Optimal Search in *Ex situ* Collections of Wheat Genetic Resources

Douglas Gollin, Melinda Smale, and Bent Skovmand

CIMMYT

Printed in Mexico.

Contents

Abstract

This paper develops a theoretical model for analyzing gene bank management decisions regarding the search for traits of economic value in *ex situ* collections of wheat. The model is applied to data on the probability of finding useful sources of resistance to Russian wheat aphid (*Diuraphis noxia*) and septoria tritici leaf blotch, using Monte Carlo simulations for sampling distributions, simulations of varietal diffusion paths, and actual cost data from searches. Three specific questions are posed and answered: (1) what is the optimal size of search among genetic resources of a given type for a trait of economic value? (2) what is the value of specialized knowledge about which genetic resources are most likely to display resistance? and (3) how should search resources be allocated across types of genetic resources? Results demonstrate the sensitivity of the optimal size of search to the economic importance of the problem, the probability distributions of the trait in different types of genetic resources, and the costs and time lags associated with transferring the trait into usable breeding material. The costs and time lags involved with conventional pre-breeding techniques imply that in some searches, certain categories of genetic resources (such as landraces) will be systematically ignored. The fact that they may be rarely utilized does not imply that collections of landraces have no value, however, as shown in the case of Russian wheat aphid. Though applied here to data on insect and disease resistance, the model can be adapted to search decisions for other types of traits.

Tables

Figures

Optimal Search in *Ex situ* Collections of Wheat Genetic Resources

Douglas Gollin, Melinda Smale, and Bent Skovmand

Introduction

Much public and scientific attention in recent years has focused on the protection and conservation of genetic resources. The threats posed to a handful of wildlife species—such as elephants and condors—have attracted the greatest attention, along with the loss of pristine habitats, such as rainforests and oceans. Increasingly, however, genetic diversity for agriculture has surfaced as a cause for public concern. Questions have been raised about the adequacy and management of *ex situ* collections of crop germplasm.

There are two principal strategies for preserving genetic resources for crop species, which are increasingly viewed as complements rather than as exclusive alternatives. One strategy is to preserve different varieties or species *in situ*. For cultivated species, *in situ* conservation, as currently understood, implies the management of traditional varieties and wild materials by farmers in centers where the crop has evolved (Maxted et al. 1997). The other strategy is to preserve seeds or other propagative materials *ex situ*, in a collection that is physically separated from the environment of origin. Historically, *ex situ* collections of economically important crops date back hundreds of years. Early collections consisted of botanical gardens or simple fields in which different plant varieties were maintained through cultivation. Today, *ex situ* collections are more technologically sophisticated. For many crops, present-day gene banks are essentially huge refrigerators designed to store samples of seed under low temperature and humidity for long periods of time.

For most cultivated species, extensive *ex situ* collections already exist. In wheat, the Food and Agriculture Organization (FAO) of the United Nations estimates that 95% of wheat landraces and 60% of wild species have been collected, although the origin of these figures is unclear. Similarly high figures are reported for other major crops, such as rice and maize, although wild species coverage is lower for these crops (FAO 1996). Since the 1970s, many national and international research programs have upgraded their gene banks, but Wright (1997) argues that the utilization of these resources has not kept pace with their expansion.

The issues investigated in this paper concern the economics of gene bank utilization. What is the optimal "size" of a search for useful materials? Do large collections ever have value? In a collection of 35,000 landraces, are marginal accessions likely to be redundant for all practical purposes, or might they be valuable under some circumstances? Is low utilization by breeding programs of existing gene banks an indication of inefficiency or excessive capacity? Are breeders and other scientists wastefully ignoring the materials in gene banks?

To answer these questions, we propose a relatively simple statistical framework developed from search theory. We apply the model to data on resistance to Russian wheat aphid (RWA) (*Diuraphis noxia*) and septoria tritici leaf blotch, using frequency distributions

1

obtained from the Genetic Resources Information Network (GRIN)[1] and CIMMYT scientists. The frequency distributions are used to estimate "smooth" representations of the hypothetical underlying distributions for resistance. We combine this information with estimates of the benefits that can be attained from finding resistant materials and information on the cost of searching for resistant materials, both from CIMMYT databases. The model then guides us in identifying optimal strategies for searching the gene bank and for utilizing different materials. We use Monte Carlo simulations to solve the model computationally.

Three specific questions are posed and answered in the analysis. First, what is the optimum number of gene bank accessions that should be included in a search for a useful trait, given a category of genetic resources? Second, what is the value of specialized knowledge about the distribution of desirable traits across types of germplasm? In this case, we compute the value of knowing that resistance to RWA is more common among a set of bread wheat (*Triticum aestivum*) landraces from Iran than among the general population of bread wheats. Third, we ask how should an optimal search proceed when scientists can hunt for a particular trait among two distinct populations with differing distributions of resistance (using the case of septoria tritici leaf blotch as an illustration)?

The next section discusses briefly two search models that economists have previously applied to the analysis of genetic resources. Our search theoretic framework is then presented, followed by a description of the computational methods used to apply it. Results of the experiments are then presented and discussed. Limitations of the method, implications, and further research directions are provided in the final section.

Related Search Models

There is little confusion or disagreement over which concepts to apply in economic analysis of genetic resources; simple marginal analysis is appropriate (Brown 1990; Evenson 1993; Pearce and Moran 1994; Swanson 1995). Genetic resources should be collected so long as the marginal expected benefits exceed the marginal costs of collection and storage. Searches of a collection should proceed until the marginal expected value of success is outweighed by the search cost.

Evenson and Kislev (1975) used a simple search model to illustrate the economics of agricultural research. In their model, new crop varieties were "discovered" through a costly search (research) process. Using data on sugarcane research, they answered questions about optimal search strategies, investigated the effects of changes in research technology, and explored other policy issues. A general implication of their model was that additional investments in research would be likely to pay off, either by "discovering" improved varieties or reducing search costs.

---

[1] Genetic Resources Information Network (GRIN) of the U.S. National Small Grains Collection, U.S. Department of Agriculture, Agricultural Research Service.

More recently, Simpson et al. (1996) applied a search model to the problem of valuing a marginal species in a tropical rainforest.[2] In that model, pharmaceutical researchers test a large number of species for a particular trait of economic value. The trait is assumed to be distributed randomly and uniformly across the entire population; with a given probability, each species either possesses the trait or fails to possess the trait. The search process consists of repeated and independent Bernoulli trials. The authors arrive at the striking conclusion that under most plausible specifications, the expected value of a marginal species is minuscule: "*regardless* of the probability with which the discovery of a commercially useful compound may be made, if the set of organisms that may be sampled is very large, the value of the marginal species must be very small" (emphasis in the original).

Simpson et al. thus take a relatively pessimistic view towards the marginal value of genetic resources, in contrast to the generally optimistic view of Evenson and Kislev. This essential difference indicates the sensitivity of search theoretic models to important assumptions concerning the *redundancy* of materials in a collection of objects to be searched and to empirical issues about the "rareness" of valuable traits.

Suppose that a desirable trait is found with some distribution within a particular population and that it is equally useful wherever it is found. For example, assume we are seeking flowers that are yellow and that any shade of yellow is equally useful. In this case, there is perfect substitutability among the subpopulation of species that possesses the trait (all yellow flowers) and zero substitutability with the remainder of the population. Once a single yellow flower is found, all further search becomes redundant. This is an assumption in Simpson et al. (1996).

Alternatively, suppose that the desired trait can be found in varying intensities or forms, so that it can be conceptualized as a continuous variable. For example, gold might be found in different deposits of ore at varying degrees of concentration, or a number of different plants might have antibacterial properties of varying usefulness. In this case, there is imperfect substitutability among materials in the population. Different materials are more or less desirable, according to some index, and a distribution of "desirability" is found across the population. Also, additional search is always expected to offer some marginal benefit unless an extreme value has been obtained. This is the result obtained by Evenson and Kislev (1975) and reiterated in Evenson (1996).

In addition to this theoretical difference between the two models, there are underlying differences in their views concerning the distribution of useful traits and the costs and benefits of search. Simpson et al. note that their results are sensitive to the relative magnitudes of costs and benefits as well as to the distribution of a trait within a population.

To answer questions about searching *ex situ* collections of crop genetic resources, a model is needed that does not arbitrarily impose any assumption about the nature or distribution of desirable traits. Data is also required on the actual distributions of traits within the relevant

---

[2] In a similar paper, Simpson and Sedjo (1996) apply the same general model to conserving genetic resources for agriculture.

population and on the costs and benefits of searching for such traits. Within the literatures on agricultural science and applied genetics, relatively little recent work has looked at the overall variation in traits across large populations of crop varieties. Recent studies focus on molecular differences among varieties and compare small segments of DNA (Autrique et al. 1996; Tsegaye et al. 1994; Pecetti and Damania 1996; Chen et al. 1994), although molecular analysis of genetic diversity in wheat has not yet advanced to the point where it can tell us much about the distributions of rare alleles within large populations of wheat varieties. While Tanksley and McCouch (1997) have argued that phenotypic evaluation to determine the breeding value of an accession is likely to be misleading (especially with respect to quantitative characteristics), phenotypic evaluation is the best tool now available. The following model and analysis are most relevant to searches involving phenotypic analysis of economically useful traits.

A Search Theoretic Model
of an *Ex situ* Collection of Crop Genetic Resources

Searching for Resistance or Tolerance: Minimum Values
Many traits of economic importance are found in varying degrees within a population of varieties or within a category of a crop's genetic resources. For many descriptive traits, such as plant height or leaf thickness, the distribution across the population approximates normal. For other traits, the distribution within the population may follow another pattern. With diseases and pathogens, many populations may display little resistance, while the remainder display varying degrees of resistance.

Typically, scientists score varieties on a three-, five-, six-, or nine-point scale in evaluation and characterization trials. Let s denote the score achieved by a variety, where $s \in S$. For simplicity, let $S = [0, 1]$. In keeping with common agricultural practice, let low scores be considered "better" than high scores; in other words, a variety with $s = 0$ is superior to one with $s = 1$.

Let the distribution of varieties within subpopulation j be denoted by $\Delta^j(S)$, where $j \in J$ denotes a particular subpopulation of varieties from the set J of possible subpopulations. For example, we might be interested in differences in distributions across landraces (LR), obsolete lines (OB), synthetics (SY), and elite lines (EL). Then,

(1) $J...\{LR,OB,SY,EL\}$.

