Ecological Applications, 12(4), 2002, pp. 1119–1124

IMPROVING RISK ASSESSMENT FOR NONTARGET SAFETY OF TRANSGENIC CROPS

MICHELLE MARVIER[1]

Department of Biology and Environmental Studies Institute, Santa Clara University, Santa Clara, California 95053 USA

Abstract. In many countries, government regulations require environmental risk assessment prior to commercial sale and widespread planting of transgenic crops. Here I evaluate the design and statistical rigor of experiments used by industry to assess the safety of transgenic plants for nontarget organisms, as required under U.S. regulations. This review reveals that a few simple improvements in experimental design could greatly increase the rigor and information content of studies required under current regulations. For example, although most experiments were conducted for 1–4 wk, some of the tested species can live a year or more and could experience much longer periods of exposure. Moreover, the number of replicates used in these studies was generally quite small (usually 2–6 replicates per treatment), resulting in experiments that had little chance of detecting real effects. Clearly, sample sizes should be bolstered, and nonsignificant results should be accompanied by an analysis of statistical power. In addition, information readily available over the Internet is insufficient for a quantitative assessment of a transgenic crop's safety. Improved access to information regarding the details of risk assessment studies could greatly increase the public's ability to evaluate industry's claims of safety.

Key words: Bacillus thuringiensis; Bt *toxin; experimental design; genetically engineered crops; insecticidal properties; nontarget organisms; risk analyses; sample size; statistical power; transgenic crops.*

INTRODUCTION

Genetically engineered crops have prompted often acrid criticism from environmental and food safety groups. Although the benefits of this new technology are potentially large (more food from less land, cheaper medicine, reduced pesticide use), the risks remain uncertain because the technology is relatively new and the variety of traits that could be produced is enormous (Dunwell 1999). For reviews of potential risks, see Wolfenbarger and Phifer (2000) and Marvier (2001). Policy makers have responded to this uncertainty by imposing health and environmental safety regulations in the form of experimental tests of ''effects'' to be performed before each new transgenic variety is approved for commercial release. Environmental advocates complain that the regulations are not sufficiently precautionary. Are there ways of making the scientific testing that is currently required more compelling and informative? I address this question by focusing on one particular dimension of environmental safety: unintended effects on nontarget organisms stemming from plants genetically modified to have insecticidal properties.

As of January 2001, the U.S. Department of Agriculture (USDA) had approved 15 petitions for deregulation involving crops genetically modified for enhanced insect resistance (Table 1). All 15 of these deregulated varieties were modified to include genes from the bacterium *Bacillus thuringiensis*, thereby causing the crops themselves to produce *Bt* toxin. I focus on tests conducted on these *Bt* crops under the auspices of the Environmental Protection Agency (EPA) because *Bt* crops are common and likely to become even more important on a global basis (James 2000), and because EPA has well-developed procedures for testing pesticides (Code of Federal Regulations, Title 40, Part 158), which it has modified for insecticidal transgenic crops. Here I evaluate the experimental design for a suite of studies that examined the effects of *Bt* crops on nontarget invertebrate species. These analyses point toward some specific recommendations regarding how we might better perform risk analyses for genetically modified crops.

Manuscript received 27 March 2001; revised 5 November 2001; accepted 14 November 2001.

[1] E-mail: mmarvier@scu.edu

EPA FRAMEWORK FOR NONTARGET TESTING

Under the Federal Food, Drug, and Cosmetic Act and the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), the EPA has the authority to regulate pesticides, and this authority has been interpreted to extend to crops genetically engineered to produce pesticidal compounds. The EPA is responsible for determining the environmental and health risks that a plant pesticide might pose and for registering compounds deemed unlikely to cause unreasonable harm. Thus, one step toward commercial sale, or deregulation, of a *Bt* crop is registration of the crop's *Bt* toxin with the EPA. The environmental safety of each *Bt* crop variety is individually assessed because transgenic *Bt* crops vary

TABLE 1. Petitions for deregulation of engineered transgenic *Bt* crops that have been approved by the USDA as of January 2001.

