

Searching a Gene Bank: The Case of Wheat

D. Gollin, M. Smale, and B. Skovmand

Paper to be presented at:

American Agricultural Economics Association annual meeting

Salt Lake City, Utah

July 1998

ABSTRACT:

This paper presents theoretical and empirical results on the optimal search for genetic traits within a crop gene bank. Three specific questions are posed and answered, using data on the distributions of resistance genes and on the costs and benefits of search. First, we ask what is the optimal size of search for a trait of economic value? Second, what is the value of specialized knowledge about likely sources of resistance? Third, how should search resources be allocated across different types of genetic resources? Results demonstrate that for some traits, large searches are justified. In other cases, the costs and time lags associated with conventional pre-breeding techniques may make it efficient to ignore certain categories of genetic resources. In general, the infrequency of searches does not imply that additional accessions are valueless; low utilization of the gene bank by breeders does not imply that banks are “seed morgues.”

* Douglas Gollin is Assistant Professor of Economics at Williams College and Affiliate Scientist, CIMMYT Economics Program. Melinda Smale is CIMMYT economist, and Bent Skovmand is the Head of Wheat Genetic Resources, CIMMYT. Among those who have contributed helpful comments and suggestions are: Norman Borlaug, Cheryl Doss, Jesse Dubin, Bob Evenson, Paul Heisey, Lucy Gilchrist, Michael T. Jackson, Stephane Lemarié, Sanjaya Rajaram, Matthew Reynolds, Greg Traxler, and Maarten van Ginkel. Harold Bockelman, director of the U.S. Department of Agriculture’s National Small Grains Collection, made available part of the GRIN data base. Will Chang provided research assistance. The views represented here are those of the authors, and do not represent the official views of CIMMYT.

Introduction

The conservation of genetic resources for crop improvement has emerged as a controversial policy issue in recent years. In response to concerns about the perceived narrowing of the genetic base of crop plants (Shiva, 1993; Raeburn, 1995), the international community has spent millions of dollars to collect and conserve varieties of rice, wheat, maize and other food crops. International institutions have devoted substantial resources to developing policies and protocols for the protection and exchange of crop genetic resources.

There are two principle strategies for preserving genetic resources for crop species, although the strategies are increasingly viewed as complementary rather than as substitutes (Maxted *et al.* 1997). One strategy is to preserve different varieties or species *in situ*. For cultivated species, *in situ* conservation as it is currently understood implies the management of traditional varieties and wild materials by farmers in centers where the crop has evolved.

Another 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. The early collections consisted of botanical gardens or simple fields in which different plant varieties were physically cultivated. 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 at conditions of low temperature and humidity for long periods of time. Such collections are principally intended to maintain genes for future crop improvements through plant breeding.

Ex situ storage remains the principal means of genetic resource conservation. 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 percent of wheat landraces and 60 percent of wild species have been collected, although it is unclear on what data these figures are based, since the actual total number of populations or species is unknown. 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 expanded and upgraded their gene banks.

Wright (1997) has argued that the utilization of these resources has not kept pace with their expansion. Critics have argued that *ex situ* collections are wasteful and poorly managed. One criticism, drawn from evolutionary biology, is that the materials stored in gene banks are “frozen” at the time of their collection, losing their potential to adapt to the changing natural and economic environments in which farmers produce their crop (Guldager, 1975; see Frankel and Soulé 1981). A second criticism, grounded in microeconomic principles, is that low utilization of accessions by breeding programs indicates that banks have excess capacity and marginal accessions are valueless. Most genetic resource specialists would agree that “accessions should be *used*, and breeders need to know what the packets or bottles of seeds on the shelves contain.” (Plucknett *et al.* 1987: 74).

Do large collections ever have value? Are breeders and other scientists wastefully ignoring the materials in gene banks? Is low utilization by breeders of existing collections an indication of economic inefficiency? Our work investigates such issues. We use a basic search model, and apply it to data on resistance to Russian wheat aphid and *Septoria* leaf blotch to identify optimal strategies for searching the gene bank and for utilizing different categories of materials.