Suppose that n draws are taken from the distribution $\Delta^j(S)$. Let y_1 be the score achieved on the ith draw from this distribution. Then define the order statistic Z_n^j as follows:

(2) $Z_n^j = \min[y_1, y_2, ..., y_n]$.

The *minimum value*, in most scoring systems, is the most resistant—and hence most desirable material.[3]

Note that the expected minimum value depends on the number of materials drawn from the distribution. Given a distribution of scores, $\Delta^j(S)$, it is a straightforward process to derive an expression for the distribution of the minimum value.

Let z be a particular score, such that $z \in S$. Then the probability that the minimum value attained in n draws from the distribution exceeds z is given by:

(3) $\quad \Pr[y_1 \geq z; y_2 \geq z; \ldots y_n \geq z].$

Let $f_j(s)$ denote the probability density function for the distribution $\Delta^j(S)$, and let $F_j(s)$ denote the cumulative density function for this distribution. Then the probability that $y_1 \geq z$ can be written, for all i, as $[1 - F_j(z)]$. Assuming that the n draws are independent and identically distributed, the probability that all the draws will exceed z can be written simply as:

(4) $\quad \Pr\{z \leq \min [y_1, y_2, \ldots, y_{n,}] \} = [1 - F_j(z)]^n.$

We can think of the probability that z will be a minimum value from a given number of draws, n, as being related to a cumulative density function for z. Let $H_n^j(z)$ designate this cumulative density function; thus,

(5) $\quad H_n^j(z) = \Pr\{z \geq \min [y_1, y_2, \ldots, y_{n'}]\}$

It follows directly from equation (5) that $H_n^j(z) = 1 - [1 - F_j(z)]^n$. From this cumulative density function, we can also derive a probability density function for z. Let $h_n^j(z)$ represent the probability that z is the minimum value attained in n draws from the distribution $\Delta^j(S)$. Then it follows that $h_n^j(z)$ can be derived from $H_n^j(z)$, so that:

(6) $\quad h_n^j(z) = n[1 - F_j(z)]^{l-1} f_j(z).$

Note that $h_n^j(z)$ is defined for all possible values $z \in S$. This function gives the probability that z will be the minimum value attained from n draws from the distribution $\Delta^j(S)$. Also note that we can derive an expression for $h_n^j(z)$ simply from information about the underlying distribution $f_j(z)$ and the number of draws made from subpopulation j. Note that $h_n^j(z)$ is a well-behaved distribution. In particular, it has a well-defined mean and

---

[3]   The most resistant materials are not always the ones with the greatest value to breeders. In some instances, breeders may find that varieties with slightly higher susceptibility to a stress are, nonetheless, effective sources of useful traits. Moreover, breeders may occasionally opt for varieties that display moderate susceptibility to a problem, rather than those that are fully resistant, as part of a dynamic breeding strategy in which they seek to limit future evolution of disease and pest biotypes. For example, CIMMYT breeders have for many years sought to breed wheat varieties that display moderate susceptibility to rusts, which are common fungal diseases. For the purposes of this model, however, it is a reasonable and useful generalization to imagine that breeders are seeking the "best" materials by their defined criteria.

variance. Assuming that all the functions are continuous and integrability conditions apply, we get the expected value:

(7) $\quad E[z_j(n)] = \int\limits_{s \in S} s \, h_n^j(s) \, ds = \int\limits_{s \in S} s \, n \, [1 - F_j(s)]^{l-1} \, f_j(s) \, ds.$

Note that the expected value of this minimum varies with n. The more draws that are taken from the distribution, the smaller the expected value of z becomes. The variance of this distribution is also well defined.

For some traits the actual minimum score attained may be of interest. For other traits, it may matter only whether some threshold value is obtained. Either type of trait can be represented with this model; thus, the Bernoulli case considered by Simpson et al. (1996) can be viewed as a special case of this more general framework. For the Bernoulli case, we simply define some threshold of usefulness, saying that any variety with a resistance score below s^* is useful, and anything with a score above s^* is not useful. The Bernoulli probability, p, is thus the probability that a variety has a resistance score below s^*, or in other words,

(8) $\quad p = F_j(s^*)$

The probability that a "useful" variety will be found in a search of size n is given by $H_n^j(s^*)$, which reduces here to $1 - (1 - p)^n$.

The Value of Successful Search
Suppose that world agriculture can be divided into a set M of agroclimatically similar growing "mega-environments." CIMMYT, for example, classifies the world bread wheat growing areas into "mega-environments" that span different continents and regions within the developing world. CIMMYT views these mega-environments as distinct, from the point of view of technology development. Suppose that a particular mega-environment, m, has an average score of $\bar{z}$ for a given trait. The value of production given this level of the trait is $V_m(\bar{z})$.[4] If instead agriculture in this mega-environment obtained a higher trait score of $\hat{z}$, the single-period gain would be given by $v_m(\hat{z}) - v_m(\bar{z})$. Typically, plant breeders would expect that for some traits, the productivity advantage would fade over time through a kind of depreciation. For example, resistance to pests and diseases typically breaks down over a period of years as new biotypes of pathogens emerge. In contrast, resistance to abiotic stresses (such as drought or salinity) is maintained for longer time periods or may last indefinitely in a fixed environment. Let $\delta(t)$ represent the depreciation path for a particular trait, and let it be considered as a known parameter of the model. In addition, assume that the discount factor β corresponds to the rate of time preference. Then the returns at date t obtained from achieving the score $\hat{z}$ today can be written as:

(9) $\quad \sum\limits_{m \in M} \beta^t \delta(t) [v_m(\hat{z}) - v_m(\bar{z})].$

---

[4] Although we do not make any explicit assumptions about the form of v at this stage, we would not generally expect that v is linear in z. In other words, the value of achieving a particular degree of resistance is not linear in the degree of resistance. Usually we would expect that v is increasing but concave with respect to z.

Now consider how the timing of the search process works for plant breeding. The initial search for resistance takes place at time $t = 0$. Suppose that this is an instantaneous event. Two types of costs are incurred at date zero. The first is the fixed cost of searching among materials of type j. This cost, C_j, is incurred at date zero, and it may include the costs of developing search techniques or setting up a new experiment. The cost might differ across the type of material if, for example, different techniques are used to screen for resistance among wild species. For instance, tall and late-maturing landrace materials are harder to screen for septoria tritici leaf blotch than short, early maturing elite lines, because their growth pattern normally precludes exposure to the disease. Thus, screening landrace materials for *Septoria tritici* resistance would require a more complicated—and more expensive—technique than screening elite lines. The second relevant category of costs is the variable cost c_j, which is the cost of screening one more variety of type j. These costs may also differ across types of material if, for example, molecular methods are needed to screen one type of material and conventional screenhouse techniques can be used to screen another type.

Upon completion of the search, but still at date zero, scientists observe the best material of each type and select one to use in breeding. There is a time lag, however, associated with transferring resistance genes from materials in the gene bank into a breeding program, which varies by type of material and by the trait.[5] Let T_j denote the time lag from discovery until a material of type j can be used by a breeding program. Once the material has reached the breeding program, there is a second time lag associated with developing a usable variety. This time lag differs with the target mega-environment. For favorable areas, varieties may be developed more quickly than for marginal environments. Further, because of other factors related to adoption of new varieties by farmers, the diffusion of materials carrying resistance may begin later in some areas than in others. Let T_m denote the time lag from the entry of the material into the breeding program until the beginning of its diffusion in a given mega-environment. Finally, let τ denote the relevant time horizon to breeders.

Finally, there is a transfer cost, K_j, that is incurred in transferring resistance from a material of type j into a breeding program. This cost reflects the differing techniques that may be needed to incorporate desired traits into usable breeding materials. These are "pre-breeding" costs.

With these definitions in place, consider the costs and benefits of search. Let $R_j(n_j)$ denote the benefit stream from using the best material found in a search of n_j accessions of type j. These benefits are given by:

10) $$R_j(n_j) = \sum_{m \in M} \sum_{t=T_j+T_m}^{T_j+\bar{T}_m+\tau} \beta^t \delta(t)[v_m(\hat{z}_j) - v_m(\bar{z})].$$

Note that $R_j(n_j)$ is a distribution over returns. Depending on the actual draw of n_j materials, the returns will vary. Let $F_{R_j}^{n_j}(y)$ denote the cumulative density function for the distribution $R_j(n_j)$. In other words, $F_{R_j}^{n_j}(y)$ represents the probability that $R_j(n_j) \leq y$.

---

[5] For example, some traits are highly heritable; this allows them to be transferred more quickly than traits that are less readily heritable. It is also possible that in some cases, the time lag may be so great that a pathotype will have already evolved!

The researcher's problem is to select the material that gives the highest overall returns from the realization of a vector of draws, n, where $n = (n_1, n_2 \ldots n_j)$. The researcher will choose the maximum value from J independent draws from J different distributions:

(11) $\quad R^*(n) = \max[R_1(n_1)\text{-}K_1, R_2(n_2)\text{-}K_2, \ldots, R_j(n_j)\text{-}K_j]$.

Because $R_j(n_j)$ is a distribution, the researcher's total returns $R^*(n)$ are also a distribution. As above, we are interested in an order statistic—the maximum return attained through drawing once from each of the distributions $R_j(n_j)$. Thus, we can define a cumulative density function for overall returns. This is given as $F_{R^*}^n(y)$, where:

(12) $\quad F_{R^*}^n(y) = \Pr[R^*(n) \le y] = \Pr[R_J(n_J) \le y, \ \forall j \in j] = \prod_{j=1}^{j} F_{R_j}^{n_j}(y).$

We can also define a probability density function $f_{R^*}^n(y)$ that is derived from $F_{R^*}^n(y)$. Assuming that $f_{R^*}^n(y)$ is integrable, the expected value of the researcher's returns can be written as $E[R^*(n)]$, where:

(13) $\quad E[R^*(n)] = \int_0^\infty y \, f_{R^*}^n(y) \, dy.$

Since the distributions $F_{R^*}^n(y)$. are different from one another, we cannot easily derive a closed form expression for $f_{R^*}^n(y)$, but it can be approximated computationally by differencing $F_{R^*}^n(y)$.

Optimal Search Strategies
Now suppose that researchers face an optimization problem of the following kind. They seek to sample n materials in a way that maximizes expected returns subject to a research budget constraint. Suppose for simplicity that researchers are risk-neutral; in other words, their preferences are linear in the expected return obtained. The researcher's problem can then be written in a straightforward fashion as:

(14) $\quad \max_n E[R^*(n)]$
$\qquad s.t. \ \sum_{j \in J}(C_j + c_j n_j)I_j \le B$

$\qquad \text{where } I_j = \begin{cases} 1 \text{ if } n_j > 0 \\ 0 \text{ if } n_j = 0 \end{cases}$

and where B represents the total budget available for search, c denotes the variable costs associated with screening materials, as described above, and C denotes the fixed costs of search, also described above.

