assortative mating. Model 3 did not accurately predict genotype frequencies. The predicted genotype frequencies were significantly different from the observed genotype frequencies (Table 2), and the model had a very low Bayesian degree of belief at all three loci examined.

Table 3 does reveal differences among loci in Bayesian degree of belief. Under the assumption of independence of loci, we can provide a simple multi-locus estimate for each model by multiplying the posterior probabilities for each locus and normalizing this product by dividing by the sum of the products over all models. With this approach, the multi-locus probability for model 1 is >0.999 and is <0.001 for the other models. As mentioned in the *Methods*, MNR and 6-PGD could be loosely linked and hence dependent; if this proved to be true, this simple multi-locus approach would not be valid.

DISCUSSION

Persistence of crop alleles in wild populations

Frequencies of crop-specific genetic markers decreased dramatically from the 1997 seedlings (25% crop alleles) to the population of nondamaged seeds that were produced by these plants (5% crop alleles) and the seedlings that emerged in 1998 (3% crop alleles). This reflects the fact that wild plants produced far more seeds than F_1 hybrids. However, crop-specific allele frequencies remained near 5% in the seeds produced by 1998 plants, suggesting that after an initial drop in frequency, selectively neutral crop alleles may continue at low but stable levels in wild populations. It is likely that backcross genotypes (crop–wild F_1 ×

TABLE 3. Bayesian ''degree of belief'' (posterior probability) for each model for each locus.

Locus	Model	Posterior probability
MDH	Hardy-Weinberg	1.84×10^{-252}
MDH	fecundity (model 1)	0.2888
MDH	seed predation (model 2)	0.7112
MDH	assortative mating (model 3)	2.28×10^{-112}
PGD	Hardy-Weinberg	3.13×10^{-220}
PGD	fecundity (model 1)	0.9974
PGD	seed predation (model 2)	0.0025
PGD	assortative mating (model 3)	5.28×10^{-103}
MNR	Hardy-Weinberg	1.55×10^{-242}
MNR	fecundity (model 1)	0.9890
MNR	seed predation (model 2)	0.0100
MNR	assortative mating (model 3)	1.19×10^{-90}

Note: The closer the posterior probability for a model is to 1.0, the greater support there is for that model, relative to the other models being considered.

wild) had less of a fitness disadvantage than F_1 plants, as seen by Morán Palma (1998), such that the allozyme markers were no longer associated with deleterious crop traits (i.e., less branching and fewer heads per plant). Repeated episodes of hybridization followed by backcrossing could allow higher levels of selectively neutral crop alleles to accumulate over time in wild populations.

It is not easy to compare our study of crop allele frequencies to others. For example, many studies focus on crop-to-wild gene flow within a single generation (e.g., Kirkpatrick and Wilson 1988, Langevin et al. 1990, Klinger et al. 1991, Wilson and Manhart 1993). In sunflower, Arias and Rieseberg (1994) found high levels (27%) of crop–wild hybridization in experimental sunflower plants that were adjacent to sunflower cultivars. Persistence of crop alleles in later generations has been documented, but the rate at which this occurs is not known. After a single hybridization event, Whitton et al. (1997) documented 42% wild–crop hybrids at a crop margin near a wild sunflower population in California, and crop-specific markers were still present five years later. Linder et al. (1998) found >30% of the seeds in natural sunflower populations had crop-specific markers after up to 40 yr of contact with the crop. With such strong evidence that crop alleles will occur in wild sunflower populations, we focused on a more predictive approach for evaluating crop allele frequencies.

