

ISB NEWS REPORT

COVERING AGRICULTURAL AND ENVIRONMENTAL BIOTECHNOLOGY DEVELOPMENTS

JANUARY 2004

In This Issue:

WORKSHOP REPORT: EXTENDING THE NET FITNESS MODEL TO CONSIDERATIONS OF CROP GENE FLOW

The Muir risk assessment model and its application to crop plants	3
Using the E(NK) model to study genotype by environment (GxE) interactions	4
How to incorporate changes due to natural selection into the fitness model	5
Stochastic events, error in parameter estimation, and credibility intervals in predictions	6
Modeling needs within a risk assessment context	6
Adapting the net fitness model for plants	7
Data needs for net fitness modeling	8
Differing needs for modeling crop-crop versus crop-native plant gene flow	9
Model refinement and verification	9
Toward an integrated risk assessment model for plant transgene spread	10
Workshop Participants	10

SPECIAL ISSUE

WORKSHOP REPORT: EXTENDING THE NET FITNESS MODEL TO CONSIDERATIONS OF CROP GENE FLOW

Information Systems for Biotechnology (ISB) hosted the Environmental Risk Assessment Modeling Workshop II on October 9 and 10 in Colorado Springs, Colorado. The Workshop was convened to evaluate and extend the net fitness approach of Muir and Howard (2001, 2002) to plants for use in predicting the spread of a transgene in a wild population. William Muir and Richard Howard have described and developed a net fitness model to investigate the ecological risks of species extinction or invasiveness posed by deployment of transgenic organisms in nature (Muir and Howard, 2001; 2002). Their work, considering a model system of Japanese medaka modified through insertion of a growth hormone gene, gave rise to the Trojan Gene Hypothesis, which posits, "Pleiotropic effects of transgenes that have antagonistic effects on net fitness components can result in unexpected hazards, such as local extinction of the species containing the transgene (Muir and Howard, 1999)." An earlier ISB workshop addressed the use of the net fitness approach for assessing risks posed by transgenic animals (Hallerman, 2002). A forthcoming white paper discusses suggestions from that workshop to enhance the utility of the model as a tool for regulatory assessment of the risks posed by transgenic animal release to the environment.

For this second Environmental Risk Assessment Modeling Workshop, an invited panel of experts in evolutionary biology, population and quantitative genetics, and ecology, as well as specialists in regulatory risk assessment, considered the net fitness model for its applicability to plants. Discussion centered on applications and limitations to plants, suggested improvements that extend the model to transgenic plant species, recommendations for testing the model, and the potential for integration of the model into regulatory decision-making. The discussion was prefaced by a series of talks focused on key aspects of extending the net fitness approach. Dr. Muir described the net fitness model and provided examples of how it might be applied to considerations of crop-to-wild gene flow. Dr. Mark Cooper of Pioneer Hi-Bred International introduced the E(NK) model and discussed how it could be linked to the net fitness model to provide context-dependent information to address genotype by environment (GxE) interactions. Dr. Bruce Walsh of the University of Arizona discussed incorporating concepts of natural selection into gene frequency considerations. Dr. Guilherme Rosa of Michigan State University overviewed methods to improve stochastic utility of the net fitness model using Bayesian inference and Monte Carlo methods.

THE ISB NEWS REPORT

The material in this News Report is compiled by Information Systems for Biotechnology, funded as the National Biological Impact Assessment Program by a grant from USDA/CSREES to Virginia Tech. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the US Department of Agriculture or of Virginia Tech. The News Report may be freely photocopied or otherwise distributed with attribution.

Current and past issues of the ISB News Report are available at <http://www.isb.vt.edu>.

Editor: Ruth Irwin
rirwin@vt.edu

To order your free subscription:

Email: Send to news@nbiap.biochem.vt.edu
Type *subscribe newsreport*
[your name] in the message section. Do not include any additional text.

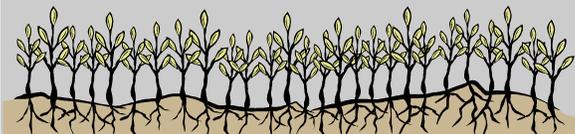
Phone: Call 540-231-3747

Web: Connect to <http://www.isb.vt.edu>.
Select "News Report," Subscribe."

ISB welcomes your comments and encourages article submissions. If you have a suitable article relevant to our coverage of the agricultural and environmental applications of genetic engineering, please email it to the Editor for consideration.

Information Systems for Biotechnology

207 Engel Hall
Virginia Tech
Blacksburg, VA 24061
Tel: 540-231-3747
Fax: 540-231-9070
Email: isb@vt.edu



Muir and Howard describe six net fitness components that allow quantitative prediction of the outcome of natural selection: juvenile viability, age to sexual maturity, mating success, female fecundity, male fertility, and adult viability. Estimates of these net fitness components for alternative genotypes can be incorporated into a model that predicts changes in gene frequency and population size. This information can be used to quantify the probability for transgene spread from domestic to wild species.