This problem is difficult to solve using standard optimization techniques, and without closed form expressions for the distributions of interest, it is difficult to arrive at an analytical solution. Instead, we must rely on computational techniques.

Solution Approach

One way to solve the problem computationally would be to evaluate the expected returns from all the possible vectors n that exhaust the research budget. For a small number J of subcollections, this may be a reasonable strategy given modern computational techniques. Analytically, we might be interested in expected returns and also the distribution of returns associated with a particular search strategy. These can be estimated through Monte Carlo simulations, in which we repeatedly choose n^* materials at random and observe the actual returns obtained. With a sufficiently large number of iterations, we can obtain good estimates for the underlying distribution. In the experiments described herein, we equated the marginal benefits to marginal costs of search, using: (1) the probability distributions for "useful" resistance generated with Monte Carlo simulations from smoothed, actual distributions for the trait; (2) estimates of benefit streams; and (3) representative cost data. Methods used to develop each of these are summarized below and described in greater detail in Appendix 1.

Probability distributions for traits. The population of wheat accessions found in national and international collections can be thought of as a (large) draw from the distribution of wheat populations found worldwide. The subpopulations evaluated for particular disease or pest problems are then second-stage samples from the gene banks. Given an actual discrete distribution of scores for disease or pest resistance over any subcollection of wheat varieties, we can calculate the distribution $h_n^j(z)$, which gives the probability that a given minimum value will be obtained from n draws from distribution j.

The GRIN provides data on discrete distributions for certain traits within the population of wheat varieties, taken from performance scores in various agronomic trials, represented in scales. These scales are essentially imposed for convenience by agricultural scientists on an underlying, "true" distribution that has been reconstructed. A least squares technique was used to fit a smooth, beta distribution to each discrete distribution of disease and pest resistance scores.[6] The parameters of a beta distribution are generally written as a, b, and β. Only two of the parameters are independently determined; the third can be computed from the other two. Here, we searched for the value of (a, b) that minimized the sum of squared residuals when the discrete distribution was compared with the smoothed distribution. The beta distributions can take on a wide range of shapes and forms, and can be used with data that range from approximately symmetric to severely skewed. Beta distributions can also allow for tails of varying thickness, which is important for the types of analysis undertaken here. Figures 1, 2, and 5 illustrate some of the fitted distributions for specific traits. The beta distributions offer good representations of the data, although the "goodness of fit" varies.

Benefits. We estimated the areas affected by diseases, average annual yield losses, and expected lifetime of resistance based on published literature and personal communication (CIMMYT 1985; Dubin and Rajaram 1996; Great Plains Agricultural Council on Russian

---

[6]   We can work either with the original distribution or fit a smoothed distribution. The discrete distribution is in some sense more accurate; it gives the only data that we have. At the same time, the discrete distribution is itself an approximation of an underlying continuous distribution; it reflects a (somewhat arbitrary) scoring procedure used by researchers to translate field observations into data.

Wheat Aphid 1989; Marasas et al. 1997; Rajaram and van Ginkel 1996; Scharen and Sanderson 1983; Robinson 1994; Gilchrist and Mujeeb-Kazi 1996; J. Dubin, personal communication; L. Gilchrist, personal communication). Average yields by environment were obtained from the Wheat Impacts databases maintained by the CIMMYT Economics Program, and adjusted for yield potential according to estimates provided in Sayre et al. (1998) and Byerlee and Moya (1993). Details of the methods and parameters used in the scenarios are provided in Appendix 1.

"Crop losses averted" were used as an estimate of the benefits associated with achieving an improved level of resistance. Production losses in each time period have been calculated as

(15) $q_t = \sum_m a_{mt} \gamma_{mt} y_{mt}$

where a_{mt} is the area planted to varieties carrying resistance in environment m, in time t; γ_{mt} is the average annual percent yield losses due to disease over the area usually affected by disease in that year, and y_{mt} is the average wheat yield in farmers' fields.

The percent yield loss γ_{mt} that is averted with a new source of resistance decreases over the time period t as the source of resistance depreciates or decays with the evolution of new disease pathotypes. The path of depreciation is given by $\delta(t)$. The areas planted to varieties carrying resistance follow a diffusion path that differs by environment, with adoption ceilings based on the proportion of materials grown in that environment in 1990 that were CIMMYT crosses or had at least one CIMMYT parent. The period used to calculate the benefits, τ, depends on the type of material from which the source of resistance was obtained and the environment in which the resistant varieties are grown. The growing environment influences how much time elapses until the adoption of resistant varieties begins. The average research lag in each mega-environment covers the time from entry of the resistant lines into the breeding program, through the development of finished varieties, until the time farmers begin to grow them ($\bar{T}_m$).

Our estimates are based on a number of assumptions, most of which result in an understatement of the magnitude of benefits. First, we computed the benefits of finding useful materials on the assumption that CIMMYT performs the search, incorporates the materials into breeding lines, and disseminates the resulting breeding lines into national programs. This implies that benefit streams are based only on the major environments in which spring bread wheats are grown in the developing world, ignoring: (1) the potential for incorporating resistance into other wheat types; (2) other avenues through which useful materials might pass into national programs and farmers' fields; (3) benefits obtained through the subsequent diffusion of varieties that incorporate the resistance trait through more distant CIMMYT ancestors; and (4) benefits transferred through trade to consumers in countries not included in CIMMYT's mandate.

We also assume that the diffusion of varieties carrying a new source of resistance would follow a trajectory similar to the diffusion pattern of semidwarf wheats that was observed between 1967 and 1990. Today's diffusion paths, however, would probably have a steeper

slope because of improvements in seed systems and increased commercialization of wheat growers in many environments. In our simulations, we assume the same differences in diffusion between marginal and favorable environments that occurred historically.

Further, we consider only a 35-year time horizon, because depreciation of resistance and discounting imply that benefits obtained beyond this period are of little immediate consequence. It is possible to extend the analysis to incorporate longer durations, but there are few data on which to base this analysis, and quantitatively, benefits occurring outside this time period are of little consequence given standard discount rates.

As compared to a more comprehensive social impact analysis that would consider the effects of technical change on consumers, related factor or output markets, and health or environmental amenities, our analysis considers only production benefits. Nevertheless, genetic resistance to disease that circumvents the negative health and environmental effects of chemical control would generate benefits.

Finally, we assume that the evolution of new disease and pest biotypes can be adequately captured through the simple structure used here to model the breakdown of resistance over time as a constant rate of depreciation. The diffusion of resistant varieties can actually contribute to the emergence of new disease and pest pathotypes, which we treat as an exogenous event. We have also assumed that adoption ceilings are exogenous, when, in fact, the area planted to wheat adjusts to changes in prices induced by technical change (Jansen and Barker 1990). How the assumptions of exogeneity affect the benefits streams is unclear. Moreover, the simplicity of our approach reflects its objective: to analyze search economics as opposed to assess research impact.[7]

Costs. Data on fixed and variable costs associated with screening different types of materials and on likely rates of depreciation can be obtained from plant pathologists, gene bank managers, and other agricultural scientists. Cost estimates and the conceptual basis of the cost structure in this study are based on historical experience with the CIMMYT wheat germplasm bank.

For our purposes, most of the costs of operating a gene bank—including the costs of construction, power, and day-to-day management—can be viewed as fixed, in the sense that they do not vary with the size of a search or with the type of materials searched. For the purpose of computing the optimal size of a gene bank, these are important factors, but for the questions posed by this paper, relevant costs are those associated with evaluating wheat varieties for specific traits.

These costs have three components: general program costs, which are roughly proportional to the size of the search; evaluation costs, which are also linear in the size of the search; and transfer or "pre-breeding" costs, the costs of transferring useful genes from the varieties where they are initially found into high-quality breeding lines. The latter is a fixed cost incurred on a one-time basis when material is transferred into breeding lines.

---

[7]   For Russian wheat aphid, a more comprehensive impact analysis is Marasas et al. (1997).

For purposes of economic analysis, it is useful to distinguish between fixed and variable costs. The variable costs of search are the sum of experimental costs and transfer costs per accession. Experimental costs, which include expenditures for land preparation, preparation of insects or inoculum, crop management, and note-taking, are constant across materials and vary with n.

Pre-breeding is costly. Relevant expenses include the costs of evaluation, crossing, and genetic studies. The shares of these costs vary over the time period through T_j (see Appendix 1, Figure 1). The cost of pre-breeding varies depending on the type of material, j, in which the trait was found, but generally decreases with the improvement status of the source material. Simple traits and those that are highly heritable are relatively easy to transfer, while complex traits and those that are not highly heritable are difficult to transfer. At one extreme, for example, no pre-breeding is required to transfer a trait that is found within a released variety or an elite line. At the other extreme, a wild relative of wheat has high transfer costs because (at present) special techniques are required to make it useful to a crossing program. Applied molecular biology techniques are expected to reduce the costs of pre-breeding for such materials in the near future, but there will always be a range of pre-breeding costs for different sources of genetic resistance and traits.

In the work that follows, variable costs are assumed to fall in a lump sum in the first year of the time period ($t=0$), rather than at T_j, leading us to slightly overestimate variable costs. Transfer costs are treated as fixed costs incurred at time zero, but these fixed costs ultimately have little significance in the analysis; they can be netted out of the total benefit stream, as they are quantitatively insignificant. The *time* to transfer and the *time* in the breeding program, however, are critical in determining the discounted net benefits stream. These clearly vary by type of material, transfer technique, and breeding program.

Results

Optimal Search with Perfect Information
What is the optimal size of search for resistance when the probability distribution of resistance is "known"? We used the example of RWA to illustrate the solution to the problem. From its center of origin in the Caucasus and Central Asia, RWA has emerged as a pest of some importance in the United States, the Republic of South Africa, parts of the Southern Cone of Latin America, and North and East Africa. The pest is potentially important in Australia and parts of the People's Republic of China (see Robinson 1994). Some, but not all, of the wheat lines from the countries of origin of the pest have resistance (Marasas et al. 1997).

