Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild

Steven D. Tanksley, *et al.*

Science **277**, 1063 (1997);
DOI: 10.1126/science.277.5329.1063

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

ARTICLE

Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild

Steven D. Tanksley and Susan R. McCouch

Nearly a century has been spent collecting and preserving genetic diversity in plants. Germplasm banks—living seed collections that serve as repositories of genetic variation—have been established as a source of genes for improving agricultural crops. Genetic linkage maps have made it possible to study the chromosomal locations of genes for improving yield and other complex traits important to agriculture. The tools of genome research may finally unleash the genetic potential of our wild and cultivated germplasm resources for the benefit of society.

As we head into the next millennium, the world faces a greater demand on agricultural output than at any time in history. Despite efforts to curb birthrates, the Earth's human population is expected to rise to 8.9 billion by the year 2030, corresponding to more than a 50% increase from the current population of 5.7 billion (*1*). In the past, we have met the demand for increased agricultural productivity by a combination of genetic improvements, greater farming inputs (fertilizers, pesticides, and water), and cultivation of more land. With dwindling freshwater reserves and petroleum resources (on which fertilizers and pesticides are based) and increased problems caused by agricultural pollution, we can hardly expect to increase or even maintain our current levels of agricultural inputs. Similarly, much existing farmland is falling victim to urban expansion, and it is unlikely that new farmland will become available in the near future. That leaves the genetic improvement of crops as the most viable approach by which food production can attempt to keep pace with the anticipated growth of the human population. For the genetic approach to succeed, we must harness the wealth of genetic variation provided by nature and currently warehoused in our seed repositories. Until now we have been only modestly successful in utilizing these resources for plant improvement. New findings from genome research indicate that there is tremendous genetic potential locked up in seed banks that can be released only by shifting the paradigm from searching for phenotypes to searching for superior genes with the aid of molecular linkage maps.

The Narrow Genetic Base of Crop Plants

Today, modern agriculture—and, for that matter, human existence—is dependent on the cultivation of a few highly productive crop species. These food crops were first domesticated from wild species about 10,000 years ago during the transition from nomadic hunter-gatherers to life in agrarian societies. Considering that flowering plants first evolved over 150 million years ago, crop plants as we know them have existed for the mere blink of an evolutionary eye.

Although the exact series of steps by which plants were domesticated is unknown, it is likely that strong selection pressure exerted by humans on the genetic diversity found in the wild resulted in rapid and radical changes in plant species (*2*). Certain traits, such as nonshattering of seeds, compact growth habit, or loss of germination inhibition, would have been selected by early agriculturists (*3*). Selective propagation of lines containing these favorable mutations would have resulted in a progressive narrowing of the genetic base of subsequent populations (Fig. 1).

Following domestication, the genetic variation in crop plants has continued to be reduced by another force—modern plant breeding. Over the past century, the development and successful application of plant-breeding methodologies has produced the high-yielding crop varieties on which modern agriculture is based. Yet, ironically, it is the plant-breeding process itself that threatens the genetic base on which breeding depends. Because new varieties are usually derived from crosses among genetically related modern varieties, genetically more variable, but less productive, primitive ancestors are excluded. Soybeans and wheat are good examples of crops with very narrow genetic bases. Virtually all modern U.S. soybean varieties can be traced back to a dozen strains from a small area in northeastern China, and the majority of hard red winter wheat varieties in the United States originated from just two lines imported from Poland and Russia (*4*, *5*).

The limited genetic diversity of crops renders them more vulnerable to disease and insect epidemics and jeopardizes the potential for sustained genetic improvement over the long term (*5*). This risk was brought sharply into focus in 1970 with the outbreak of Southern corn leaf blight. This disease drastically reduced corn yields in the United States and was attributed to extensive use of a single genetic male sterility factor that, unfortunately, was genetically linked to disease susceptibility (*6*).

World Germplasm Resources

Over six decades ago, Vavilov (*2*) first called attention to the potential of crop relatives as a source of genes for improving agriculture. It was this promise that motivated the establishment of gene banks, living seed collections focused primarily on races and species that are closely related to crop plants (collectively referred to as "exotics"). In addition, a report commissioned by the National Academy of Sciences in response to the Southern corn leaf blight disaster recommended placing much greater emphasis on collecting and preserving the genetic diversity in crop species, especially that present in the remaining populations of wild ancestors (*7*). One result of that report was the formation of the International Plant Genetic Resources Institute, which was charged with overseeing worldwide efforts in plant collection and preservation (*8*).

Worldwide, there are more than 700

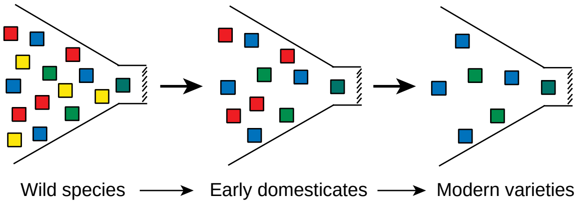

Fig. 1. Genetic bottlenecks imposed on crop plants during domestication and through modern plant-breeding practices. Boxes represent allelic variations of genes originally found in the wild, but gradually lost through domestication and breeding. Such lost alleles can be recovered only by going back to the wild ancestors of our crop species.