Using ecological data to predict allele persistence

In this study, the lifetime fecundity (survival × seed production) of F_1 hybrid genotypes relative to wild genotypes was a useful predictor of the frequencies of crop-specific alleles in the next generation. A simple, one-locus selection model accurately predicted wild, crop–wild, and crop genotype frequencies in the 1998 seedlings based on fecundity differences between cross types in 1997. The 1997 wild and crop–wild F_1 plants had similar survival, but wild plants had nine times more heads, three to five times more seeds per head, and much less predispersal seed predation compared to hybrid plants. In previous experiments, we found a great deal of variation in the fecundity of F_1 hybrids relative to wild plants (Snow et al. 1998). For example, in a common garden experiment in Kansas, hybrid plants produced about half as many flower heads as wilds, and wild plants from Nebraska produced similar numbers of flower heads as their corresponding hybrids. This variation among experiments and populations would presumably influence frequencies of crop alleles that introgress into wild populations, resulting in higher levels of introgression than occurred in the present study.

Model 1 (fecundity selection) and model 2 (fecundity + seed predation) gave accurate predictions of crop and wild genotype frequencies, but it is curious that adding information to our selection model to improve our ability to predict genotype frequencies was largely unsuccessful. Addition of predispersal seed predation levels in wild–crop and wild plants did not seem to improve the fit of the model (except for genotypes based on examination of MDH, Table 3). Addition of information on flowering phenology, suggesting a positive assortative mating scenario rather than a random mating scenario, substantially reduced the accuracy of the prediction of genotype frequencies.

There are two general explanations for the lack of predictive power of the assortative mating model. On one hand, it is possible that a model of complete positive mating is too extreme, despite the large differences in phenology for the two types. Alternatively, assortative mating is occurring, but significant events occurring at other life stages acted against the tendency of assortative mating to reduce the frequency of crop alleles. For example, the effects of male fecundity differences between genotypes might have acted to increase crop allele frequency if crop–wild hybrid flowers were relatively more likely to contribute pollen than wild flowers. However crop–wild hybrid plants had open flowers early in the season (Fig. 1), releasing pollen before many other plants had flower heads, and thus seem unlikely to have relatively greater chances at being fathers of the next generation. Alternatively, the absence of seed dormancy in our models could be significant. Wild sunflower seeds can remain dormant in the soil for several years (Burnside et al. 1996, Teo-Sherrell 1996). Seed dormancy differences in crop–wild hybrid (90–95% seed germination) and wild (64%–72%) seeds have been reported (Snow et al. 1998). In the present study, greater germination of seeds produced by crop–wild hybrids would have the same result as increased survivorship of crop–wild hybrid seeds or seedlings, increasing the number of crop–wild hybrid seedlings in the population and the frequency of crop alleles, and thus acting in opposition to the effects of assortative mating on allele frequencies.

Other researchers have also examined the fecundity and viability of F_1 crop–wild hybrid plants (e.g., Wilson and Manhart 1993, Klinger and Ellstrand 1994, Snow et al. 1998, 2001). Quite often these studies make qualitative statements about the likelihood of crop allele persistence based on findings of high or low F_1 hybrid performance. Although these studies are valuable (e.g., sterile F_1 crop–wild hybrids would indicate a situation in which crop allele persistence is very unlikely), there is a need for a more quantitative framework to predict allele persistence. It is also vital to have a methodology in place to statistically test the quantitative predictions made by models incorporating F_1 fecundity and viability measurements. This study introduces both a quantitative framework for predicting crop allele persistence using the types of ecological data that are often collected, and a method for testing these predictions.

An overall objective of our studies on sunflower is

to predict crop allele frequencies in real-world situations of wild sunflowers growing near crop sunflower populations. Our experimental populations did experience natural weather variation, pollination, and insect predation. However, our experimental design did not mimic natural conditions of plant density, herbivory, postdispersal seed predation, or successional change. The high survival rate of all genotypes may have been due to the experimental environment in which seedlings were initially watered after transplantation and competing vegetation was trimmed. In order to obtain more realistic survival rates between genotypes, tagged genotyped individuals could be transplanted into wild populations and allowed to grow under more natural conditions. Such survival measurements would likely vary with density; in high density conditions, seedling mortality in sunflowers increases (Watkinson et al. 1983). Sunflowers are also affected by successional stage (Heiser 1954), and postdispersal seed predation (Robel and Slade 1965, Alexander et al. 2001; J. Nash and H. M. Alexander, *unpublished data*). All of these factors could affect F_1 vs. wild sunflowers differently and thus have the potential to change the relative survival and fecundity terms in the selection models. For example, postdispersal seed predators have been shown to take seeds of F_1 hybrid sunflowers at a higher relative frequency than seed produced by wild plants (Alexander et al. 2001).