Searches among bread wheat varieties in the USDA collection yielded virtually no useful material, although a few durum wheats displayed effective resistance. Of 41,109 wheat accessions evaluated by the USDA—most of them elite lines and released varieties—slightly more than 100 displayed useful resistance, and of these, none was a spring-habit bread wheat. Literature summarizing searches for resistance also reports the near absence of sources of resistance in improved materials or any materials originating outside of Central Asia (Robinson and Skovmand 1992; Souza et al. 1991; Harvey and Martin 1990; du Toit 1987). The question facing CIMMYT researchers was how best to search their collection to find little-known sources of resistance.

A researcher who knows the distribution of resistance would determine the search size so that the expected marginal benefit of search equals the marginal search cost. Remember that the function $H_{n_j}^j(s^*)$ gives the probability that a search of size n among materials of type j will result in a "useful" discovery, where the usefulness of resistance is defined by having a score below s^*. We also know that H depends on the parameters of the underlying distribution of resistance, $\Delta^j(S)$. Denote these parameters by θ. To emphasize the function's dependence on n, the function can be written as $H_{n_j}^j(n; s, \theta)$. Then the optimal search is given by:

(16) $$\sum_{m \in M} \sum_{t=T_j+T_m}^{T_j+\overline{T}_m+\tau} \beta^t \delta(t)(Ev_m(z_j)-v_m(\overline{z}))] \times \frac{\partial H^j(n; s^*,\theta)}{\partial n}=c.$$

In this equation, the left-hand side represents the marginal benefit of expanding the search, with the first term giving the total benefit stream and the second term giving the marginal change in the probability of successful search when n is increased. On the right-hand side of the equation, c is the marginal cost of search. Note that under our earlier simplifying assumptions, $H_{n_j}^j(n; s^*, \theta)$ can be reduced to the Bernoulli expression $1-(1-p)^n$.[8]

Using data on the actual distribution $\Delta^j(S)$ and the actual benefit streams and search costs, we can compute the optimal search size, n^*, for the case of RWA. Figure 1 shows both the raw histogram and the smoothed beta distribution that approximates the underlying distribution of resistance among 10,190 landraces for which data were available in the GRIN. The distribution has a very thin left-hand tail; almost no landraces were found.

Table 1 shows total discounted net benefits for the incorporation of resistance to RWA into CIMMYT bread wheat materials. We present estimates of net benefits under a range of different assumptions about areas affected by the pest, average yield losses, longevity of resistance, and the time lags associated with pre-breeding and breeding (details on parameter assumptions are in Appendix 1, Table 1). The difference between the total affected areas assumed in the conservative and intermediate scenarios solely reflects uncertainty about their actual extent. Benefits in both of these scenarios are modest because in CIMMYT's mandate area, RWA is a major threat in only limited portions of marginal production environments, where yield losses inflicted by the pest are small because yield levels are

Parameters of the distribution
area= 6.276,
b=1.1078,
beta=0.122726.

Resistance level (0= resistant, 1=susceptible)
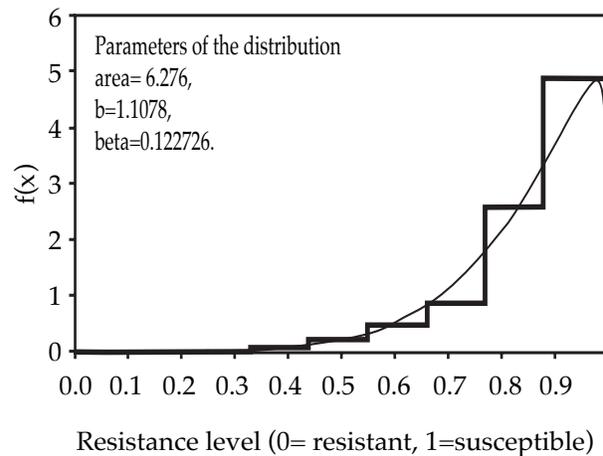
Figure 1. Actual distribution of resistance to Russian wheat aphid in landrace accessions, with smooth fitted approximation.
Note: Actual distribution based on GRIN data for 10,190 landraces of *Triticum aestivum*. Smooth approximation is a beta distribution fitted by least squares.

---

8   In particular, this reflects our treatment of resistance as a "threshold" characteristic, rather than as a continuous variable. We prefer to keep the more general form of the expression in equation (16) to illustrate the flexibility of our model in this respect.

generally low. Benefits are much larger in magnitude when based on the full global area delineated by the CLIMEX model (reported in Robinson 1994). Our estimation methods, however, remain very conservative. For example, it has been reported that the pest caused a cumulative loss of US$ 890 million in the United States alone from 1987 to 1993. Only 39% of the loss was attributed to lost production. About 9% was spent on chemical control, and 52% was estimated as lost in other economic activities.[9]

A key analytical result demonstrated in Table 1 is that the time lag for transfer, breeding, and adoption affects the economic return by a large order of magnitude. Transferring a source of resistance from a landrace using conventional breeding would undoubtedly require considerably longer than one or two years. The breeding lag usually assumed for CIMMYT is five years, while 10 years may be a more reasonable approximation for CIMMYT materials incorporated into and adapted by national breeding programs.

Table 1. Benefit streams (in US$ million 1990) associated with finding a source of resistance to Russian wheat aphid in a bread wheat landrace, under alternative assumptions

Assumptions and parameters	Transfer lag: 2 yr Breeding lag: 5 yr	Transfer lag: 5 yr Breeding lag: 10 yr	Transfer lag: 10 yr Breeding lag: 10 yr
Global	165.82	48.97	18.61
23 m ha affected			
8.3 m ha ceiling adoption			
5% initial average yield loss			
15 yr duration of resistance			
variable adoption lag			
Major CIMMYT spring wheat mega-environments	29.2	9.97	3.77
9 m ha affected			
3.4 m ha ceiling adoption			
5% initial average yield loss			
15 yr duration of resistance			
No adoption lag			
Major CIMMYT spring wheat mega-environments	15.67	3.315	1.2
4 m ha affected			
1.6 m ha ceiling adoption			
5% initial average yield loss			
15 yr duration of resistance			
No adoption lag			

---

[9] In their comprehensive analysis of the impact of the Russian Wheat Aphid Control Research Program in the Republic of South Africa, Marasas et al. (1997) estimated a 34.6% internal rate of return to the yield savings associated with the investment in developing resistant cultivars, for the period 1980-2005.

In all scenarios, the cost of search is estimated at US$ 82.97 per landrace screened. Based on the experience that variable search costs do not change substantially with the size of the search, we have assumed that all costs are variable costs, resulting in a constant average search cost equivalent to a constant marginal cost.

From these data, it is possible to compute the optimal search size for alternative scenarios. Figure 2 shows the relationship between search size and the probabilities of successful search for RWA among accessions of bread wheat landraces. Figures 3a-c combine this information with the associated marginal benefits and marginal costs. When the least favorable assumptions about the benefit stream are used, optimal search size is about 4,700 landrace accessions (illustrated by the point at which the marginal benefit curve intersects the marginal cost curve in Figure 3a). The most conservative scenario assumes initial average yield losses of 5% on a total of 4 million affected hectares, with an adoption ceiling of only 1.6 million hectares, and a 15-year longevity of resistance. The time lag for transferring resistance is 10 years, with an additional 10-year research lag for breeding (Table 1). In this scenario, at the optimal size of search, the total benefits of finding a landrace with resistance to RWA are US$ 0.865 million, with a total cost of US$ 0.406 million, for total net benefits of US$ 0.459 million.

The intermediate scenario presented in Figure 3b also assumes a long transfer and breeding lag, a 5% annual average yield loss, and 15 years duration of resistance, but the adoption ceiling in this scenario is 3.4 million hectares. The optimal size of search is about 10,000 landraces, with a total benefit of US$ 3.452 million, and a total cost of US$ 0.830 million for a total net benefit of US$ 2.622 million.
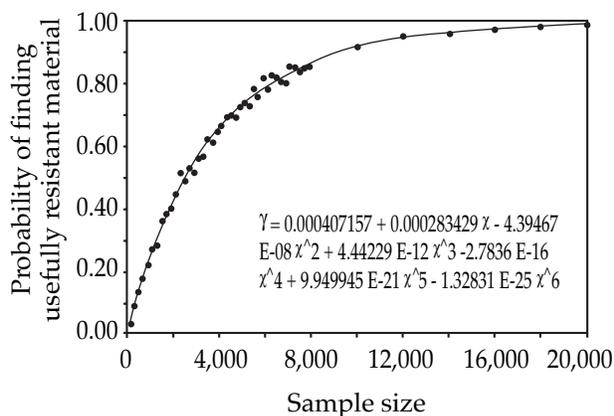
Figure 2. Probability of finding usefully resistant material from draws of given sample size, Monte Carlo simulations and smoothed function.

Note: Data from Monte Carlo simulations are based on repeated draws from the actual distribution of resistance to Russian wheat aphid among landrace accessions, with smoothed function. The smoothed function is a sixth-order polynomial, fitted by least squares.
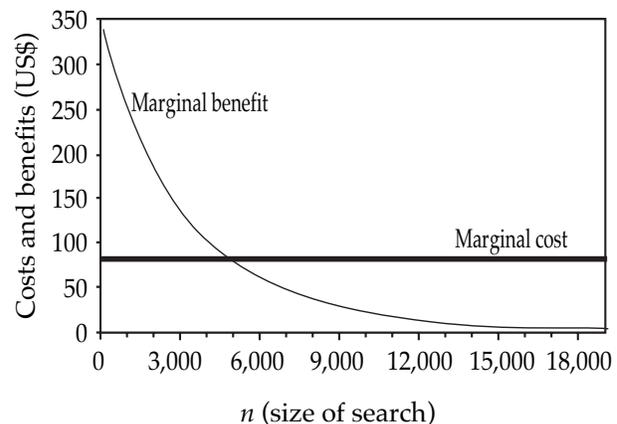
Figure 3a. Marginal costs and marginal benefits of seraching for Russian wheat aphid resistance in a sample of *Triticum aestivum* landraces, Low Scenario (total benefits = US$ 1.2 million).

The benefit streams in both the least favorable and intermediate scenarios are based exclusively on CIMMYT spring wheat mandate areas, and the longest research lag. Under the global benefits scenario with the longest research lag, the benefits are large enough to justify a search of approximately 18,000 landraces (Figure 3c). At the optimum, total search costs are US$ 1.493 million and expected benefits are US$ 18.228 million, for an expected net benefit of over US$ 16.735 million.