Three specific questions are posed and answered in the analysis. First, we ask how many gene bank accessions should optimally be included in a search for a useful trait. Second, we estimate the value of specialized knowledge about the distribution of desirable traits. In this case, we compute the value of knowing that resistance to Russian wheat aphid is more common among a set of bread wheat landraces from Iran than among the general population of bread wheats. Third, we ask how an optimal search should proceed when scientists can hunt for a particular trait among two distinct populations with differing distributions of resistance, using the case of *Septoria* leaf blotch as an illustration.

Data and methods

There is little confusion or disagreement over the concepts to apply in economic analysis of genetic resources; simple marginal analysis suffices (Brown 1990; Evenson 1993; Pearce and Moran 1994; Swanson 1995). Searches of a collection should proceed until the marginal expected value of success is outweighed by the search cost.

To answer questions about searching *ex situ* collections of crop genetic resources, we need a model that does not arbitrarily impose any assumption about the nature or the distribution of desirable traits. The approach used here builds on work of Evenson and Kislev (1975) and Simpson *et al.* (1996). The underlying model allows for continuous distributions for the genetic expression of a trait and implies imperfect substitutability among bank accessions. (A more complete version is presented in Gollin *et al.* 1998). To represent the specific case of screening for disease resistance in wheat, however, we identify a level of “useful” resistance from a distribution of scores. We define the probability of success as the probability of finding an accession with a “useful” level of resistance, and the draws of accessions resemble repeated Bernoulli trials.

What is the optimal size of search for resistance when the probability distribution of resistance is known? A researcher who knows the distribution of resistance would choose the size of the search such that the expected marginal benefit of search equals the marginal search cost. Let the function $H_{n_j}^j(s^*)$ give 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^* . Material j is a particular subpopulation of materials from the set J of possible subpopulations, such as a subpopulation of landraces, obsolete lines, wild relatives, or breeding lines. 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 , we can write the function as $H_{n_j}^j(n; s^*, \theta)$.

Then the optimal size of search within a subpopulation j is given by:

$$(1) \quad \left[\sum_{m \in M} \sum_{t=T_j+\bar{T}_m}^{T_j+\bar{T}_m+\tau} \beta^t \delta(t) (Ev_m(z_j) - v_m(\bar{z})) \right] \times \frac{\partial H^j(n; s^*, \theta)}{\partial n} = c.$$

The left-hand side in this equation 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.¹ Note that under our assumptions regarding the search for resistant materials, $H_{n_j}^j(n; s^*, \theta)$ can be reduced to the Bernoulli expression $1 - (1 - p)^n$. Resistance is treated here as a “threshold” characteristic, rather than as a continuous variable.

The total benefit stream represents the sum of the discounted annual savings. As new biotypes of pathogens emerge over time, the productivity advantage of a novel source of resistance follows a depreciation path $\delta(t)$. The discount factor β corresponds to the rate of time preference of the decision-makers represented in the model. The timing of the search process, varietal release and diffusion have implications for the benefits stream. There is a time lag associated with transferring resistance genes from materials identified in the evaluation process into a breeding program, which varies by the type of material and by the heritability of the trait (T_j). This period is often called “pre-breeding”. A second time lag marks the time from transfer to the breeding program to the time at which diffusion begins in farmers’ fields (T_m). This time lag differs by mega-environment. For favorable areas, varieties may be developed more quickly than for marginal environments. Further, because of other factors related to farmer adoption of new wheat varieties, the diffusion of materials that carry the resistance may begin later in some areas than in others. Finally, let τ denote the time horizon of relevance to breeders.

The right-hand side of the equation represents the marginal cost of search, c . For the choice of the optimal number of accessions to search within a category of material j , the relevant costs are those associated with screening and evaluating an additional accession. For the questions posed here, 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. The costs of the breeding program are also fixed from the standpoint of the search decision.

¹ Note that benefits are summed across a set of regions, indexed by m .

In the experiments described here, 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 Gollin *et al.* 1998.

The actual distributions of traits were obtained from the Genetic Resource Information System (GRIN) of the U.S. Department of Agriculture, which contains performance scores in various agronomic trials. A least squares technique was used to fit a smooth, beta distribution to each distribution of disease and pest resistance scores.