Crop and *Bt* protein	Corporation	USDA petition no.	Engineered trait(s)†
Potato			
Cry3a	Monsanto	99-173-01p	IR (beetle), VR‡
Cry3a	Monsanto	97-339-01p	IR (beetle), VR§
Cry3a	Monsanto	97-204-01p	IR (beetle), VR‖
Cry3a	Monsanto	95-338-01p	IR (beetle)¶
Cry3a	Monsanto	94-257-01p	IR (beetle)#
Tomato			
Cry1Ac	Monsanto	97-287-01p	IR (lepidopteran)
Cotton			
Cry1Ac	Calgene	97-013-01p	IR (lepidopteran), HT
Cry1Ac	Monsanto	94-308-01p	IR (lepidopteran)
Corn			
Cry9c	AgroEvo	97-265-01p	IR (lepidopteran), HT
Cry1Ab	Monsanto	96-317-01p	IR (lepidopteran), HT
Cry1Ac	DeKalb	96-291-01p	IR (lepidopteran)
Cry1Ab	Monsanto	96-017-01p	IR (lepidopteran)
Cry1Ab	Northrup King	95-195-01p	IR (lepidopteran)
Cry1Ab	Monsanto	95-093-01p	IR (lepidopteran)
Cry1Ab	Ciba-Geigy	94-319-01p	IR (lepidopteran)

† Abbreviations: IR, insect resistant; VR, virus resistant; HT, herbicide tolerant.
‡ Similar to a line covered in petition 97-204-01p, except transformed with a different plasmid (the marker gene is resistance to glyphosphate instead of kanamycin).
§ Resistant to potato virus Y.
‖ Resistant to potato leaf roll virus.
¶ Transformed ''Superior'' and ''Atlantic'' cultivars (male fertile).
# Transformed ''Russet Burbank'' cultivar (male sterile).

in the particular type and amount of *Bt* protein produced (e.g., Cry1Ab, Cry3A, Cry9C). Certain *Bt* proteins target lepidopteran pests whereas others target coleopteran or dipteran pests. Even within an insect order, *Bt* toxins affect some species more than others, depending, in part, on the pH of the particular insect's gut (Peferoen 1997).

In 1994, the EPA published proposed regulations (Federal Register Volume 59:60495, 23 November 1994) regarding data requirements for the environmental safety assessment of transgenic plants expressing pesticidal traits (now called plant incorporated pesticides). After receiving considerable public and scientific feedback, the EPA issued final versions of these rules on 17 January 2001, but these rules are currently under review by the new administration. The proposed data requirements are generally similar to those required by the EPA for microbial pesticides (Code of Federal Regulations, Title 40, Part 158, Section 740).

BASIC EXPERIMENTAL DESIGN OF RISK ASSESSMENT STUDIES

To assess ecological risks of transgenic crops, it is important to know how effective the experimental and statistical protocols are. One important consideration is that the pattern, duration, and extent (dosage) of exposure must at least equal those experienced by the organisms in nature. Table 2 summarizes sample sizes, number of individuals per experimental replicate, *Bt* concentrations, and duration for each of 29 individual experiments concerning the effects of *Bt* toxin on 10 different nontarget invertebrate species. These experiments were drawn from six different petitions for deregulation, involving four different crop species and three different corporations.

Testing generally involves exposing nontarget organisms to high concentrations of *Bt* toxin, usually 10–100 times the concentration that is lethal to 50% of target individuals (LC_{50}). The assumption underlying this ''maximum hazard approach'' is that if no effect is detected while using such high doses, then the chance of effects at the anticipated lower doses should be miniscule. It is worth noting that recent studies have demonstrated that *Bt* toxin can bind to certain types of soil particles, depending on soil pH, and remain biologically active for extended periods (Tapp and Stotzky 1998, Saxena et al. 1999). Soil binding could allow *Bt* toxin to build up to much higher concentrations than previously expected. However, with respect to dosage, testing methodologies generally appear to be conservative: assessing concentrations that are 1–2 orders of magnitude above those lethal to the target species will probably err on the side of demonstrating nontarget effects, if any exist.

IMPORTANCE OF STATISTICS: SAMPLE SIZE AND STATISTICAL POWER

Scientific inquiry typically involves formulation of a null hypothesis (a hypothesis of no effect), followed

TABLE 2. Experimental design of studies used to assess *Bt* toxicity to nontarget invertebrates.