When global benefits streams are assumed with a total transfer and breeding lag of only seven years (Table 1), the economic problem becomes trivial—the optimal size of search is larger than the number of landraces in the CIMMYT gene bank. The benefits are so great relative to costs that a search of all existing accessions would be justified.

Given any total benefit stream, we can compute the optimal search size n^* and the associated expected net benefits. As Table 2 indicates, optimal search size increases with the total benefit stream.

The Value of Specialized Knowledge
The next experiment examines the value of specialized knowledge in regards to the distribution of desirable traits across various subpopulations of the set J. Figure 4 displays the actual distribution of resistance in a subset

Table 2. Optimal search size and expected net benefits as determined by the size of total benefit stream, for the case of Russian wheat aphid

Total benefits	Approximate optimal search size n^*	Expected net benefits
500,000	1,750	$46,388
1,000,000	4,100	321,428
5,000,000	10,500	3,752,204
10,000,000	12,650	8,456,173
15,000,000	14,150	13,236,039
20,000,000	18,900	18,153,771
25,000,000	19,150	23,086,572
30,000,000	14,000	28,022,376

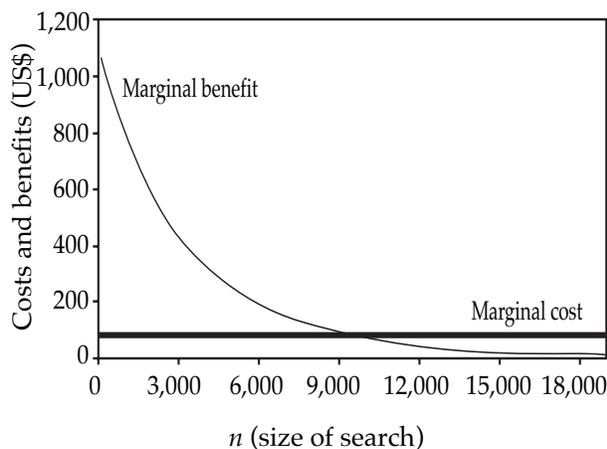**Figure 3b. Marginal costs and marginal benefits of searching for Russian wheat aphid resistance in a sample of *Triticum aestivum* landraces, Medium Scenario (total benefits = US$ 3.77 million).**
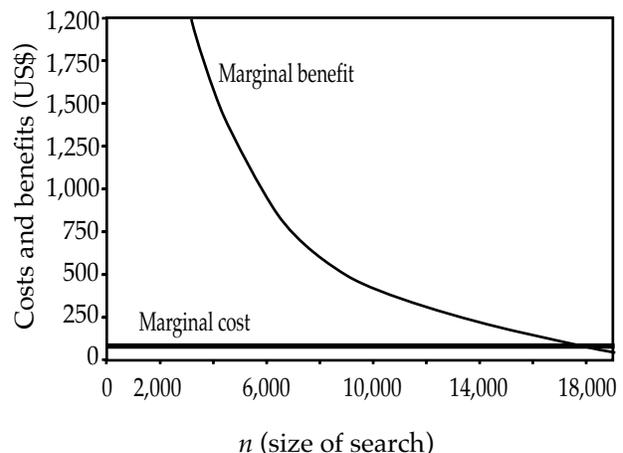
Figure 3c. Marginal costs and marginal benefits of searching for Russian wheat aphid resistance in a sample of *Triticum aestivum* landraces, High Scenario (total benefits = US$ 18.23 million).

of 1,089 Iranian landraces evaluated by the CIMMYT germplasm bank. This distribution implies that a very small search is likely to yield resistant materials; as shown in Figure 5, searches sizes of 15 were almost certain to result in usefully resistant materials. For the most favorable scenario, expected net benefits rise by US$ 1.69 million by searching Iranian landraces, rather than the entire landrace population. For the intermediate scenario, the expected net benefits rise by US$ 1.11 million with specialized knowledge; for the lowest case, the expected net benefits rise by "only" US$ 0.73 million. In this framework, two benefits are obtained if a gene bank manager knows how to focus a search. Specialized knowledge simultaneously allows researchers to save on search costs and to increase the probability of finding useful material. Both of these contribute to the increase in expected net benefits.

It is clear that specialized knowledge can be extraordinarily valuable. We do not make any claim, however, about the "uniqueness" of this specialized knowledge. Possibly, many people share the specialized knowledge, so that we are measuring the value of publicly available information. In many cases, though, it seems reasonable to assume that the specialized knowledge is held by a relatively small number of scientists.

Searching for Resistance in Multiple Categories of Material
In most cases, researchers have the option of searching for desirable traits in more than one category of germplasm such as landraces, elite lines, and uncultivated species. This experiment asks: how can search resources best be allocated among types of germplasm?

If more than one category of material is searched, then the efficiency conditions of economic theory require that the expected marginal benefits of search should be equalized across categories of germplasm. Commonly, however, such problems have corner solutions: when the distribution in each material is known, optimal search will omit all but the category with the highest expected marginal benefits.
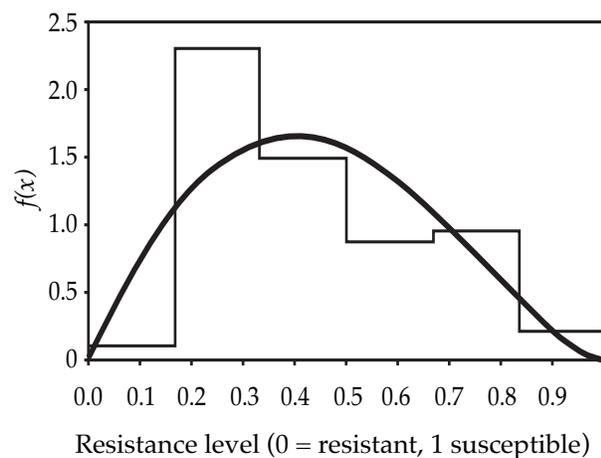
Figure 4. Beta distribution to approximate actual pattern resistance to Russian wheat aphid in 1,089 Iranian landraces of *Triticum aestivum,* as fitted by least squares approximation technique.
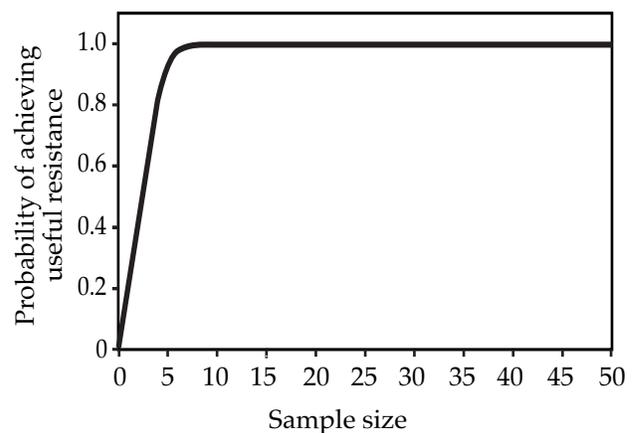
Figure 5. Probability of finding materials resistant to Russian wheat aphid in draws from a population of Iranian landraces.
Note: Based on Monte Carlo simulations using a smoothed version of the actual distribution of resistance in a population of 1,089 Iranian landraces screened by the CIMMYT wheat gene bank.

To analyze this question, we considered the case of *Septoria tritici*, a leaf blotch that affects wheat on over 10 million hectares worldwide. Most of the affected area is found in the CIMMYT mandate areas (9 million hectares), including large portions of the Southern Cone of South America (Brazil, Chile, Argentina, and Uruguay). In a research initiative to diversify the genetic basis of resistance to septoria tritici leaf blotch, CIMMYT scientists searched for new sources of resistance in breeding lines, landraces, and other materials, including emmer wheat, *T. diccoccon* (Fuentes and Gilchrist 1994; Gilchrist and Skovmand 1995; Gilchrist and Mujeeb-Kazi 1996). Here, we consider the comparison of breeding lines and emmer wheat only. Data on resistance is drawn from searches conducted by the CIMMYT wheat gene bank. As above, it is assumed that researchers have full information about the distributions of resistance within these two populations.

To determine how the search would optimally proceed with two types of materials, it is necessary to know the distributions of resistance, benefit streams, and search costs associated with each type. The distributions of resistance to septoria tritici leaf blotch are shown in Figures 6 and 7. Almost all of the emmer wheat tested, but only a few of the breeding lines, displayed useful resistance. The distribution of resistance in emmer wheat dominates the distribution of resistance in breeding lines, in the first-order stochastic sense. In other words, for any level of resistance, the probability of obtaining resistant material among emmer materials is greater than it is among breeding lines. While this suggests that scientists should ignore the breeding lines altogether, there are important differences in benefit streams between the two materials (Tables 3 and 4) that essentially result from differences in transfer time. The higher benefit stream is associated with finding resistance in breeding materials, because this resistance can be transferred almost immediately into breeding programs —in two years as compared to a minimum of five years for emmer wheat.

The variable costs of searching for resistant breeding materials are also much lower than the comparable costs of searching among emmer materials. It is relatively quick and easy to evaluate breeding materials for resistance: they can be subjected to disease stress, and resistant materials selected. Accessions of emmer wheat, on the other hand, must first be
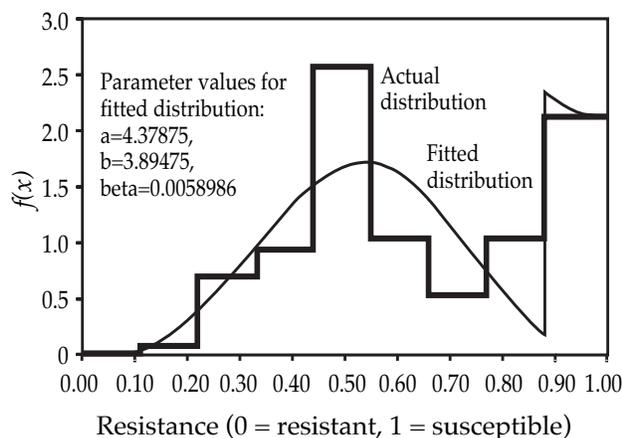

Figure 6. Actual and fitted distributions of resistance for septoria tritici leaf blotch in 1,834 breeding lines.
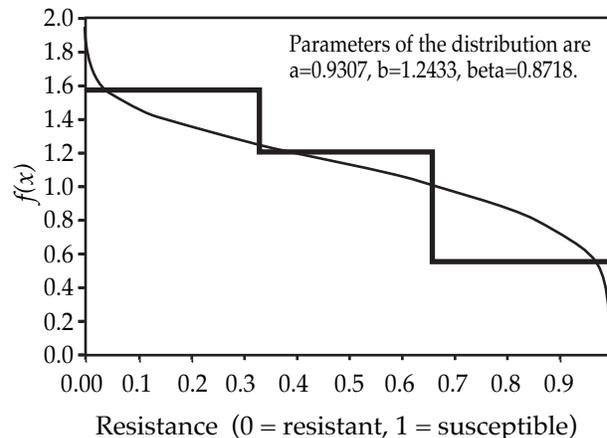
Figure 7. Actual distribution of resistance to *Septoria tritici* in 1,729 accessions of emmer, with smooth fitted approximation.

head-selected to remove heterogeneity, and it is more difficult and time-consuming to grow the plants and subject them to the necessary stresses because they are taller and less uniform. For the septoria case, average variable search costs have been estimated at about US$ 6 per accession for breeding material and approximately US$ 80 per accession for emmer.