Benefit streams were estimated from the areas affected by diseases, average annual yield losses, and expected lifetime of resistance based on published literature and personal communication. Average yields by environment were obtained from the Wheat Impacts databases held by CIMMYT Economics Program, and adjusted for yield potential according to estimates provided in Sayre *et al.* (1998) and Byerlee and Moya (1993). “Crop losses averted” were used as a measure of the benefits associated with achieving an improved level of resistance.

Cost estimates were generated from the historical experience of the CIMMYT wheat germplasm bank. Search costs were recorded in terms of three components: general program costs, evaluation costs, and the costs of transferring useful genes from the varieties where they are initially found into breeding lines. Program costs, such as salaries and equipment, are roughly proportional to the size of the search undertaken. Evaluation costs include expenses for land and crop management, preparation of insects or inoculum, and note taking. These costs are also linear in the size of search. They generally do not differ with the types of materials being evaluated, but they may vary according to the characteristics of the disease or pest. By contrast, transfer costs—also known as “pre-breeding” costs — vary with the different types of materials used and the nature of the trait. They usually decrease 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 traits that are not highly heritable are difficult to transfer.

Results

(1) Perfect information

We use the example of Russian wheat aphid to illustrate the solution to the search problem with perfect information. From its center of origin in the Caucasus and Central Asia, Russian wheat aphid (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.

Searches among bread wheat varieties in the USDA collection yielded almost no useful material, although a few wheats displayed effective resistance. Of 41,109 wheat accessions evaluated by the USDA — most of them elite lines and released varieties — just over 100 displayed any useful resistance.

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 resistant landraces were found.

In all scenarios, the cost of search was estimated at \$82.97 per landrace screened (Skovmand, unpublished data). 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. When the least favorable assumptions about the benefit stream are used, optimal search size is about 4,700 landrace accessions. In this scenario, the total benefits of finding a landrace with resistance to Russian wheat aphid are \$865,000 with a total cost of \$406,000, for total net benefits of \$459,000. The intermediate scenario yields an optimal search size of about 10,000 landraces, with a total benefit of \$3,452,000, and total cost of \$830,000 for a total net benefit of \$2,622,000. Under the assumptions that

give the largest benefit stream, the payoffs are large enough to justify a search of approximately 18,000 landraces. Total search costs are \$1,493,000 and expected benefits are \$18,228,000, for an expected net benefit of over \$16,735,000.

For sufficiently high estimates of benefits, 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. This result does not tell us whether it would be worthwhile to collect additional accessions, however, since we have not considered the costs of collection and storage.

(2) The value of specialized knowledge

Our next experiment asks the value of specialized knowledge about the distribution of desirable traits across various subpopulations of wheat varieties. In a subset of 1,089 Iranian landraces evaluated by the CIMMYT germplasm bank, the distribution of resistance was such that searches of size 15 were almost certain to result in usefully resistant materials. For the most favorable scenario, expected net benefits rise by \$1,690,000 when Iranian landraces are searched instead of the entire landrace population. For the intermediate scenario, the expected net benefits rise by \$1,110,000 with specialized knowledge; for the lowest scenario, the expected net benefits rise by \$73,000. There are two benefits obtained if a gene bank manager knows how to focus a search: (1) savings on search costs, and (2) greater probability of finding useful material. Both of these contribute to the increase in expected net benefits.

Although it is clear that specialized knowledge can be extraordinarily valuable, we do not make any claim as to its “uniqueness”. It may be that 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 imagine that the specialized knowledge is held by a relatively small number of scientists.

(3) 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 final experiment asks: how best can search resources be allocated among types of germplasm?

If more than one category of material is searched, or an interior solution holds, 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. The researcher's problem is to select the material that gives the highest overall returns.

To analyze this question, we consider the case of *Septoria tritici*, a leaf blotch that affects 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* leaf blotch, CIMMYT scientists searched for new sources of resistance in breeding lines, landraces and other materials, including emmer wheat, *T. diccoccum*. Here, we consider the comparison of breeding lines and emmer wheat only². We use data on resistance from searches conducted by the CIMMYT wheat gene bank. As above, we assume that researchers have full information about the distributions of resistance within these two populations.