Nontarget species, grouped by USDA petition no. and crop	Common name and life stage tested	No. replicates	No. nontarget individuals exposed per replicate	Concentration of *Bt* protein (μg/g)†	Duration (d)
99-173-01p, potato					
Folsomia candida	collembola	5	10	200	21
Xenylla grisea	collembola	6	10	200	21
Apis mellifera	honey bee larvae	4	20	100	18
Eisenia foetida	earthworms	4	10	100	14
94-257-01p, potato					
Hippodamia convergens	ladybird beetles	6	25	100	‡
Nasonia vitripennis	parasitic wasps	2	25	100	‡
Apis mellifera	honey bee larvae	4	50	100	until emergence
Apis mellifera	honey bee adults	3	40	100	‡
Chrysopa carnea	green lacewing larvae	30	1	417	‡
97-287-01p, tomato					
Apis mellifera	honey bee larvae	4	50	20	11
Apis mellifera	honey bee adults	3	40	20	7
Nasonia vitripennis	parasitic wasps	2	25	20	23
Hippodamia convergens	ladybird beetles	2	25	20	30
Chrysopa carnea	green lacewing larvae	30	1	20	11
97-013-01p, cotton					
Eisenia foetida	earthworms	4	10	10 000	14
Folsomia candida	collembola	4	10	1000	28
96-317-01p, corn					
Apis mellifera	honey bee larvae	4	50	20	11
Apis mellifera	honey bee adults	3	40	20	9
Chrysopa carnea	green lacewings	30	1	16.7	7
Brachymeria intermedia	parasitic wasps	2	25	20	30
Hippodamia convergens	ladybird beetle	2	25	20	9
Eisenia foetida	earthworms	4	10	200	14
Xenylla grisea	collembola	6	10	200	28
Folsomia candida	collembola	5	10	200	21
94-319-01p, corn					
Daphnia magna	daphnia	2	10	19, 32, 54, 90, and 150 mg pollen/L	2
Eisenia foetida	earthworms	4	10	500	14
Coleomegilla maculata	ladybird beetle	3	15		1st instar to adults
Apis mellifera	honey bee larvae	4	25	1 mg pollen	until emergence
Folsomia candida	collembola	4	10	125, 250, and 500	28

Notes: The studies included here are representative of a larger set of studies for which not all details were available. Because of lengthy procedures at the EPA for releasing these studies, I obtained information about studies submitted to EPA indirectly from 15 USDA petitions for deregulation and directly from the Monsanto Corporation. The number of replicates is high for green lacewings because they are cannibalistic and therefore cannot be tested in batches.

† Units are μg/g except where noted otherwise.

‡ Until there is 20% mortality in the control group.

by an experiment that rigorously strives to disprove, or reject, the null hypothesis. Where the null hypothesis is not rejected, either there truly is no effect *or* an effect exists, but the experiment and statistical tests were simply not powerful enough to detect this effect. In the case of nontarget effects of transgenic crops, the companies performing the experiments have invested a great deal in research and development and would, in all likelihood, prefer that the null hypothesis were not rejected. In light of industry's conflicting interests, it is perhaps not terribly surprising that few experiments used more than EPA's recommended minimum number of replicates (Table 2). The use of few replicates makes it less likely that an effect could be detected, even if it does, in fact, exist.

A recent report by the FIFRA Scientific Advisory Panel (FIFRA SAP 2000) recommended that EPA pay more attention to certain elements of experimental design when offering guidelines for nontarget testing. In particular, the group pointed out that, rather than requiring a particular number of samples, a far more appropriate approach would involve selecting a desired level of statistical power and an effect size that we wish to be able to detect, collecting preliminary data to estimate the amount of within-treatment variability and then calculating the required sample size (FIFRA SAP 2000). I have followed up on the SAP recommendations and calculated the required sample size for five industry studies of nontarget effects, which were the only studies for which I could obtain the infor-

mation necessary for the calculations. For these calculations, I chose the standard 5% chance of wrongly rejecting the null hypothesis (α) and a 90% chance of detecting a true difference ($\beta = 0.10$). For purposes of comparison, I assumed that the mean of the control group remained fixed and then calculated the effect sizes (δ) corresponding to a 50% and a 20% difference between the treatment and control groups. For instance, if survival in the control treatment were 90%, then a 50% difference would correspond to only 45% survival of the *Bt*-treated organisms, and a 20% difference would mean that *Bt*-treated organisms experience 72% survival. Clearly, these differences are substantial. Starting with an arbitrary initial number of samples (n), I solved iteratively for the required sample size:

$$n \geq (2s_p^2/\delta^2)(t_{\alpha(2),\nu} + t_{\beta(1),\nu}) \quad (1)$$

where s_p^2 is the pooled variance, the t values are critical values of Student's t distribution and, for studies with balanced sample size, $\nu = 2(n - 1)$ (Zar 1996).