Conclusions

Although we acknowledge that any experimental study will not accurately reflect all natural conditions, we have been able to predict crop allele frequencies in sunflower populations using ecological data and test our predictions statistically. The next step will be to follow crop allele frequencies in natural sunflower populations (as Whitton et al. 1997 have done) and collect more accurate fecundity and viability estimates for plants growing under natural conditions. It would then be possible to use these parameter estimates in a similar selection model to test the robustness of this approach. An extension to a multi-locus approach would be straightforward to develop and useful when trying to predict the fate of alleles with known selective advantages.

This study was motivated by concerns about the persistence of crop transgenes in wild populations. We used allozyme markers to trace the movement of crop genes; use of such markers was necessary given the difficulty in obtaining permission to use actual crop transgenes in field experiments. We presume these markers are selectively neutral; thus our study provides a minimum estimate of movement of crop alleles into wild populations. As noted by Rieseberg and Burke (2001), much greater gene movement is likely with selectively advantageous genes, such as transgenes conferring insect or disease resistance. With sunflower, the current study and results from past studies lead to the following scenario: (1) crop and wild sunflowers can easily cross, (2) the resulting crop–wild hybrids can vary greatly in fitness, from relatively low in this study to comparable fitness to wild plants in studies in other years/environments (Snow et al. 1998), and (3) because of the variation in hybrid fitness, the frequencies of crop alleles in wild populations will also vary. However, such crop alleles are likely to be maintained in wild populations over time (see *Results*; Arias and Rieseberg 1994, Whitton et al. 1997). Repeated hybridization events and backcrossing into wild populations will be important factors in this process.

An important contribution of this study is the use of an explicit quantitative approach for integrating ecological data on hybrids and wild plants with independent genetic data on crop allele frequencies. Future work is now needed to evaluate the population ecological consequences of crop alleles in wild populations; i.e., how does the presence of crop genes conferring insect or disease resistance specifically affect plant population size, persistence at a site, or ability to invade new habitats? Questions on crop-to-wild hybridization thus provide a practical illustration of the need for greater integration of population ecology and population genetics disciplines (Brussard 1978, Lewontin 1979, Antonovics and Via 1988, Alexander et al. 1996).

Acknowledgments

We thank the University of Kansas Field Station and Ecological Reserves staff for their assistance in the set-up and maintenance of this study and the many students who worked in both the field and laboratory. N. Slade, R. Holt, M. Orive, and two anonymous reviewers provided useful comments on this research. This study was supported by USDA BRARG 9601405 received by A. Snow, L. Rieseberg, and H. Alexander.

Literature Cited

Alexander, H. M., C. L. Cummings, L. Kahn, and A. A. Snow. 2001. Seed size variation and predation of seeds produced by wild and crop-wild sunflowers. American Journal of Botany **88**:623–627.

Alexander, H. M., P. H. Thrall, J. Antonovics, A. Jarosz, and P. Oudemans. 1996. Population dynamics and genetics of plant disease: a case study of anther-smut disease. Ecology **77**:990–996.

Al-Khatib, K., J. R. Baumgartner, D. E. Peterson, and R. S. Currie. 1998. Imazethapyr resistance in common sunflower (*Helianthus annuus*). Weed Science **46**:403–407.

Antonovics, J., and S. Via. 1988. Genetic influences on the distribution and abundance of plants. Pages 185–203 *in* A. J. Davy, M. J. Hutchings, and A. R. Watkinson, editors. Plant population ecology. Blackwell Scientific, Oxford, UK.

Arias, D. M., and L. H. Rieseberg. 1994. Gene flow between cultivated and wild sunflowers. Theoretical and Applied Genetics **89**:655–660.