In this experiment, the optimal search strategy is surprising. Despite the superior distribution of resistance among the accessions of emmer, it is optimal to search only within the category of breeding materials (i.e., to select a corner solution). To demonstrate this, suppose that we could be certain of finding an accession of emmer with useful resistance in a

Table 3. Benefit streams (in US$ million 1990) associated with finding a source of resistance to septoria tritici leaf blotch in breeding materials of spring bread wheat

Assumptions and parameters	Transfer lag: 2 yr Breeding lag: 5 yr	Transfer lag: 2 yr Breeding lag: 10 yr
Major CIMMYT mega-environments	15.022	7.989
9 m ha affected		
5.5 m ha adoption ceiling		
2% initial avg. yield loss in favorable areas		
1% initial avg. yield loss in marginal areas		
10 yr duration in favorable areas		
15 yr duration in marginal areas		
7 m ha affected		
4.1 m ha adoption ceiling		
2% initial avg. yield loss in favorable areas		
1% initial avg. yield loss in marginal areas		
7 yr duration in favorable areas		
10 yr duration in marginal areas	6.993	4.25

Table 4. Benefit streams (in US$ million 1990) associated with finding a source of resistance to septoria tritici leaf blotch in emmer wheat (assuming the resistance is transferred only into spring bread wheat)

Assumptions and parameters	Transfer lag: 5 years Breeding lag: 5 years	Transfer lag: 5 years Breeding lag: 10 years
Major CIMMYT mega-environments		
9 m ha affected	11.854	4.358
5.5 m ha adoption ceiling		
2% initial avg. yield loss in favorable areas		
1% initial avg. yield loss in marginal areas		
10 yr duration in favorable areas		
15 yr duration in marginal areas		
7 m ha affected		
4.1 m ha adoption ceiling	6.376	2.316
2% initial avg. yield loss in favorable areas		
1% initial avg. yield loss in marginal areas		
7 yr duration in favorable areas		
10 yr duration in marginal areas		

draw of size one. Using the most favorable set of estimates would yield benefits of US$ 6.376 million (less US$ 80 in search costs). But within the collection of breeding materials, a search of 220 varieties will have a 99% probability of finding usefully resistant materials.[10] The costs of this search are trivial (at most US$ 912), and the benefit stream is higher, because the yield gains are attained sooner due to the ease of transferring the resistance into new varieties. The net benefit is US$ 6.992 million, exceeding the expected benefits from the search of emmer accessions. It does not make sense to search for resistance in the emmer accessions if the same trait can be found in the breeding lines.

Although the distribution of *S. tritici* resistance in emmer is more favorable than that of breeding materials, the difference in benefit streams is of overriding importance. From an economic perspective, the breeding materials strictly dominate the emmers for resistance. In other cases, however, it may make economic sense to search in multiple categories of genetic resources, even if the payoffs are similar to those described for septoria tritici leaf blotch resistance. For example, it may be less expensive to search a second category of materials, given that a search is already being conducted. Moreover, multiple objectives can be pursued in a search; alternative objectives may include a scientific advance, the testing of a new technique, or the simultaneous screening of a material for several traits.

It will not always be advantageous for researchers to work with breeding materials. For some traits, such as those presented here, the data suggest strongly that the distributions of resistance are markedly superior in landraces or wild species. But for other traits, such as resistance to the rust diseases, the GRIN data show that the breeding materials stochastically dominate landraces and ancestral materials. Given the search cost differentials and time lags, it is frequently rational for researchers *not* to use the gene bank. Only when a desired trait is absent among breeding materials, or when resistance distributions are distinctly better for unimproved materials, does it make sense for researchers to focus their search on unimproved materials. This finding explains why plant breeders, pathologists, and other scientists sometimes appear reluctant to use unimproved materials from the gene bank. In their explanations, scientists often refer explicitly or implicitly to costs and time lags associated with using unimproved materials.

In some cases, however, researchers turn to unimproved materials in an effort to "broaden the base of resistance" to a particular disease or pest problem—in other words, to find alternative resistance genes that would substitute for an identified source of resistance. At first this seems to violate the principle described above. In fact, however, it adheres closely to it. To seek a new source of resistance is essentially to search for a new trait. The expectation is that the new trait will be distributed more favorably among unimproved materials than among breeding lines, which share a different source of resistance. Therefore, it makes sense to search among unimproved materials despite the longer time lags and higher costs. This was the thinking behind the search for resistance to septoria tritici leaf blotch: the goal was to find new sources of resistance that might allow breeders to diversify the genetic base.

---

[10] Using the actual discrete distribution, a draw of size 152 breeding lines will have a 99% probability of achieving useful resistance. Using the smoothed beta distribution, a draw of only 95 breeding lines will have a 99% probability of achieving useful resistance. "Useful resistance" is defined to be a resistance score below 2.5 on a 10-point scale.

Conclusions, Implications, and Directions
for Further Research

The results of this study allow us to draw several conclusions relating to the management and valuation of *ex situ* collections of genetic resources. One conclusion is that the optimal scale of a search for desirable traits is very sensitive to the size of the economic problem, as well as to the probability distribution for the trait. For some traits, the payoffs are simply not large enough to justify exhaustive searches. For other traits, the distributions are such that small searches will suffice. There are occasional situations, however, where the distribution of resistance and the payoffs to discovery are such that large searches are justified—under these circumstances large collections will prove quite valuable. Russian wheat aphid is one such example.

A second conclusion is that differences across types of genetic materials in the cost of searching and in the associated time lags can lead to optimal search strategies in which some materials are systematically ignored. Unless the probability of success with emmer is dramatically higher than with breeding lines, so long as the time lags associated with incorporating desirable traits from emmer are higher than the time lags associated with breeding lines, accessions of emmer will not be searched. Given conventional breeding techniques, the fact that many gene banks have been little used is economically rational—yet it says nothing about their long-term value as technologies and breeders' demand for traits evolve. Until new wide cross and molecular techniques can substantially reduce the cost and time constraints on evaluation and pre-breeding, we should expect that collections of landraces and wild relatives will be seldom used. Moreover, it is rational to turn to unimproved materials only after breeding lines have been searched extensively without success and the economic problem is large.

A third conclusion is that even in large collections, there are non-trivial benefits associated with marginal accessions. We have not yet attempted to model the entire marginal value of accessions, but the results presented here are suggestive. There are some situations where large searches are economically profitable, and by extension, large collections will be valuable. The question that remains is *how often* are large searches warranted, and with what expected payoffs. Our preliminary reading of the recent history of breeding is that marginal accessions are not redundant, and the conclusion of Simpson et al. (1996) is almost certainly not empirically relevant for a collection of genetic resources in an important economic crop like wheat.

The notion that unused gene banks are "morgues" and lack value is a total misreading of the economics of search. Our model implies that, with current search and transfer techniques, collections of landraces will indeed be used only on rare occasions—but high values will be associated with those uses. In other situations, it will be "efficient," in an economic sense, to keep landraces sitting unused in banks.

Further, it is important to recognize that we have used the term "utilization" in its narrowest sense. Even if breeding programs infrequently place direct demands on gene banks, each year large numbers of accessions are sent out on request to scientists for

genetics research and to increase the knowledge of the biochemical and molecular basis of traits. Gene banks are "utilized" for the accumulation of scientific knowledge, which in turn renders their use by breeding programs more strategic.

The results provided in this paper are illustrative. The main conclusion is that it is possible, using a straightforward process, to arrive at empirically sensible answers to questions about the valuation and management of *ex situ* collections of genetic resources. To answer such questions, we propose a relatively simple framework based on search theory. Our analysis depends on a two simplifying assumptions, neither of which is particularly restrictive. First, we have modeled the use of an *ex situ* germplasm collection as a source of resistance to a disease and pest, rather than as a source of other traits of agronomic importance. In the past, breeders' demand for new sources of resistance to diseases and pests appear to have motivated most of the searches of *ex situ* wheat collections. Many experts believe, however, that future increases in yield potential in bread wheat, and hence the demand for genetic resources, is most likely to arise from quantitative characteristics that are more costly to evaluate (Reynolds et al. 1996). Our model can be adapted for analyzing issues related to the search for such traits.

Our second simplifying assumption is that scientists "know" the form of the distribution of useful traits across all relevant populations when often they do not. In principle, however, our analysis would be essentially identical if scientists simply "guessed" the distributions on the basis of all available information. We do not, however, model the process by which scientists may *learn* about the distributions of useful traits. We assume here that researchers make a one-time decision about how many materials to evaluate, and that the search is conducted at a moment in time. More realistically, however, we could allow for a sequenced search in which different numbers of materials can be screened at different points in time. Sequential search could allow researchers to improve their prior information about distributions of resistance, to identify subpopulations with different resistance distributions, and to allow for repeated draws (for example, see Rausser and Small 1997). Does it make sense to initially search through a small sample of genetic resources, as a way of assessing the likely differences across types of material, or as a way of acquiring specialized knowledge about where to find desirable traits?

It is feasible that questions of this kind can be answered using the techniques and approaches outlined in this paper. Given a sufficiently large set of distributions for traits of economic value, the model proposed here could also be used to generate estimates of the expected value of adding accessions to a collection. Some questions are harder. How, for example, will new technologies affect the materials that can be searched or the costs of search? How will they change optimal search strategies? Will the current wave of biotechnological change raise or lower the marginal value of genetic resources? These are questions of substantial policy importance which remain to be addressed in future research.