Only one of the five industry studies included a sufficient number of replicates to have had a 90% chance of detecting a 50% reduction in the survival (or fecundity, etc.) of organisms exposed to a *Bt* crop compared to organisms exposed to the unmodified form (Fig. 1). None of the five studies included a sufficient number of samples to have had a 90% chance of detecting a 20% difference between treatments. In fact, the sample sizes required to detect a 20% effect with high consistency ranged from 16 to 39 replicates and were far larger than any of the sample sizes used in industry testing (Fig. 1). To detect a smaller difference between treatments 90% of the time would require even more replicates. Exactly what size of an effect we should expect to be able to detect is a policy decision that must weigh the increase in experimental rigor against the increased cost of adding more replicates to an experiment. However, policy makers should remain mindful that, of all the factors that can reduce the overall error rate for an experiment, sample size is the factor most directly under the experimenter's control.

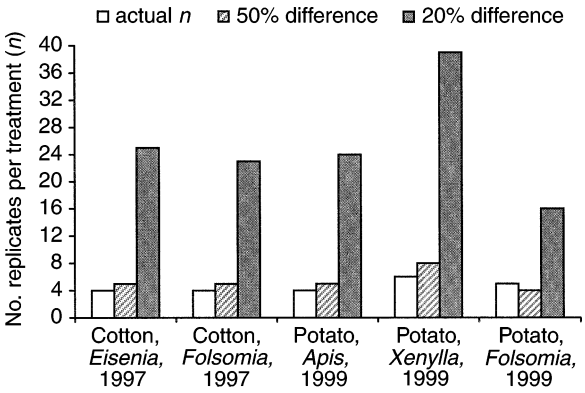

FIG. 1. Actual and required sample sizes for studies of *Bt* toxicity of transgenic crops. For five experiments examining nontarget effects, the columns represent the actual sample size used, the sample size required to detect a 50% difference between control and experimental treatments, or the sample size required to detect a 20% difference between the control and experimental treatments (for calculation, see *Importance of statistics: Sample size and statistical power*). Calgene studies on the effect of *Bt* cotton on earthworms (*Eisenia foetida*) and collembola (*Folsomia candida*) were part of USDA petition 97-013-01p. Data used to assess impact were untransformed percentage gain in body mass for earthworms and untransformed percentage surviving for collembola. Monsanto studies on the effects of *Bt* potato on honey bee larvae (*Apis mellifera*) and collembola (*Xenylla grisea* and *Folsomia candida*) were part of USDA petition 99-173-01p. Data used were the number of honey bees that emerged and progeny production for the two collembola species. For cases in which studies included additional control treatments, I selected only the nontransgenic version of the *Bt* treatment. The Monsanto collembola studies compared three different *Bt* toxins to a single control treatment. In this case, I used the variance from the control group as the estimate of pooled variation.

How Can Risk Assessment for Transgenic Crops Be Improved?

A few simple improvements in experimental design could greatly increase the rigor and information content of studies required under current EPA regulations. First, the duration of experiments assessing the risks to nontarget organisms should be extended to more accurately reflect the pattern and duration of exposure that these organisms are likely to experience under field conditions. In particular, the duration of studies is, in many cases, quite short relative to the life spans of the test organisms. For example, studies involving earthworms typically lasted 14 d, but earthworms can live >4 yr in the laboratory and ~1 yr in nature (Edwards 1998). Similarly, studies on toxicity to ladybird beetles ranged from 9 to 30 d, but *Hippodamia convergens* live anywhere from 3 mo to 1 yr, depending on the climate (Gordon 1985). It has been demonstrated previously that duration of exposure to toxic compounds such as lead acetate and copper is negatively related to the performance of organisms as diverse as earthworms and rainbow trout (Marr et al. 1996, Saint-Denis et al. 2001). Thus, if a nontarget organism is potentially exposed for several months, experiments should test the impact of similarly prolonged exposures. The SAP recommended allowing experiments to continue until 20% mortality of individuals in the control treatment is attained (FIFRA SAP 2000). This would be an important improvement in the design of these studies, because brief exposures with limited follow-up are unlikely to detect effects. In the field, exposure is more likely to be chronic and effects may be slow to materialize.