Arriola, P. E., and N. C. Ellstrand. 1996. Crop to weed gene flow in the genus *Sorghum* (Poaceae): spontaneous interspecific hybridization between Johnsongrass (*Sorghum halapense*) and crop sorghum, *S. bicolor*. American Journal of Botany **83**:1153–1160.

Bodmer, W. F. 1965. Differential fertility in population genetics. Genetics **51**:411–424.

Brussard, P. F. 1978. Ecological genetics: the interface. Springer-Verlag, New York, New York, USA.

Burnside, O. C., R. G. Wilson, S. Weiseberg, and K. Hubbard. 1996. Seed longevity of 41 weed species buried 17 years in eastern and western Nebraska. Weed Science **44**:74–86.

Cummings, C. L. 2000. Population genetic and population ecological consequences of crop–wild sunflower hybridization. Dissertation. University of Kansas, Lawrence, Kansas, USA.

Cummings, C. L., H. M. Alexander, and A. A. Snow. 1999. Increased pre-dispersal seed predation in sunflower crop-wild hybrids. Oecologia **121**:330–338.

Dietz-Pfeilstetter, A., and M. Kirchner. 1998. Analysis of gene inheritance and expression in hybrids between transgenic sugar beet and wild beets. Molecular Ecology **7**: 1693–1700.

Ellstrand, N. C., and C. A. Hoffman. 1990. Hybridization as an avenue of escape for engineered genes. BioScience **40**: 438–442.

Ellstrand, N. C., H. C. Prentice, and J. F. Hancock. 1999. Gene flow and introgression from domesticated plants into their wild relatives. Annual Review of Ecology and Systematics **30**:539–563.

Frello, S., K. R. Hansen, J. Jensen, and R. B. Jorgensen. 1995. Inheritance of rapeseed (*Brassica napus*)–specific RAPD markers and a transgene in the cross *B. juncea* × (*B. juncea* × *B. napus*). Theoretical and Applied Genetics **91**:236–241.

Geir, P. W., L. D. Maddux, J. Moshier, and P. W. Stahlman. 1996. Common sunflower (*Helianthus annuus*) interference in soybean (*Glycine max*). Weed Technology **10**:317–321.

Gillespie, G. R., and S. D. Miller. 1984. Sunflower competition in wheat. Canadian Journal of Plant Science **64**:105–111.

Hails, R. S. 1999. Genetically modified plants—the debate continues. Trends in Ecology and Evolution **15**:14–18.

Hedrick, P. W. 1985. Genetics of populations. Jones and Bartlett, Boston, Massachusetts, USA.

Heiser, C. B. 1954. Variation and subspeciation in the common sunflower, *Helianthus annuus*. American Midland Naturalist **51**:287–305.

Hilborn, R., and M. Mangel. 1997. The ecological detective. Princeton Press, Princeton, New Jersey, USA.

Jorgensen, R. B., and B. Andersen. 1994. Spontaneous hybridization between oilseed rape (*Brassica napus*) and weedy *B. campestris* (Brassicaceae): a risk of growing genetically modified oilseed rape. American Journal of Botany **81**:1620–1629.

Keeler, K. H., and C. E. Turner. 1990. Management of transgenic plants in the environment. Pages 189–281 *in* M. Levin and H. Strauss, editors. Risk assessment in genetic engineering: environmental release of organisms. McGraw-Hill, New York, New York, USA.

Kirkpatrick, K. J., and H. D. Wilson. 1988. Interspecific gene flow in *Cucurbita*: *C. texana* vs. *C. pepo*. American Journal of Botany **75**:519–527.

Klinger, T. D., R. Elam, and N. C. Ellstrand. 1991. Radish as a model for the study of engineered escape rates via crop–weed mating. Conservation Biology **5**:531–535.

Klinger, T., and N. C. Ellstrand. 1994. Engineered genes in wild populations: fitness of weed–crop hybrids of *Raphanus sativus*. Ecological Applications **4**:117–120.