To ask how search would proceed optimally with two types of materials, we need to know the distributions of resistance, benefit streams, and search costs associated with each type. Almost all the emmer wheats tested displayed useful resistance, whereas relatively few of the breeding lines were usefully resistant. The distribution of resistance in emmer wheat thus dominates the distribution of resistance in breeding lines, in the first-order stochastic sense. This might suggest that scientists should ignore the breeding

² Emmer is a tetraploid relative of bread wheat that is still found in wild and cultivated forms.

lines altogether. But there are important differences in benefit streams between the two materials, resulting essentially from differences in time lags associated with transferring the resistance into breeding programs. The higher benefit stream is associated with finding resistance in breeding materials, where the transfer into breeding programs is almost immediate — two years, by a conservative estimate, as compared to 5 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 simply be subjected to disease stress, and resistant materials can be selected. Accessions of emmer wheat must first be 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 case of *Septoria*, average variable search costs have been estimated at about \$6 per accession for breeding material and about \$80 per accession for emmers.

The optimal search strategy in this case, 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 see this, suppose that we could be certain of finding an accession of emmer with useful resistance in a draw of size one. Using the most favorable set of estimates, this would yield benefits of \$6,376,000 (less \$80 in search costs). But within the collection of breeding materials, a search of at most 220 varieties will have a 99 percent probability of finding usefully resistant materials.³ The costs of this search are trivial (at most \$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 \$6,992,000, exceeding the expected benefits from the emmer wheats. Under this scenario, then, it does not make sense to

³ Using the actual discrete distribution, a draw of size 152 breeding lines will have a 99 percent probability of achieving useful resistance. Using the smoothed beta distribution, a draw of only 95 breeding lines will have a 99 percent probability of achieving useful resistance. “Useful resistance” is defined to be a resistance score below 2.5 on a 10-point scale.

search for resistance among the emmer accessions if the same trait can be found in the breeding lines. Other benefit scenarios give rise to the same conclusion.

It will not always be true that researchers are better off working with breeding materials. For some traits, the distributions of resistance are superior in landraces or wild species. When a desired trait is absent among breeding materials, or when the resistance distributions are distinctly better for unimproved materials, it makes sense for researchers to focus their search on landraces and breeding lines. Given the search cost differentials and time lags, however, it is frequently rational for researchers *not* to use unimproved materials from the gene bank. This finding explains why plant breeders, pathologists, and other scientists sometimes appear reluctant to use such materials.

Conclusions

We have demonstrated several results in this paper that relate 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. These will be the situations when large collections are valuable. Russian wheat aphid is one such example.

A second conclusion is that differences across types of genetic materials in the cost of search 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 wheats is dramatically higher than with breeding lines, so long as the time lags associated with incorporating desirable traits from them are higher than the time lags associated with breeding lines, accessions of emmer wheat will be excluded. Given conventional breeding techniques, the fact that many gene banks have been little used is economically rational — and 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 used seldom. It makes economic sense to turn to landraces and wild relatives primarily in situations where breeding lines have been searched extensively without success and the economic problem is large. An implication of this is to refute the notion that unused gene banks are “seed morgues” and lack value. This view represents 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 occasions. In other situations, it will be “efficient,” in an economic sense, to keep landraces sitting unused in banks.