Second, a rigorous test of safety must include a substantial number of replicates. There is always some chance that the ''answer'' emerging from a statistical test of a hypothesis may be wrong, but statistical analyses quantify the probability that the difference between treatments is real, rather than an artifact of mere chance. Scientists typically set the probability of

wrongly rejecting the hypothesis of no effect (the Type I error rate, α) to 5%. This broadly accepted standard allows a reasonably conservative test of the null hypothesis. However, in risk assessment, a difference between treatments often signifies the potential for irreparable harm. In such cases, wrongly concluding that the treatments differ (a Type I error) would result in less harm than would an incorrect conclusion of ''safety'' (a Type II error, β). Scientists and science philosophers involved in risk assessment have recently called for more attention to be paid to the probability of wrongly accepting the hypothesis of no effect (Underwood 1997, Pool and Esnayra 2001, Shrader-Frechette 2001). To appreciate the importance of large sample sizes, consider a study that documented the effects of a biological control agent, *Rhinocyllus conicus*, on a native nontarget thistle, *Cirsium canescens*. Louda et al. (1997) compared seed production in 181 thistle heads infested *with R. conicus* and 40 uninfested thistle heads. In contrast, the majority of studies of nontarget effects for genetically modified crops reviewed here have employed fewer than five samples per group (Table 2). If Louda et al. (1997) had examined only five thistle heads per group, (assuming the same 86% reduction in seed production and same pooled variance), they would have found no significant difference in seed production between the two groups, with and without the introduced weevil. In this report, I have shown for some real examples of nontarget studies how increases in sample size can reduce the chances of Type II errors. Another means of achieving reduced Type II error rates, which might be combined with moderate increases in sample size, would be to increase the Type I error rate (e.g., set α at 10% so $P < 0.10$ would be deemed statistically significant).

Third, access to information regarding experimental studies of nontarget effects should be improved (National Research Council 2000). EPA biopesticide fact sheets are posted on the Internet and would provide an ideal way to disseminate information regarding nontarget testing. These fact sheets currently do not provide the information needed for an assessment of experimental rigor. For example, in regard to effects on nontarget insects, the biopesticide fact sheet for a type of *Bt* cotton stated only that the toxin ''. . . caused no adverse effects to the parasitic wasp, *Nasonia vitripennis*, green lacewing larvae, honeybee larvae, honeybee adults, and adult ladybird beetles when fed at 1700 and 10 000 times the levels found in pollen and nectar'' (EPA 2000). Although these fact sheets are meant only to summarize information, they could be modified to include sample size, duration, dosage, measured variables (e.g., survival, fecundity), and the mean and standard deviation for each variable for each treatment. Only with this information in hand can anyone assess whether the conclusion of no adverse effects is credible. Industry would probably benefit if these data were summarized in a single database in which the weight of evidence supporting industry's claim of safety might accrue to the advantage of the industry.

Fourth, at least at a research level (as opposed to a regulatory level), there needs to be a broader examination of ecological risks. Ecological studies examining the nontarget effects of biological control agents suggest specific complications for which we should be on the lookout. For example, the time between the introduction of *Rhinocyllus conicus* and detection of its host range expansion was ~20 yr (Louda et al. 1997). Nontarget effects of transgenic crops might take similarly long periods to become detectable, and long-term monitoring should be required for commercially produced transgenic varieties (Pool and Esnayra 2001). Second, *R. conicus* exerted both direct effects on nontarget thistles and indirect effects on nontarget tephritid flies (Louda et al. 1997). Because indirect effects are generally harder to anticipate, monitoring efforts should involve a wide variety of species, including those that are unlikely to interact directly with the transgenic crop.

In summary, ecological risk assessment generally entails the uncertainty of extrapolating from small, lab-based studies to more variable and open natural systems, the possibility of evolution and environmental change altering the circumstances, intrinsic biological time lags, and a high degree of stochastic inputs. Reassurance of safety is no simple matter under these circumstances. In the studies reviewed here, the lack of a statistically significant effect was invariably interpreted as evidence of the crop's safety, with no exploration of the statistical power of the experimental design. Unfortunately, a nonsignificant statistical test that is not accompanied by a power analysis provides little insight as to whether or not an effect really exists (Underwood 1997). Given the enormous challenges that risk analysis entails, the use of sufficiently large sample sizes, combined with an explicit quantification of statistical power, is a necessary step toward increasing our confidence in a verdict of ''safety.'' Finally, a transparent and full disclosure of safety data from industry tests in an easily accessible database may go a long way toward quieting the public's uneasiness regarding the safety of this new technology.