Langevin, S. A., K. Clay, and J. B. Grace. 1990. The incidence and effects of hybridization between cultivated rice and its related weed red rice (*Oryza sativa* L.). Evolution **44**:1000–1008.

Lewontin, R. C. 1979. Fitness, survival, and optimality. Pages 3–21 *in* D. J. Horn, R. Mitchell, and G. R. Stairs, editors. Analysis of ecological systems. Ohio State University Press, Columbus, Ohio, USA.

Linder, C. R., and J. Schmitt. 1995. Potential persistence of escaped transgenes: performance of transgenic, oil-modified *Brassica* seeds and seedlings. Ecological Applications **5**:1056–1068.

Linder, C. R., I. Taha, G. J. Seiler, A. A. Snow, and L. H. Rieseberg. 1998. Long-term introgression of crop genes into wild sunflower populations. Theoretical and Applied Genetics **96**:339–347.

Luby, J. J., and R. J. McNicol. 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.

Metrowerks. 1999. Code warrior. Metrowerks, Austin, Texas, USA.

Mikkelsen, T. R., B. Andersen, and R. B. Jorgensen. 1996. The risk of crop transgene spread. Nature **380**:31.

Morán Palma, P. 1998. Case studies on reproductive ecology of lakeside daisy (*Hymenoxys heracea* = *H. acaulis* var. glabra) and evolution of weeds in *Helianthus annuus*. Dissertation. Ohio State University, Columbus, Ohio, USA.

Oka, H. I., and W. T. Chang. 1959. The impact of cultivation on populations of wild rice, *Oryza perennis* f. *spontanea*. Phyton **13**:105–117.

Oka, H. I., and W. T. Chang. 1961. Hybrid swarm between wild and cultivated rice species *Oryza pernnis* and *O. sativa*. Evolution **15**:418–430.

Penrose, L. S. 1949. The meaning of fitness in human populations. Annals of Eugenics **14**:301–304.

Rieseberg, L. H., and J. M. Burke. 2001. The biological reality of species: gene flow, selection, and collective evolution. Taxon **50**:47–67.

Rieseberg, L. H., and G. J. Seiler. 1990. Molecular evidence and the origin and development of the domesticated sunflower (*Helianthus annuus*). Economic Botany **44**:79s–91s.

Rissler, J., and M. Mellon. 1996. The ecological risks of engineered crops. MIT Press, Cambridge, Massachusetts, USA.

Robel, R. J., and N. A. Slade. 1965. The availability of sunflower and ragweed seeds during fall and winter. Journal of Wildlife Management **29**:202–206.

Small, E. 1984. Hybridization in the domesticated weed-wild complex. Pages 195–210 *in* W. F. Grant, editor. Plant biosystematics. Academic Press, Toronto, Ontario, Canada.

Smith, D. L. 1978. Planting and seed production. Pages 371–372 *in* Sunflower science and technology. Agronomy 19, American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, Madison, Wisconsin, USA.

Snow, A. A., and P. Morán Palma. 1997. Commercialization of transgenic plants: potential ecological risks. Bioscience **47**:86–96.

Snow, A. A., P. Morán Palma, L. H. Rieseberg, A. Wszelaki, and G. J. Seiler. 1998. Fecundity, phenology, and seed dormancy of F_1 wild-crop hybrids in sunflower (*Helianthus annuus*, Asteraceae). American Journal of Botany **85**:794–801.

Snow, A. A., K. L. Uthus, and T. M. Culley. 2001. Fitness of hybrids between weedy and cultivated radish: implications for weed evolution. Ecological Applications **11**:934–943.

Teo-Sherrill, C. P. A. 1996. The fates of weed seeds. Dissertation. Department of Agronomy, University of Nebraska, Lincoln, Nebraska, USA.

Till-Bouttrand, I., X. Reboud, P. Brabant, M. Lefranc, B. Rherissi, F. Vedal, and H. Darmency. 1992. Outcrossing and hybridization in wild and cultivated foxtail millets: consequences for the release of transgenic crops. Theoretical and Applied Genetics **83**:940–946.