This Workshop considered the conceptual approach and data needs for extension of the net fitness approach for addressing crop to wild gene flow. Presentations and discussion also addressed the broadening and strengthening of the approach through integration of the net fitness model with consideration of

- (1) The consequences of gene by environment (G×E) interactions;
- (2) The within and between generation changes occurring with multi-gene selection in the presence of gene correlations;
- (3) The degree to which the transgene behaves as a major gene; and,
- (4) Statistical approaches to deal with the uncertainties in prediction of model outcomes.

Discussants recognized the value of the net fitness approach as a tool for assessing exposure (that is the probability for transgene spread in a wild population) as an aspect of the long-term risk of transgenic plant deployment. There was further recognition of the general applicability of the net fitness approach to plants, although the nature of the model structure and parameterization will prove case dependent based on the particular crop and receiving population that is considered. Integration of the net fitness model with models considering consequences G×E was deemed of high value.

The ability to acquire sufficient data to parameterize models was seen as the greatest difficulty to integration and extension of the net fitness approach. Discussion led to recognition that verification of the approach perhaps could be achieved through

- (1) Consideration of a genetic "lab rat" such as *Drosophila* for which existing data could be used to populate the model;
- (2) Development of a model system such as a glasshouse system with *Arabidopsis* where environmental zones and spatial structure provide for determination of net fitness within the context of a diverse set of environments;
- (3) Use of fundamental data for antagonistic genes (such as insecticide resistance, sickle cell, or other major gene effects) for determining the degree to which major gene/polygene concepts may apply to the spread of a transgene into a new background;



- (4) Consideration of how the model predicts behavior of well-studied invasive plant species;
- (5) Consideration of existing data for closely related crop and wild species occupying similar populations of environments (examples such as sugar-beet–sea beet, canola–wild mustard, cultivated–wild sun flower may prove useful); and/or,
- (6) Development of verification datasets for transgenic crops which are actively of concern relative to issues of crop-to-wild gene flow (cotton, *Brassica*, turf grasses, and conifers were suggested).

Each of these suggested approaches has limitations in terms of the data richness that may be achieved and the degree to which outcomes may be usefully applied to other plant species. The greater the number of these diverse approaches that can be used for verification, the greater will be the certainty of model predictiveness.

If verification of the integrated net fitness approach can be satisfactorily developed from one or more of these approaches, subsequent extension of the approach to regulatory risk assessments will require development of simplified screening level models that can address

- (1) Less rich data sets with high associated uncertainties; and,
- (2) Representative environmental scenarios, rather than the total population of receiving environments into which a transgenic plant may be deployed.

This may be achievable using probabilistic approaches that capture and integrate uncertainties into the overall evaluation of transgene spread. Workshop participants were introduced to concepts of Monte Carlo analysis and Bayesian statistics as a means whereby the integrated net fitness model could become stochastic.

Current regulatory schema within North America (principally Appendix II of the Canada and US Bilateral Agreement on Agricultural Biotechnology) provide a process for development of the requisite data to parameterize the net fitness model. Workshop participants recognized an important justification for the data requested in the Canada–US bilateral agreement on ecological data could be its usefulness in assessing gene flow through an integrated net fitness model.

Development of transgenic plants has a strong genetic thrust. Broadening the genetic research thrust of development activities to more explicitly recognize and encompass ecological consequences of transgenic plant release to the

environment in terms of the long-term consequences of gene flow on wild receiving populations is an area where further research support will advance understanding of transgenic plant risks and benefits.

The Muir risk assessment model and its application to crop plants

Muir and Howard described a first generation model predicting the effects of net fitness components on changes in transgene frequencies within a natural receiving population of con- and heterospecific organisms. This model provides insight as to whether a transgene introduced into a population of closely related organisms may spread and become fixed and predominate (that is, if the transgene confers invasive characteristics to the host) or whether that gene will become extinct. A key objective in advancing and expanding the net fitness approach is to develop ways whereby risk can be quantified in a manner that is transparent, objectively formulated, and to the greatest extent possible removes perception from the formulation, assessment, and communication of risk.

A leading encumbrance to the use of quantitative risk assessment for evaluation of transgenics is that definitions of risk are not universal. Muir has described risk and its components in a manner consistent with the National Research Council's "Red Book" (NRC, 1983) (and its quantitative application in a context consistent with ecological risk assessment). Thus, harm (an undesirable outcome) is considered from the standpoint of hazard (harm as affected by the potential environmental stressor) and exposure (the environmental occurrence of the stressor). Risk is the likelihood of harm resulting from the hazard being manifested under environmentally relevant conditions. Thus, in the case of a transgene as stressor, the risk is a joint probability of harm given the transgene spreads, $P(H/E)$, and the probability of the transgene spreading, $P(E)$:

$$\text{Risk} = P(H/E) * P(E)$$

Recognizing that because of the near infinite number of biotic interactions possible in and between ecosystems, it is not realistic to anticipate all possible harms. For those harms that can be anticipated, such as species displacement or extinction, the probability of such an event resulting cannot be predicted with our current state of the science even with foreknowledge that invasion by a given species or GM organism will occur. As such, the current focus in net fitness modeling is to quantify exposure—the ability for the transgene to escape, become fixed within a natural population, and spread.