Acknowledgments

Sincere thanks to Kay Peterson of the USDA for rapid response to requests for copies of petitions for deregulation. Comments by Peter Kareiva, Angelika Hilbeck, Matt Liebman, and two anonymous reviewers greatly improved an earlier version of this manuscript.

Literature Cited

Dunwell, J. M. 1999. Transgenic crops: the next generation, or an example of 2020 vision? Annals of Botany **84**:269–277.

Edwards, C. A. 1998. Earthworm ecology. St. Lucie Press, Boca Raton, Florida, USA.

EPA (Environmental Protection Agency). 2000. Biopesticide fact sheet: *Bacillus thuringiensis* subsp. *kurstaki* Cry1Ac Delta-Endotoxin and its controlling sequences as expressed

in cotton (006445). EPA Publication Number EPA 730-F-00-007. EPA, Washington, D.C., USA.

FIFRA SAP. 2000. FIFRA (Federal Insecticide, Fungicide, and Rodenticide Act) Scientific Advisory Panel Meeting, 8–9 December 1999, Arlington, Virginia, USA. SAP Report Number 99-06. EPA, Washington, D.C., USA. [Online, URL: ⟨http://www.epa.gov/scipoly/sap/1999/december/report.pdf⟩.]

Gordon, R. D. 1985. The Coccinellidae of North America. Journal of the New York Entomological Society **93**:1–912.

James, C. 2000. Global status of transgenic crops: challenges and opportunities. Pages 1–6 *in* A. D. Arencibia, editor. Plant genetic engineering: towards the third millennium: Proceedings of the International Symposium on Plant Genetic Engineering, 6–10 December 1999, Havana, Cuba. Elsevier, New York, New York, USA.

Louda, S. M., D. Kendall, J. Connor, AND D. Simberloff. 1997. Ecological effects of an insect introduced for the biological control of weeds. Science **277**:1088–1090.

Marr, J. C. A., J. Lipton, D. Cacela, J. A. Hansen, H. L. Bergman, J. S. Meyer, and C. Hogstrand. 1996. Relationship between copper exposure duration, tissue copper concentration, and rainbow trout growth. Aquatic Toxicology **36**:17–30.

Marvier, M. 2001. Ecology of transgenic crops. American Scientist **89**:160–167.

National Research Council. 2000. Genetically modified pest-protected plants: science and regulation. National Academy Press, Washington, D.C., USA.

Peferoen, M. 1997. Insect control with transgenic plants expressing *Bacillus thuringiensis* crystal proteins. Pages 21–48 *in* N. Carozzi and M. Koziel, editors. Advances in insect control: the role of transgenic plants. Taylor and Francis, London, UK.

Pool, R., and J. Esnayra. 2001. Ecological monitoring of genetically modified crops: a workshop summary. National Academy Press, Washington, D.C., USA.

Saint-Denis, M., J. F. Narbonne, C. Arnaud, and D. Ribera. 2001. Biochemical responses of the earthworm *Eisenia fetida andrei* exposed to contaminated artificial soil: effects of lead acetate. Soil Biology and Biochemistry **33**:395–404.

Saxena, D., S. Flores, and G. Stotzky. 1999. Insecticidal toxin in root exudates from *Bt* corn. Nature **402**:480.

Shrader-Frechette, K. 2001. Ecology. Pages 304–315 *in* D. Jamieson, editor. A companion to environmental philosophy. Blackwell: Malden, Massachusetts, USA.

Tapp, H., and G. Stotzky. 1998. Persistence of the insecticidal toxin from *Bacillus thuringiensis* subsp. *kurstaki* in soil. Soil Biology and Biochemistry **30**:471–476.

Underwood, A. J. 1997. Environmental decision-making and the precautionary principle: what does this principle mean in environmental sampling practice? Landscape and Urban Planning **37**:137–146.

Wolfenbarger, L. L., and P. R. Phifer. 2000. The ecological risks and benefits of genetically engineered plants. Science **290**:2088–2093.

Zar, J. H. 1996. Biostatistical analysis. Third edition. Prentice Hall, Upper Saddle River, New Jersey, USA.