Watkinson, A. R., W. M. Lonsdale, and L. G. Firbank. 1983. A neighbourhood approach to self-thinning. Oecologia **56**: 381–384.

Whitton, J., D. E. Wolf, D. M. Arias, A. A. Snow, and L. H. Rieseberg. 1997. The persistence of cultivar alleles in wild populations of sunflowers five generations after hybridization. Theoretical and Applied Genetics **95**:33–40.

Wilson, H., and J. Manhart. 1993. Crop/weed gene flow: *Chenopodium quinoa* Willd. and *C. berlandieri* Moq. Theoretical and Applied Genetics **86**:642–648.

APPENDIX

Frequencies of crop- and wild-specific allozyme markers of seeds from 1997, seedlings in the 1998 populations, and seeds from 1998. Numbers of individuals in each category are listed in parentheses.

	MDH				6-PGD			
Source	Wild (ww)	Both (wc)	Crop (cc)	Crop allele frequency	Wild (ww)	Both (wc)	Crop (cc)	Crop allele frequency
1997 seeds								
Site 1	0.852 (368)	0.148 (64)	0.000 (0)	0.0741 (432)	0.936 (365)	0.046 (18)	0.018 (7)	0.0410 (390)
Site 2	0.906 (319)	0.091 (32)	0.003 (1)	0.0483 (352)	0.905 (314)	0.089 (31)	0.006 (2)	0.0504 (347)
Site 3	0.896 (292)	0.101 (33)	0.003 (1)	0.0537 (326)	0.930 (293)	0.054 (17)	0.016 (5)	0.0429 (309)
1998 seedlings								
Site 1	0.952 (701)	0.048 (35)	0.000 (0)	0.0263 (361)	0.922 (636)	0.059 (41)	0.019 (13)	0.0331 (378)
Site 2	0.947 (372)	0.051 (20)	0.002 (1)	0.0280 (393)	0.962 (381)	0.035 (14)	0.003 (1)	0.0202 (396)
Site 3	0.927 (366)	0.073 (29)	0.000 (0)	0.0367 (395)	0.940 (373)	0.058 (23)	0.002 (1)	0.0315 (397)
1998 seeds								
Site 1	0.873 (448)	0.125 (64)	0.002 (1)	0.0643 (513)	0.935 (433)	0.054 (25)	0.011 (5)	0.0378 (463)
Site 2	0.889 (448)	0.100 (49)	0.014 (7)	0.0625 (504)	0.951 (423)	0.036 (16)	0.013 (6)	0.0315 (445)
Site 3	0.868 (328)	0.101 (38)	0.032 (12)	0.0820 (378)	0.971 (365)	0.021 (8)	0.008 (3)	0.0186 (376)

APPENDIX. Continued.

	MNR				
Source	Wild (ww)	Both (wc)	Crop (cc)	Crop allele frequency	Mean crop allele frequency
1997 seeds					
Site 1	0.950 (377)	0.038 (15)	0.012 (5)	0.0315 (397)	0.0489
Site 2	0.905 (324)	0.095 (34)	0.000 (0)	0.0475 (358)	0.0487
Site 3	0.903 (297)	0.079 (26)	0.018 (6)	0.0578 (329)	0.0514
					0.0497 (all sites)
1998 seedlings					
Site 1	0.942 (719)	0.048 (37)	0.010 (7)	0.0198 (379)	0.0264
Site 2	0.955 (382)	0.042 (17)	0.003 (1)	0.0238 (400)	0.0240
Site 3	0.937 (373)	0.045 (18)	0.018 (7)	0.0402 (398)	0.0361
					0.0288 (all sites)
1998 seeds					
Site 1	0.888 (396)	0.110 (49)	0.002 (1)	0.0572 (446)	0.0531
Site 2	0.925 (459)	0.075 (37)	0.000 (0)	0.0373 (496)	0.0438
Site 3	0.876 (331)	0.124 (47)	0.000 (0)	0.0622 (378)	0.0543
					0.0504 (all sites)