Quantifying the ability of the transgene to spread can be approached directly and is amenable to universal rules of population genetics dependent on natural selection; it, therefore, represents the part of the risk formulation that population biology can address. Making the case that exposure (that is, transgene spread) approaches zero indicates that risk approaches zero as a consequence. This approach cannot eliminate the need to assess for rare events with catastrophic consequences; but given that these consequences cannot be demonstrated on regulatory timescales, the emphasis must remain on cautious prediction of transgene spread as a necessary first step in risk assessment.

Predicting the outcome of natural selection is a two step process involving (1) estimation of net fitness components for alternative genotypes, and (2) incorporation of parameters into a model that predicts change in gene frequency and population size. Prout (1971) originally described the relationship of net fitness components and population prediction, stressing a general approach with a small set of components, encompassing the entire life cycle, and amenable to experimental evaluation. Net fitness components and life history characteristics are synonymous terms. As reformulated and expanded by Muir and Howard, six net fitness components allow for the quantitative description of the outcome of natural selection—juvenile viability, age to sexual maturity, mating success, female fecundity, male fertility, and adult viability. In animals it makes sense to separate these components because the act of mating not only ensures delivery of the male gametes to the receptor, but can also preclude that female from mating with other males. In plants, however, some of these components may be combined, such as mating success and fertility.

The applicability of these net fitness parameters proves case dependent when applied to considerations of crop-to-wild gene flow. Juvenile viability describes the success in proceeding from fertile seeds to adults and will reflect the nature of the trait expressed by the transgene (for instance, host plant resistance to insect pests due to Bt toxin expression or competitive ability acquired through increased growth rate both affect juvenile viability). Age to sexual maturity affects the intrinsic rate of natural increase (for example, earlier maturity may increase the time over which flowering may occur and lead to greater viable seed production). Mating success can be thought of as a measure of out-crossing success and may include ability to attract pollinators, timing of pollination, pollen density, pollen survival, distance traveled, and ability of a pollen grain to fertilize an egg given that it falls on the receptor. Fecundity

can be measured as viable seed number. Adult viability (longevity) reflects the number of reproductive opportunities within a lifetime and will differ according to the nature of the crop considered. Implementing the net fitness approach for plants seems reasonable in light of data describing attributes of invasive plant species where net fitness parameters representing age at sexual maturity, fertility, and fecundity appear to be consequential discriminators of invasive versus non-invasiveness (see for instance Rejmanek and Richardson, 1996).

Changes in net fitness components for a transgene entering into a natural population may represent effects of the transgene itself or may reflect pleiotropic effects associated with the genetic background of the transgenic crop. For instance, a crop-wild hybrid may convey traits that increase (for instance, increased seed size and seed number) or decrease (uniformity of seed set, failure of seed to disperse) the net fitness in a wild population. The degree to which these background effects are linked to the transgene will be of consequence to the evolutionary fate of the transgene in the wild. If the transgene is inserted adjacent to polymorphic genes that affect fitness in some way, the joint effects of these genes must be considered because, although linkage effects dissipate in the long term as recombination events occur, in the short term linkage disequilibrium will be high because of hybridization upon out-crossing. Because transgenes are usually inserted into domesticated plants, which have been selected for increased productivity but are poorly adapted to natural environments, the transgene will most likely be linked with maladaptive alleles under natural conditions. Thus, linkage disequilibrium presents a challenge and an opportunity. A challenge is presented because use of the net fitness approach would underestimate the true benefit of the transgene in natural environments. An opportunity because such linkage may cause such a severe fitness disadvantage that the transgene would be eliminated by natural selection before recombination had an opportunity to separate the genes.

Using the E(NK) model to study genotype by environment (GxE) interactions

Ecological factors are important determinates for predicting the effects of net fitness parameters on natural selection. Thus, spatial explicitness may be necessary to account for the context dependence of genetic models and their application to risk assessment. Mark Cooper's presentation emphasized the value added to plant breeding strategies when variation across the target population of environments (TPE) was considered. Additive models for recruitment selection gains uniformly tend to over-predict the



realized gain. Focusing on GxE interactions (and GxExM—genotype by environment by management—interactions in the case of cultivated crops) allows for an understanding of how GxE may shift a trait (or suite of traits) within the population over time. Tightness in the GxE interaction explains the contribution of the environmental component and, thus, indicates when environmental considerations may drive natural selection to secondary optima in the response landscape.

The E(NK) model of Cooper (Cooper and Podlich, 2002) represents the extension to diploid plants of a model first formulated by Kaufman (1993), which describes environmental variance as related to the number of genes and interconnections involved in traits that may be components of fitness. The E(NK) model evaluates topological features of networks in a way that is useful for making connections to gene networks that interact. It provides insight as to how often we might be surprised at the outcomes of crop gene entry into a native population because of genotype by environment interactions that result in natural selection for a secondary optimum.

The environmental component E within the E(NK) model is a way to capture the environments that underlie the gene response and describe effects of genes as they are switched on within a TPE. These can be described as epistatic networks, $P = E(NK) + e$, where P is phenotype, N is gene number, K is the average number of connections among other genes in the genetic background, and e is error. The E(NK) model allows for consideration of the likelihood of correlation between gene responses in different environments. E(NK) is a “shrinkage parameter” where the shrinkage corrects for the observed versus predicted response and is dynamically pegged to genetics.