Appendix 1. Methods

Probability Distributions for Traits

This paper draws on characterization and evaluation data for wheat varieties from the Genetic Resources Information Network (GRIN) compiled by the United States Department of Agriculture's (USDA) Agricultural Research Service. Data are collected by crop on a number of traits. Scientists have screened large numbers of varieties in the USDA germplasm collections for certain attributes, or "descriptors." These descriptors can be categorized in several ways. Typically, scientists distinguish between descriptors that are useful in identifying a particular variety through phenotypic characteristics and those that are of economic significance. The term "characterization" refers to screening for identifying phenotypic attributes; the term "evaluation" refers to screening for economically important traits, such as resistance to biotic or abiotic stresses.

In the GRIN data on wheat, a set of descriptors has been proposed by the Wheat Crop Germplasm Committee. For different descriptors, different numbers of varieties have been characterized or evaluated. For some descriptors, no systematic data are available. Thus, in the GRIN data, there are no data available for resistance to cereal leaf beetle (*Oulema melanopus*) or to bacterial leaf blight (*Pseudomonas syringae*). For other descriptors, enormous amounts of evaluation data are available. For example, between 10,000 and 20,000 varieties of bread wheat have been evaluated for resistance to each of three biotypes of Hessian fly (*Mayetiola destructor*). For some descriptors, the data include multiple observations on specific varieties. Where a particular variety has been tested in several trials or in several locations, it may show several observations.

For our work, we are interested in relatively crude measures of the distributions of traits within the population of wheat varieties. In particular, we relied on summary data compiled by the GRIN search software. This software provides a count of the number of varieties meeting specified search criteria. We recorded the numbers of varieties that achieved given scores for certain descriptors. We also disaggregated the data to consider possible differences within the collection of wheat varieties among landraces, cultivars, breeding lines, and other types of material.[11]

For purposes of time and interest, we focused our attention on the descriptors referring to economically useful traits—primarily, disease and pest resistance—rather than on phenotypic identifiers. This decision reflects our interest in the "economically important" traits, but it probably has little impact on our conclusions: in general, the two categories of descriptors are distributed in similar fashion. We also chose to consider the *performance* of varieties under evaluation conditions, rather than the *genetic composition* of those varieties, because this is the relevant criterion for economically important traits. We are interested in the frequency with which resistance to common bunt occurs in the population; we are much less interested in the underlying genetic mechanism of resistance. Thus, it is not the actual allele frequencies that we measure or describe here; instead, it is the phenotypic outcome.

---

[11] In the data, these various classifications of varieties are said to describe the "improvement status" of varieties.

With data on allele frequencies, we could arrive at estimates of how much genetic material is actually gained with a new variety of wheat, allowing us to address more directly questions about the benefits and costs associated with expanding the global collection of wheat germplasm.

Benefits Streams
The estimated diffusion paths for wheat varieties carrying resistance or tolerance to disease i in environment m were generated using logistic functions. Ceiling cumulative adoption levels were computed as

(1.1) $K_{im} = \alpha_m \theta_m D_{im}$,

where α_m is the weighted average percent of semidwarf varieties grown in 1990 that were CIMMYT crosses or had at least one direct CIMMYT parent, θ_m is the weighted average percent adoption of semidwarf varieties in that year, D_{im} is estimated percent of area affected by disease i in environment m, and weights used were regional areas planted to spring bread wheat by environment m (Byerlee and Moya 1993).

The initial year of the diffusion path, $(T_j + \overline{T}_m)$, was set seven years later in marginal than in favorable environments. This implies a breeding research lag $\overline{T}_m$ of seven years for favorable environments and 14 years for marginal environments. The D_{im} estimates, like the disease loss data, were developed from secondary sources and personal communication with CIMMYT staff.

Average annual percent yield losses due to disease over geographical areas (g) were estimated from a combination of losses reported in secondary literature sources, losses summarized in CIMMYT (1985), and personal communication of CIMMYT staff.[12]

For *Septoria tritici*, we assumed a 2% loss (Scharen and Sanderson 1983). The losses avoided in the first year of diffusion of varieties carrying new resistance to Russian wheat aphid (RWA) were estimated as 5% of yield in affected areas, as assumed in the report by the Great Plains Agricultural Council on Russian wheat aphid (1989). Marasas et al. (1997) report trial data, researchers' estimates, and farmers' estimates of yield losses to RWA in the Republic of South Africa. These estimates range from 9 to 13% in trials, and 2 to 20% among farmers surveyed in 1995. The variation in the percent of yield lost in farmers' fields is clearly very high.

The expected lifetime of a given source of resistance to RWA is not yet known both because the genetic basis of resistance is not yet fully understood and because cultivars with new sources of resistance have only recently been grown commercially. Marasas et al. (1997) report that since 1993, eight resistant cultivars have been released in the Republic of South Africa. Most contain the same single gene (*Dn1*); the longevity of resistance based on this gene is not yet known. In the baseline scenarios, we have assumed a lifetime of 15 years for a new source of resistance to RWA, and 10 years for *Septoria tritici*. Depreciation rates for

---

[12] Scharen and Sanderson (1983); Dubin and Rajaram (1996); Marasas et al. (1997); Great Plains Agricultural Council on Russian Wheat Aphid (1989); Robinson (1994); and J. Dubin and L. Gilchrist (personal communication).

resistance were estimated using straight-line depreciation from the initial percent yield savings to zero yield savings over the lifetime or resistance.

In general, disease pressures are lower in dry environments than in wet, warm environments, although this is not always true. In all scenarios we assumed a longer lifetime of resistance to *Septoria tritici* in the semi-arid areas, as well as a lower initial yield loss due to disease. The spring wheat environments in which RWA has caused relatively large losses are semi-arid, including parts of the Southern Cone of South America (in Argentina and Chile) and West Asia (in Iran and Turkey). Other countries in which the incidence and magnitude of losses are relatively high (the U.S., Canada, South Africa, Spain, and nations of Central Asia) or where the environment is favorable for the disease (parts of Australia and East Africa) are not within the major environments for growing spring bread wheat that are included in CIMMYT's traditional mandate. The areas of highest disease pressure for *Septoria tritici* are found in the Southern Cone of South America (Brazil, Chile, Argentina, and Uruguay).

Yields by environment and year (y_{mt})were generated from average farm-level yields in 1990 by environment, with an annual reduction of 0.5% (backwards in time) from 1990 for favorable environments and an annual reduction of 0.165% in marginal environments to remove the effect of gains in yield potential through breeding that occurred during that period (Sayre et al. 1998; information on differences in environment based on Byerlee and Moya 1993).[13]

Dat on total areas and average 1990 bread wheat yields in environment *m* were obtained from databases of the CIMMYT Economics Program, and the definition and description of the environments used in the analysis are found in Rajaram and van Ginkel (1996). The spring bread wheat area included in the analysis represents approximately 70-80% of the traditional mandate area of CIMMYT.[14]

For a summary of the base parameters used in the benefit simulations, see Appendix Table 1.

Cost Structure
Appendix Figure 1 illustrates how the distribution of costs changes over time in a search process. In the early years of the search, relatively more is spent on evaluation. In the case of the search for resistance to septoria tritici leaf blotch, transfer of resistance began before genetic studies commenced. At what point spending on genetic studies begins depends on the objective of the search. If the search for alternative sources of resistance is more important than the time lapse until a new source of resistance (whatever its genetic base) enters the breeding program, genetic studies will begin sooner.

---

[13] We did not correct for gains in maintenance breeding or resistance to disease. The yield series reflects actual yield levels in the presence both of disease and cultivation of resistant varieties. A more accurate series could be developed based on potential yield levels, changing disease pressures, and changing cultivation patterns.

[14] The data on which various estimates are based are least reliable for the People's Republic of China—the largest wheat producer in the developing world. The information in CIMMYT databases for China is limited relative to that of other major wheat producers in the developing world, although efforts are underway to improve it.
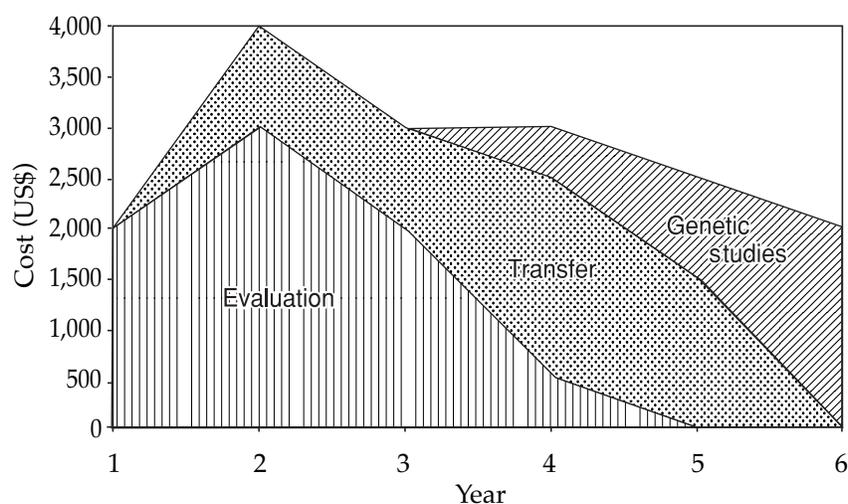
Appendix Figure 1. Costs of germplasm evaluation and utilization, for *Septoria tritici* resistance, in 215 accessions of emmer, by type of scientific activity.

Appendix Table 1. Summary of base parameters for benefits simulations

	Total area (m ha)	Affected area (m ha)	Adoption ceiling for CIMMYT-derived materials (% area)	Adoption lag (yr)	1990 yield (t/ha)	Average initial yield savings (% yield)	Expected lifetime of resistance (yr)	Annual rate of depreciation
Russian wheat aphid								
Mega-environment 1	32				2.56			
Mega-environment 2	7.5				2.66			
Mega-environment 3	1.7				1.42			
Mega-environment 4a	5.4	2.7	33	7	0.83	5	15	-0.000341
Mega-environment 4b	3.1	1.4	50	7	1.27	5	15	-0.000341
Mega-environment 4c	3				1.09			
Russian wheat aphid[a]								
Argentina/Chile		3.1	0.8	0	2.1	5	15	-0.000341
West Asia/North Africa		5.4	0.3	7	1.0	5	15	-0.000341
Mexico		0.4	1.0	0	3.485	5	15	-0.000341
Spain		0.5	0.9	0	2	5	15	-0.000341
Former Soviet Union/								
Eastern Europe		4.6	0.1	7	1.5	5	15	-0.000341
South Africa/Zimbabwe		1.2	0.6	0	1.7	5	15	-0.000341
Kenya/Ethiopia		0.45	0.1	0	1.5	5	15	-0.000341
U.S.		7	0.3	0	2.2	5	15	-0.000341
Canada		1	0.3	0	2	5	15	-0.000341
Septoria tritici leaf blotch								
Mega-environment 1	32				2.56			
Mega-environment 2	7.5	4	70	0	2.66	3	10	-0.003041
Mega-environment 3	1.7	1.7	56	0	1.42	3	10	-0.003041
Mega-environment 4a	5.4				0.83			
Mega-environment 4b	3.1	3.1	50	7	1.27	2	15	-0.001346
Mega-environment 4c	3				1.09			

Note: Some of these parameters are varied for the analysis, as reported in text tables.
[a]   Based on CLIMEX model delineation reported in Robinson (1994).