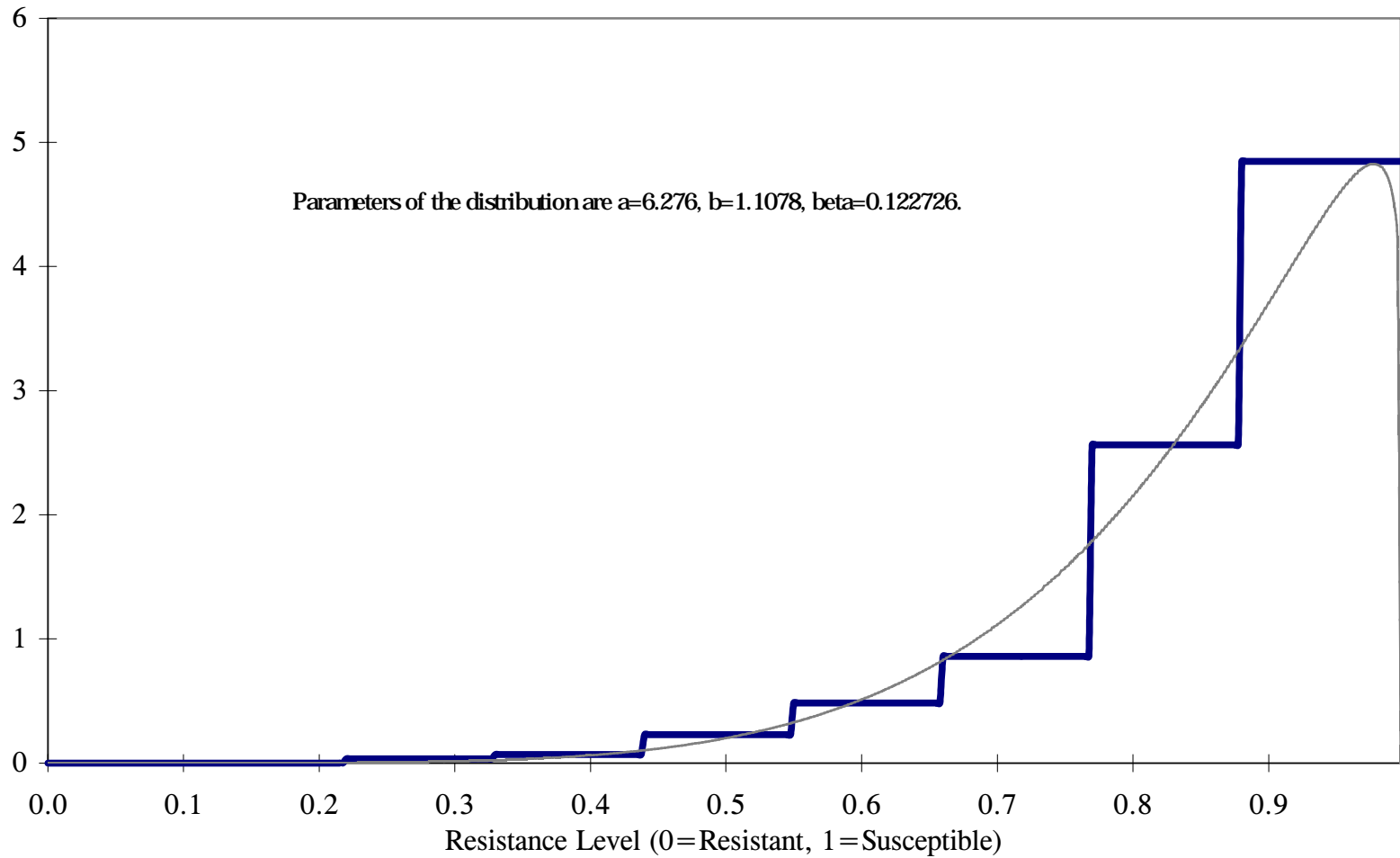
A third conclusion is that even in large collections, there are non-trivial benefits associated with marginal accessions. We have not yet attempted to model the marginal value of accessions, but the results presented here are suggestive. There are some situations where large searches are economically profitable, and by extension, large collections will be valuable. The question that remains is *how often* large searches are warranted, and with what expected payoffs.

Many further questions remain. How, for example, will new technologies affect the materials that can be searched, and how will they affect the 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.

References

- 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.
- Evenson, Robert E. and Yoav Kislev. 1975. *Agricultural Research and Productivity*. New Haven: Yale University Press.
- Evenson, Robert E. 1993. Genetic Resources: Assessing Economic Value. Manuscript. New Haven, Conn.: Yale University, Department of Economics.
- Food and Agriculture Organization 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: FAO.
- 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: Earthscan.
- Plucknett, D.L., N.J.H. Smith, J.T. Williams, and N. M. Anishetty. 1987. *Gene Banks and the World's Food*. Princeton University Press, Princeton, New Jersey.
- 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.: International Maize and Wheat Improvement Center (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.=20
- 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).
- Simpson, R. D., R. A. Sedjo, and J. W. Reid. 1996. Valuing biodiversity for use in pharmaceutical research. *Journal of Political Economy* 104:1 (February), 163-85.
- 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, International Food Policy Research Institute, Washington, DC, June 21-22, 1995. Manuscript. Cambridge University, Faculty of Economics.
- Wright, B.D. 1997. Crop genetic resource policy: the role of *ex situ* genebanks. *Australian Journal of Agricultural and Resource Economics* 41: 81-115.

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.