The net fitness model can be linked with an E(NK) model to be made context dependent. The utility of this approach may be in evaluating a trait under a broader number of environments, which globally improves the ability to predict the outcome of transgene occurrence in a natural population. Investigating the range of expectations that may be realized in practice within a context-dependent TPE for representative cases of crop-to-wild gene flow may allow for development of a screening level net fitness model that is able to differentiate gene flow events of high to moderate risk from low risk events. Because of the size of the TPE, a comprehensive analysis of all potential environments for transgene occurrence is unlikely; therefore, development of likely scenarios and representative evaluative environments will be necessary to usefully employ the linked model approach.

How to incorporate changes due to natural selection into the fitness model

As introduced in the presentations of Muir and Cooper, the direction of evolution for a transgene entering into a wild population is influenced by the genetic backgrounds into which the transgene is introduced and more importantly by changes in net fitness as a result of natural selection on that genetic background. To address this latter topic, Bruce Walsh described estimates of breeding value associated with a multi-gene trait. In selective breeding, a change in the direct selection may also drag along another trait to develop a between generation change (in the presence of genetic correlations). Therefore, selection outcomes must be considered for within and between generation changes in traits. Correlation in breeding values can be evaluated by developing a phenotypic covariance matrix, (the same as the GxE equation with the various traits treated as environments, both approaches are outgrowths of the Falconer (1961) model) and a genetic covariance matrix. This results in a multidimensional breeders’ equation that has a selection gradient ($P^{-1}S = b$, where P is the phenotypic variance covariance matrix, S is a vector of selection differential, and b is a vector of responses) which measures the direct selection on a character. The within generation response will not move in the same direction as will the underlying optimal response vector that reflects the multi-gene trait selection gradient. These observations have important correlates to understanding how a transgene (and the associated genetic background) entering into a natural population will behave within and between generations to influence the net fitness outcome.

Walsh provided further context to understanding the importance of genetic correlates in the background of a crop-wild hybrid by describing ways in which a transgene may act as a major gene. Major genes often impart negative effects on viability fitness that are later attenuated by natural selection on the background genotype to modify the effect to become less negative. Major genes include such traits as dwarfism. Thus, the degree to which a transgene behaves as a major gene is critical to net fitness considerations. Lande (1983) originally considered two (often antagonistic) fitness effects of the major gene; a major gene may improve one aspect of fitness (for instance, yield) and negatively affect another (perhaps water use efficiency). In fact, a major gene influencing one trait will likely influence a number of other secondary traits that are under polygenic control. Thus, the dynamics for the background polygenes are such that response over time for the major gene effect is lost due to the negative consequences of the polygene effects with which the major gene

is associated. If the frequency of a major allele is sufficiently rare, it may initially increase, and then be selected toward zero due to the background polygene effects. On the other hand, if the major gene is sufficiently common, it will become fixed (in most cases). Consequently, it is important to understand what constitutes sufficient rarity for a transgene frequency that it would disappear from a population. This aspect of transgene flow is not well enough understood to predict at this point.

Stochastic events, error in parameter estimation, and credibility intervals in predictions

Initial considerations of net fitness modeling for environmental risk assessment identified incorporation of stochasticity as a needed component in order that biological and statistical variance and uncertainty be addressed. There are numerous ways this may be approached. The Workshop focused on Bayesian inferences as a formal tool that is compatible with the needs of risk assessment, because the approach has formalized rules for describing variance and uncertainty in input distributions that are captured in the descriptions and communication of modeling outcomes.

Guilherme Rosa presented an overview of Bayesian inference and its applicability to model stochasticity into the net fitness model. The utility of the approach lies in making probability inferences for quantities about which we wish to learn—that is, its explicit use of probability to model uncertainty. This facilitates a common-sense interpretation of statistical conclusions (Gelman *et al.*, 2003; Shoemaker *et al.*, 1999).

Bayesian inference regarding a parameter (or a set of parameters) q is described in terms of probability statements. These probability statements are conditional on the observed data y , represented as a posterior distribution $p(q|y)$. The posterior distribution $p(q|y)$, obtained using the basic property of conditional probability known as Bayes's rule, is given by

$$p(\Theta|y) = \frac{p(\Theta, y)}{p(y)} = \frac{p(\Theta)p(y|\Theta)}{p(y)}$$

Equivalently, if the factor $p(y)$ is omitted, which can be considered a constant with y fixed, the posterior distribution can be expressed as

$$p(\Theta|y) \propto p(\Theta)p(y|\Theta)$$

where $p(q)$ is the prior distribution of q and $p(y|q)$ is the sampling distribution (or sampling model). Through $p(q)$,

any prior information about q can be incorporated into the model. Using then Bayes's rule, the prior distribution is updated with the information coming from the observed data, using the function $p(y|q)$ —which is proportional to the likelihood function.

Modern computing techniques that are based on (pseudo-) random number sequences, such as Markov Chain Monte Carlo methods, play a central role in Bayesian data analysis. Features of the posterior distribution of q can be approximated by drawing samples of q from $p(q|y)$ using, for example, a Gibbs sampler algorithm (Robert and Casella, 1999).