26

References

Autrique, E., M.M. Nachit, P. Monneveux, S.D. Tanksley, and M.E. Sorrells. 1996. Genetic diversity in durum wheat based on RFLPs, morphophysiological traits, and coefficient of parentage. *Crop Science* 36: 735-42.

Brown, G. Jr. 1990. Valuation of genetic resources. In G.H. Orians, G.M. Brown Jr., W.E. Kunin, and J.E. Swierzbinski (eds.), *The Preservation and Valuation of Biological Resources*. Seattle, Washington: University of Washington Press.

Byerlee, D. and P. Moya. 1993. *Impacts of International Wheat Breeding Research in the Developing World, 1966-90*. Mexico, D.F.: CIMMYT.

Chen, H.B., J.M. Martin, M. Lavin, and L.E. Talbert. 1994. Genetic diversity in hard red spring wheat based on sequence-tagged-site PCR markers. *Crop Science* 34: 1628-1632.

CIMMYT. 1985. *Wheat Producing Regions in Developing Countries*. Mexico, D.F.: CIMMYT.

du Toit, F. 1987. Resistance in wheat (*Triticum aestivum*) to *Diuraphis noxia* (Hemiptera: Aphididae). *Cereal Research Communications* 15: 175-179.

Dubin, H.J., and S. Rajaram. 1996. Breeding disease-resistant wheats for tropical highlands and lowlands. *Annual Review of Phytopathology* 34: 503-26.

Evenson, R.E. and Y. Kislev. 1975. *Agricultural Research and Productivity*. New Haven, Connecticut: Yale University Press.

Evenson, R.E. 1993. Genetic resources: Assessing economic value. Manuscript. New Haven, Connecticut: Yale University, Department of Economics.

Evenson, R.E. 1996. Valuing genetic resources for plant breeding: Hedonic trait value and breeding function methods. Manuscript prepared for the CEIS-Tor Vergata University Symposium on the Economics of Valuation and Conservation of Genetic Resources for Agriculture, 13-15 May, Rome, Italy.

Food and Agriculture Organization (FAO) of the United Nations. 1996. *The State of the World's Plant Genetic Resources for Food and Agriculture.* Background documentation prepared for the International Technical Conference on Plant Genetic Resources, Leipzig, Germany, 17-23 June, 1996. Rome, Italy: FAO.

Fuentes, S., and L. Gilchrist. 1994. O programa de septoria de CIMMYT. *Melhoramento* 33: 507-523.

Gilchrist, L.I., and A. Mujeeb-Kazi. 1996. Septoria tritici leaf blotch resistance germplasm derived from bread wheat/D genome synthetic hexaploids. Poster presented at the Annual Meetings of the American Society of Agronomy, August, Indianapolis, USA.

Gilchrist, L.I. and B. Skovmand. Evaluation of emmer wheat (*Triticum dicoccon*) for resistance to *Septoria tritici.* 1995. In L. Gilchrist, M. van Ginkel, A. McNab, and G.H.J. Kema (eds.), *Proceedings of a* Septoria tritici *Workshop.* Mexico, D.F.: CIMMYT.

Great Plains Agricultural Council on Russian Wheat Aphid. 1990. Russian wheat aphid economic loss summary-1989. In *Proceedings of the Fourth Russian Wheat Aphid Workshop*, 10-12 October 1990. Montana State University, Bozeman, Montana.

Harvey, T.L., and T.J. Martin. 1990. Resistance to Russian wheat aphid, *Diuraphis noxia*, in wheat (*Triticum aestivum*). *Cereal Research Communications* 18: 127-129.

Jansen, H.G.P., and R. Barker 1990. Adoption ceilings and modern coarse cereal cultivars in India. *American Journal of Agricultural Economics* 72: 653-663.

Marasas, C., P. Anandajayasekeram, V. Tolnay, D. Martella, J. Purchase, and G. Prinsloo. 1997. *Socio-Economic Impact of the Russian Wheat Aphid Control Research Program.* Gaborone, Botswana: Southern African Center for Cooperation in Agricultural and Natural Resources Research and Training (SACCAR).

Maxted, N., B.V. Ford-Lloyd, and J.G. Hawkes (eds.). 1997 *Plant Conservation: The* In Situ *Approach.* London, U.K.: Chapman and Hall.

Pearce, D. W., and D. Moran, in association with the Biodiversity Programme of the IUCN (the World Conservation Union). 1994. *The Economic Value of Biodiversity*. London, U.K.: Earthscan.

Pecetti, L., and A. B. Damania. 1996. Geographic variation in tetraploid wheat (*Triticum turgidum* ssp. *turgidum convar.* durum) landraces from two provinces in Ethiopia. *Genetic Resources and Crop Evolution* 43: 395-407.

Rajaram, S., and M. van Ginkel. 1996. *A Guide to the CIMMYT Bread Wheat Section*. Wheat Program Special Report No. 5. Mexico, D.F.: CIMMYT.

Rausser, G.C., and A.A. Small. 1997. *Bioprospecting with Prior Ecological Information.* Giannini Foundation Working Paper No. 819. Berkeley, California: Dept. of Agricultural and Resource Economics, University of California, Berkeley.

Reynolds, M.P., S. Rajaram, and A. McNab (eds.). 1996. *Increasing Yield Potential in Wheat: Breaking the Barriers.* Mexico, D.F.: CIMMYT.

Robinson, J. 1994. *Identification and Characterization of Resistance to the Russian Wheat Aphid in Small-Grain Cereals: Investigations at CIMMYT, 1990-92*. CIMMYT Research Report No. 3. Mexico, D.F.: CIMMYT.

Robinson, J., and B. Skovmand. 1992. Evaluation of emmer wheat and other Triticeae for resistance to Russian wheat aphid. *Genetic Resources and Crop Evolution* 39: 159-163.

Sayre, K.D., R.P. Singh, J. Huerta-Espino, and S. Rajaram. 1998. Genetic progress in reducing losses to leaf rust in CIMMYT-derived Mexican spring wheat cultivars. *Crop Science* (in press).

Scharen, A.L., and Sanderson, F.R. 1985. Identification, distribution and nomenclature of the Septoria species that attack cereals. In: A.L. Scharen (ed.), *Septoria of Cereals: Proceedings of a Workshop*. USDA-ARS Publication No. 12. Washington, D.C.: U.S. Department of Agriculture.

Simpson, R.D., R.A. Sedjo, and J.W. Reid. 1996. Valuing biodiversity for use in pharmaceutical research. *Journal of Political Economy* 104:(1): 163-85.

Smale, M. 1997. The Green Revolution and wheat genetic diversity: Some unfounded assumptions. *World Development* 25: 1257-1269.

Souza, E., C.M. Smith, D.J. Schotzko, and R.S. Zemetra. 1991. Greenhouse evaluation of red winter wheats for resistance to Russian wheat aphid (*Diuraphis noxia,* Mordvilko). *Euphytica* 57: 221-225.

Swanson, T. 1995. The values of global biodiversity: The case of PGRFA. Paper prepared for presentation at the Technical Consultation on Economic and Policy Research for Genetic Resource Conservation and Use, 21-22 June, 1995, International Food Policy Research Institute, Washington, D.C.

Tanksley, S.D., and S.R. McCouch. 1997. Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science* 277: 1063-1066.

Tsegaye, S., H.C. Becker, and T. Tesemma. 1994. Isozyme variation in Ethiopian tetraploid wheat (*Triticum turgidum*) landrace agrotypes of different seed color groups. *Euphytica* 75: 143-147.

Webster, J.A., and S. Amosson. 1994. *Economic Impact of the Russian wheat aphid in the Western United States: 1992-1993.* A Report by the Russian Wheat Aphid Task Force to the Great Plains Agricultural Council, 1994. Great Plains Agricultural Council Publication 152, December 1994.

Wright, B.D. 1997. Crop genetic resource policy: The role of *ex situ* genebanks. *The Australian Journal of Agricultural and Resource Economics* 41: 81-115.

Recent Economics Working Papers

96-01   Fertilizer Use and Maize Production in Sub-Saharan Africa
P. Heisey and W. Mwangi

96-02   Understanding Global Trends in the Use of Wheat Diversity and International Flows of Wheat Genetic Resources
M. Smale and contributors

96-03   Adoption and Impact of High Yielding Wheat Varieties in Northern Tunisia
M.E. Saade

97-01   The Contribution of Genetic Resources and Diversity to Wheat Productivity: A Case from the Punjab of Pakistan
J. Hartell, M. Smale, P.W. Heisey, and B. Senauer

97-02   The Maize Seed Industries of Brazil and Mexico: Past Performance, Current Issues, and Future Prospects
M.A. López-Pereira and Joao Carlos Garcia

97-03   Farmers' Use of Improved Seed Selection Practices in Mexican Maize: Evidence and Issues from the Sierra de Santa Marta
E. Rice, M. Smale, and J.L. Blanco

97-04   Towards Reduced Pesticide Use for Cereal Crops in Asia
P.L. Pingali and R.V. Gerpacio

97-05   Organization and Performance of National Maize Seed Industries: A New Institutionalist Perspective
M.L. Morris and M. Smale

97-06   Adoption, Management, and Impact of Hybrid Maize Seed in India
R.P. Singh and M.L. Morris

98-01   Institutional Change and Discontinuities in Farmers' Use of Hybrid Maize Seed and Fertilizer in Malawi: Findings from the 1996-97 CIMMYT MoALD Survey
M. Smale, A.Phiri, and contributors

98-02   An Economic Analysis of the *Abonera* Maize Production System in the Atlantic Coast of Honduras
G. Sain and D. Buckles

CIMMYT World Wheat Facts and Trends

CIMMYT World Wheat Facts and Trends, 1995/96
Understanding Global Trends in the Use of Wheat Genetic Diversity and International Flows of Wheat Genetic Resources