The process of Bayesian data analysis can be idealized by dividing it into the following three steps (Gelman *et al.*, 2003): 1) Setting up a full probability model, which comprises the choice of the sampling model and the prior distribution; 2) Conditioning on observed data, i.e., calculating and interpreting the posterior distribution; and 3) Evaluating the fit of the model and how sensitive are the results to the modeling assumptions, including the prior distribution.

An important consideration is the way priors concerning the input distributions are built. This can appear subjective without rigorous sensitivity analysis to ascertain the degree to which prior assumptions (for instance, as to distribution scale and shape) contribute to uncertainty in the model prediction (in the form of the fitness estimate, variance, and the model itself). This model uncertainty can be addressed in part through model checking of assumptions, comparisons with alternative assumptions, and averaging over alternatives.

Bayesian decision analysis is a further useful tool arising from Bayesian inference that is well suited to risk analysis. It involves optimization over decisions as well as averaging over uncertainties and can be defined by the following steps (Gelman *et al.*, 2003): 1) Enumeration of the space of all possible decisions and outcomes; 2) Determination of probability distributions of outcomes for each decision option; 3) Definition of a utility function mapping outcomes onto the real numbers; and 4) Choosing the decision with the highest expected utility as a function of the decision. This approach allows for risk managers to select the most appropriate analytical path to pursue during problem formulation for a subsequent risk assessment.

Modeling needs within a risk assessment context

Determining where to integrate a net fitness approach into regulatory risk assessments requires an understanding of



the logical endpoints for assessing gene flow risk. Given the recognized limitation in defining the effect of transgene spread on the environment, exposure endpoints that describe the ability or probability for the transgene to escape, to fix within a natural population, and to spread represent a practicable first step for current risk-driven decision-making. Part of the difficulty in arriving at a consensus endpoint relates to differing perspectives of risk assessors as to the time scale of concern for the assessment. Because regulatory decisions are made on time scales of years, and that conclusions concerning net fitness effects of the transgene are validated only over multiple generations, a stepwise use of endpoints with appropriate mitigation/monitoring options appears logical. Such a stepwise assessment must be implemented in a way that recognizes that for some transgenic plants (for instance, trees and grasses) there may be limited ability for transgene recall from the environment.

The particular endpoints selected and the method by which they are developed and evaluated are of necessity case- and model-specific due to the widely varied nature of transgenic plants that may need to be assessed. For instance, whereas viable pollen flow would appear a logical first tier endpoint for assessing transgene escape from an annual crop species, a similar consideration for a perennial crop would include both pollen flow and seed dispersal. Similarly, vegetative propagation of clonal material may be of importance for environmental dissemination of grass-expressed transgenes. Transgene escape in itself does not impact subsequent generations of a receiving population unless gene transmission occurs; therefore, measurements of out-crossing represent a second tier consideration when evidence for physical escape of transgenes (such as pollen flow) is present. Measurements of stable gene introgression and expression over subsequent generations are a third tier consideration that, if warranted, could be measured within the timeframe of regulatory decision-making.

Net fitness modeling could be most useful in terms of predicting stable gene introgression and could be usefully employed in a screening form for cases where there is demonstrated crop-wild hybridization. In this mode, the regulator would assess the probability of transgene spread to the wild population and differentiate high to moderate probability from low probability cases. For the moderate probability instances, additional evaluation of net fitness parameters would be warranted in order to refine risk. Appendix II of the Canada and United States Bilateral Agreement on Agricultural Biotechnology (<http://www.inspection.gc.ca/english/plaveg/bio/int/appenannex2e.pdf>) describes the essential elements

considered by the participating agencies for the environmental characterization of transgenic plants. This guidance encompasses many of the net fitness parameters necessary for evaluating risk of transgene spread and therefore could serve to populate a screening model. Further clarification by the regulatory community of the nature, sequence, and rationale for data generation (within the context of net fitness) would serve to guide data generators as to appropriate levels and timing of data generation.

The net fitness model used within the screening context is an obvious filter for a host of scenarios. The model itself can help to answer how much uncertainty in input parameters needs to be reduced in order that regulators may better discriminate a moderate-high level concern from a low level concern. Thus, the context-dependent aspects of transgenic plant deployment within varied environments (and management strategies) could be modeled through scenario building. In this way the variance in predicted outcomes can be described. For a transgenic plant with a high to moderate risk of spread leading to a potential adverse effect, perhaps scenario-specific cases may be identified that are low risk, allowing for deployment within a given set of environmental management constraints while eliminating others.

Uncertainty analysis for the stochastic implementation of the screening level net fitness model would provide insight into what additional information would best reduce uncertainty to manageable levels. Stepwise approaches to data generation and assessment may be amenable to formalized approaches such as sequential decision theory, which could be used to encourage data generators and regulatory decision-makers to consider development of additional data to clarify uncertainties. The desirability of further data generation needs to consider the time, cost, and regulatory burden of generating the data as well as the value of the data to the risk assessment and decision-making process.

Adapting the net fitness model for plants

In using the net fitness model for plants, there is the opportunity for significant GxExM interactions requiring consideration of the spatial implications of transgene entry into the environment. The nature of the transgenic crop deployment will govern to some degree the environmental effects on net fitness (for instance, due to the relative size of the transgenic source population compared to the non-transgenic receiving population, size of buffers, or spatio-temporal segregation). Both annual and perennial crops have important but different responses within the environment, therefore different approaches to spatially explicit

models may need to be considered for a given application of the net fitness model.

In addition to the need to incorporate considerations of environmental context, the migration-selection component of the integrated risk assessment model needs to be built. Addressing the subtleties of gene linkages and the associated within and between generation shifts that may occur is necessary to determine the various local optima within the evolutionary landscape that might be possible for a given suite of net fitness components.

Because of the case and context specific nature of the outcomes for net fitness modeling, the modeling approach will need to accommodate alternative data structures for description of fitness components. For instance, fecundity will be differentially represented for the case of a perennial tree with significant seed dispersal versus an annual crop where seeds are harvested. Different scenarios for transgene spread will arise dependent on the plant and the envisioned management. For instance, for the example of transgenic turfgrass, evaluation is keyed to the intended management scenario (intensively managed golf greens where gene transmission is by way of vegetative propagation). A secondary scenario, which considers seed production and pollen flow when a golf course is abandoned and goes to seed, is important as well. In this case, there will be differential fitness components for the clonal material (ramens) versus for crop-wild hybrids that may arise from out-crossing.

Data needs for net fitness modeling

As the foregoing examples point out, data needs will differ depending on the case and scenario modeled. For a clonally propagated plant, data on clonal rates of reproduction must be available. For certain perennials the size and nature of the seed bank must be described (for instance, for certain tree species there is a need for life history tables of reproductive efforts and mortality rates by age). For a transgenic plant that out-crosses to a wild relative, net fitness components specific to the crop-wild hybrid may be necessary.

The Canadian-US bilateral agreement on ecological data is relatively comprehensive but needs further clarification on the sequence in which data should be developed and the case specific needs. Model input must be able to capture mating success (for example, the difference in flowering time between source and receptor species, dispersal of receptors spatially, and out-crossing frequency in natural settings). The way that data are developed to address gene

flow to wild versus feral needs to be clarified. Data describing GxE is rather sparse and there is a need to develop case studies to elaborate the level of detail that may be required to address environmental interactions. This may be especially true for genes affecting plant physiology and responses to biotic and abiotic stresses.

Differing needs for modeling crop-crop versus crop-native plant gene flow

Workshop discussion centered on transgene spread into nature (that is, into unmanaged or low intensity managed ecosystems versus the agroecosystem). There is a family of related issues regarding risk of crop-to-crop gene flow that was not explicitly addressed by the Workshop. For agronomic crops, gene flow is largely managed rather than assessed (as through use of physical and biological controls to confine pollen movement and restrict out-crossing potential). Clearly, however, there is significant overlap in the data and assessment tools that are applicable to both crop-to-crop and crop-to-wild gene flow considerations. First tier assessments of transgene escape have obvious relevance to both natural and agroecosystems that go beyond transgene spread. For instance, in order to assess risks to non-target organisms or of crop contamination, data are needed to describe transgene escape. If gene transmission occurs to the breeder's seed pool (as in the case of StarLink® corn), then assessment of transgene introgression becomes an important consideration for designing gene recall strategies. Additionally, questions concerning plant invasiveness may be recast to include management practice and thus address weediness.

Caution is needed in the degree to which distinctions are made between gene flow to wild/native populations versus gene flow in agricultural landscapes, because the agroecosystems have their own diversity that needs to be managed. As certain transgenes (such as those from Bt) become commoditized over time, there will be utility in the use of net fitness modeling and linked tools to better understand breeding strategies for managing transgene distribution through the breeder's seed pool.

Model refinement and verification

A pressing need identified at the Workshop relates to the appropriate means to calibrate and validate a refined net fitness model. Model validation, at some level, constitutes a necessary first step before proceeding to development and use of a screening level model for regulatory risk assessments. Because the net fitness model looks at transgene spread in an evolutionary timeframe, validation in its fullest



sense is highly unlikely. The ability of the model to predict changes in gene frequency in the model organism *Drosophila* in laboratory environments has been demonstrated by Prout (1971). However, testing in more complicated biological systems has not been attempted. Workshop discussion centered on experimental systems and existing datasets that are most able to approach the needs for model calibration and/or validation.

As Muir and Howard have done with transgenic fish, it may be possible to use a rapidly reproducing species to develop data over many generations. *Arabidopsis* appears a logical choice from this standpoint, as the plant has a short life cycle, its genetics are well understood, and transgenics are developed readily. A glasshouse with *Arabidopsis* could be modeled as a TPE where environmental zones and spatial structure are present and net fitness could be considered within the content of GxE. A potential limitation of *Arabidopsis* is self-incompatibility; however this could be made into an advantage in the way the experiments are conceptualized. Although *Caenorhabditis elegans* or *Drosophila* are potentially effective model systems, they do not represent the total range of variables for a plant and do not constitute a useful, practical model that would allow regulators to have high certainty in the application of net fitness approaches to higher plants.

Fundamental data of use for model evaluation may be available for antagonistic gene models such as insecticide resistance, sickle cell, or other major gene effects. Antagonism may disappear once the background disappears, so these data may be useful with respect to determining the degree to which major gene/polygene concepts may apply to the spread of a transgene into a new background.

Existing data on invasive plant species will prove useful to considering high level aspects of the net fitness model with the caveat that direct comparison of invasive genomes containing tens of thousands of genes to a single gene may not be appropriate. In one case, the difference is polygenic, and in the other, monogenic. The impact of gene number on risk assessment is not clear. Datasets detailing the six net fitness components may be available and the net fitness approach also may be quite useful to invasive plant considerations. It may be possible to obtain some estimate of the genetic correlates of invasiveness across diverse environments. There is a need to be able to tap into large datasets at some point, but initially one can use model sensitivity analysis to determine the degree to which environments may drive the outcome. A more appropriate model system would entail an invasive species that out-crosses with a

wild relative in the receiving environment. Data for this system would closely approximate the case of crop-to-wild transgene flow.

It may be possible to verify the net fitness model with existing data on closely related crop and wild species occupying similar TPE but with differential environmental responses. Considerable data should be available for sugar beet and sea beet in Europe; canola (*Brassica napus*) and related weed species (field mustard (*B. rapa* L.), wild mustard (*B. kaber* (DC) L.C. Wheeler), and black mustard (*B. nigra* (L) W.J.D. Koch)) in Europe and North America; and cultivated and wild sunflower species in North America (see Stewart *et al.*, 2003).

Furthermore it may be possible to obtain (or develop) data describing the behavior over initial generations of transgene-wild crosses. As described earlier, the regulatory context of the Canadian-US bilateral agreement on ecological data provides a mechanism whereby such data may be generated as part of the deregulation petition for transgenic plants. Databases exist for transgenic *Brassica* and cotton and perhaps afford a useful first case for net fitness modeling. In addition, a number of transgenic turf grasses are in the development pipeline and therefore net fitness modeling of grasses is a possibility.

As a first instance of considering either invasive or crop-wild models, it may be possible to place life history events into the model and determine if the model predicts outcomes that have occurred in nature. More refined model verification requires data that documents what happens in the crop-wild cross (or the invasive-noninvasive cross or the transgenic-wild hybrid) over generations. In considering a transgene introduced into a receiving population, there needs to be some understanding of how many generations are needed to remove the donor background sufficiently to see the effect of transgene flow into the receiving population. Consideration must also be given to what happens to the neutral fitness factors over time, with respect to movement from crop-to-wild, and how this relates to the evaluation of the transgene that may be introduced crop-to-wild. There was agreement by Workshop participants that if data were available describing changes in any of these various hybrids through the early backcross generations to the wild parent, this might be usefully evaluated by the net fitness model. A slight degree of change from the BC2 to BC3 would lessen need for further data development; whereas, a significant degree of change from BC2 to BC3 would necessitate obtaining data on later generations in order to increase confidence in assessments using the net fitness approach.

Even with incomplete information, the net fitness model allows us to ask whether we have enough information or whether additional data may be required. Regardless, a lingering uncertainty will exist as to whether greenhouse or small field trials through early generations will be sufficient to measure the fitness components for transgene entry into the environment. The greater the number of diverse cases that can be used for verification, the greater will be the certainty of model predictiveness.

Toward an integrated risk assessment model for plant transgene spread

The Environmental Risk Assessment Modeling Workshop II focused on the use of an integrated net fitness modeling approach for assessment of transgene spread into natural environments. There was clear recognition that alternative needs (crop-to-crop gene flow) and methodologies (spatially-explicit probabilistic modeling) must be considered in the comprehensive evaluation of risks posed by transgenic plant deployment. Modification of the net fitness model of Muir and Howard to include an evaluation of transgene spread as affected by correlated genes interacting over environments within the background of natural selection could be a useful component of a modular system for risk assessment of transgenic organisms. Verification of the integrated model reliability for predicting transgene spread could enable the development of screening tools that effectively gauge the relative risks posed by transgenic plant constructs under a host of environmental scenarios. The availability of such a tool will provide transparent, quantitative rationale for the nature of data that need to be developed to support regulatory decision-making in the area of gene flow.

Within an environment where there are competing research needs for transgenics, there is a continuing need for appropriate resource prioritization. An integrated net fitness model evaluating relevant cases should use cost/benefit analysis to support whether data generation to reduce uncertainties regarding transgene spread into natural environments should be a public research priority. Probably the greatest current impact of the net fitness approach relates to transgenics in trees and grasses. These technologies tend to involve fewer companies with fewer resources; therefore, there is a need for a linkage of public and private enterprise to assure that there are no unmet needs with respect to responsibly advancing new innovations in plant biotechnology. Transgenes have been deployed in *Brassica* and cotton for many years; therefore, there are fairly rich experimental and monitoring data that make these crops valuable cases for consideration. Link-

age of the genetic thrust of research to encompass ecological consequences and context will better benefit regulatory decision-making and should attract greater support within the research community.

Workshop Participants:

Carol A. Auer
Associate Professor
Department of Plant Science, U-163
Agricultural Biotechnology Lab, Room 302C
University of Connecticut
Storrs, Connecticut 06269
Carol.Auer@uconn.edu

Mark Baird
District Sales Manager, Pioneer Hi-Bred International, Inc.
Liberal, Kansas
Mark.Baird@pioneer.com

Mark Cooper
Pioneer Hi-Bred International Inc.,
7300 N.W. 62nd Avenue, P.O. Box 1004
Johnston, Iowa 50131
mark.cooper@pioneer.com

Ruth Irwin
Editor, ISB News Report
207 Engel Hall
Virginia Tech University
Blacksburg, VA 24061
rirwin@vt.edu

Susan Koehler
Biotechnologist
USDA-APHIS
4700 River Road-Unit 133
Riverdale, MD 20737-1237
Susan.M.Koehler@aphis.usda.gov

Nick Linacre
Environmental Science, School of Botany
The University of Melbourne
n.linacre@pgrad.unimelb.edu.au

Bill Muir
Department of Animal Sciences
915 W. State Street
Purdue University
W. Lafayette IN 47907-2054
bmuir@purdue.edu



Guilherme J. M. Rosa
 Department of Animal Sciences & Wildlife
 12051 Anthony
 Michigan State University
 East Lansing, MI 48824-1225
 rosag@msu.edu

Neal Stewart
 Racheff Chair of Excellence in Plant Molecular Genetics and
 Professor.
 Department of Plant Sciences and Landscape Systems
 2431 Center Dr.
 Univ of Tennessee
 Knoxville, TN
 nealstewart@utk.edu

Dave Walker
 Sales Representative, Pioneer Hi-Bred International, Inc.
 Johnson, Kansas 67855
 DTWalker@PLD.com

Bruce Walsh
 Department of Ecology and Evolutionary Biology
 University of Arizona
 Tuscon, AZ 85721
 jbwalsh@u.arizona.edu

Arthur E. Weis
 Professor of Ecology and Evolutionary Biology
 Univ. California
 Irvine, CA 92697 USA
 aeweis@uci.edu

Claire Williams
 Duke Univ., Nicholas School of the Environment
 Box 90658
 Durham NC 27708
 claire-williams@fulbrightweb.org

LaReesa Wolfenbarger
 Adjunct Associate Professor
 Department of Biology
 Allwine Hall
 Omaha, NE 68182-0040
 lwolfenbarger@mail.unomaha.edu

Jeff Wolt
 Risk Assessment Leader
 Dow AgroSciences Global Exposure and
 Risk Assessment Bldg. 306, C-2/839
 Indianapolis, IN 46268-1053
 jdwt@ dow.com

References

- Cooper M and Podlich DW. (2002) The E(NK) model: Extending the NK model to incorporate gene-by-environment interactions and epistasis for diploid genomes. *Complexity*. Wiley Periodicals, Inc. Vol. 7 No 6.
- Falconer DS. (1961) *Introduction to Quantitative Genetics*. Oliver and Boyd: Edinburgh.
- Gelman A, Carlin JB, Stern HS, and Rubin DB. (2003) *Bayesian Data Analysis*, 2nd Ed. Chapman & Hall/CRC: New York.
- Hallerman EM. (2002) ISB Workshop suggests strengthening and broadening of net fitness model. *ISB News Report*, August 2002, <http://www.isb.vt.edu/news/2002/Aug02.pdf>.
- Kauffman SA. (1993) *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press: New York.
- Lande R. (1983) The response to selection of major and minor mutations affecting a metrical trait. *Heredity* 50: 47-65.
- Muir WM and Howard RD. (1999) Possible ecological risks of transgenic organism release when transgenes affect mating success: sexual selection and the Trojan gene hypothesis. *Proceedings of the National Academy of Sciences* 24: 13853-13856.
- Muir WM and Howard RD. (2001) Fitness components and ecological risk of transgenic release: A model using Japanese medaka (*Oryzias latipes*). *American Naturalist* 158: 1-16.
- Muir WM and Howard RD. (2002) Environmental risk assessment of transgenic fish with implications for other diploid organisms. *Transgenic Research* 11: 101-114.
- NRC (National Research Council). 1983. *Risk Assessment in the Federal Government: Managing the Process*. National Academy Press: Washington, DC.
- Prout T. (1971) The relation between fitness components and population prediction. *Genetics* 68: 127-149.
- Rejamek M and Richardson DM. (1996) What attributes make some plant species more invasive? *Ecology* 77: 1655-1661.
- Robert CP and Casella G. (1999) *Monte Carlo Statistical Methods*. Springer: New York.
- Shoemaker JS, Painter IS, and Weir BS. (1999) Bayesian statistics in genetics: a guide for the uninitiated. *Trends in Genetics* 15: 354-358.
- Stewart CN Jr, Halfhill MD, and Warwick SI. (2003) Transgene introgression from genetically modified crops and their wild relatives. *Nature Reviews Genetics* 4: 806-817.

**ISB News Report
207 Engel Hall
Virginia Tech
Blacksburg, VA 24061**

Non-Profit Org.
U.S. Postage
PAID
Blacksburg, VA 24060
Permit No. 28