The authors are in the Department of Plant Breeding, 252 Emerson Hall, Cornell University, Ithaca, NY 14853, USA.

documented seed collections holding an estimated 2.5 million entries including many exotics (Table 1) (8). The United States alone spends about $20 million per year on germplasm acquisition and preservation, a testament to the importance of these collections (Fig. 2). While it is important to collect and preserve genetic variation in seed banks, these activities are not sufficient to ensure the future productivity of agriculture. The establishment and maintenance of seed banks must be coupled with the ability to actively utilize the materials in those collections. In this regard, we have fallen far short of expectations. While there is a general belief that genes useful for improving crops are contained in these seed banks, the reality is that plant scientists have been largely unsuccessful in finding and extracting such genes. The vast majority of the accessions in gene banks make no contribution to modern varieties, particularly with respect to complex traits such as yield and nutritional quality. As a result, crop improvement is still practiced on a narrow genetic base. DNA-based genetic comparisons of modern cultivars and ancestral wild types found in germplasm banks reflect this situation (Fig. 3).

As we look toward the future, we must reflect upon the rationale for collecting and maintaining germplasm and reexamine the ways in which we are using these genetic resources. Recent applications of genome mapping suggest that the genetic diversity stored in germplasm banks can be utilized with a much higher level of efficiency than previously imagined. However, if we are to discover the true potential of the seeds carefully locked away in gene banks, we must change the way we think about our germplasm resources and be willing to approach the utilization of these living collections in new and innovative ways.

Strategies for the Evaluation and Utilization of Germplasm

The old paradigm: Looking for the phenotype. The traditional approach to the utilization of exotic germplasm is to screen entries from a gene bank for a clearly defined character, recognizable in the phenotype (physical appearance). Once a line with the desired characteristic has been identified, it is crossed with an elite cultivar in order to introduce the genes from the exotic donor into the cultivated type. This approach works well when the trait of interest is controlled by one or a few genes. For example, wild germplasm has been used with great success in breeding for simply inherited resistance to diseases and insects (8). The resistance characters selected by means of this approach are almost always conditioned by single, dominant genes that can be readily transferred into elite cultivars through established breeding procedures. Generally, resistance screening requires inoculating hundreds or thousands of entries with a given pathogen or insect. One or more highly resistant lines are identified and used as donors of the trait in subsequent crossing experiments. Though effective for certain specific characters, only a small proportion of the genetic variation inherent in exotic germplasm will ever be exploited for crop improvement as a result of this strategy.

Most traits important to agriculture, such as yield, are conditioned not by single genes, but by many genes. The yield of wild and unadapted accessions found in germplasm banks is invariably much lower than that of modern elite cultivars. For example, in rice, an accession of the wild ancestor *Oryza rufipogon* yields only a fraction of that of a modern Chinese hybrid (Fig. 4). Crop domestication and plant breeding have been extremely successful at increasing the frequency of beneficial alleles for yield at many loci. As a result, breeders have persisted in making crosses among closely related, high-yielding varieties, unable to rationalize a search for yield-enhancing genes in low-yielding ancestral types. Yet, considering how many genes are likely to influence yield, it is unlikely that modern cultivars have the best alleles for yield at all yield-related loci. Many beneficial alleles have undoubtedly been left behind because of the bottlenecks imposed by domestication coupled with years of modern breeding and selection within adapted gene pools. Thus, although wild and exotic germplasm is perceived to be a poor bet for the improvement of most traits

Table 1. Estimated number of seed bank entries worldwide for selected crops [reprinted from (*8*)].

Crop	Entries	Collections of 200+	Percent wild species
Wheat	410,000	37	60
Rice	215,000	29	10
Maize	100,000	34	15
Soybean	100,000	28	30
Potato	42,000	28	40
Tomato	32,000	28	70
Cotton	30,000	12	20

Fig. 2. Cryopreservation of seed stocks at USDA–Agricultural Research Service (ARS) National Seed Storage Laboratory, Fort Collins, Colorado. [Photo courtesy of USDA-ARS Information Services]

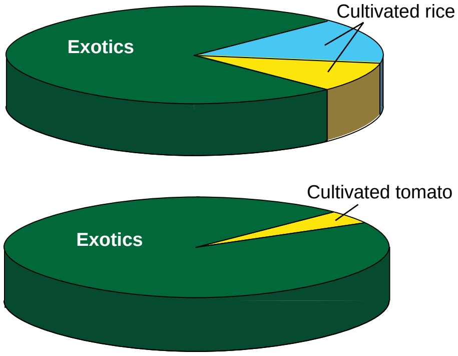

Fig. 3. Pie diagrams depicting proportion of genetic variation (as measured through DNA polymorphisms) found in cultivated rice [*Oryza sativa* spp. indica (blue) and japonica (yellow)] (*9*) and cultivated tomato (*L. esculentum*) (*10*) versus their wild relatives (exotics, green). Exotics include only those lines that cross readily with their domesticated counterparts.

Fig. 4. (Left) Wild rice species *O. rufipogon* from Malaysia. [Photo courtesy of C. Martinez, Centro Internacional de Agricultura Tropical] (Right) Modern rice variety from China. Although wild species are low yielding, they contain genes that can significantly increase the yield of modern rice varieties and provide much needed enrichment of the domestic gene pool.

based on phenotypic examination, it is quite possible that some favorable genes (alleles) lie buried amidst the thousands of accessions maintained in gene banks which, if they could be found, might be of great value to crop improvement. Implementing strategies for finding those genes requires a major shift in the paradigm for using our genetic resources.

The new paradigm: Looking for the genes. Genetic linkage maps based on molecular markers have now been developed for most major crop species, as well as for a number of minor species (http://probe.nalusda.gov). These maps have been important in a number of applications in plant science including the localization of genetic loci that condition agronomically important traits, positional gene cloning, comparative mapping, and marker-assisted selection in breeding (*12*). Perhaps their most profound impact has been in the study of complex, or quantitatively inherited, traits. Molecular linkage maps have made it possible to identify, map, and study the effects of the individual loci that control a quantitatively inherited trait (quantitative trait loci, or QTLs) (*13*). Studies from plants (mainly crop plants) have led to several interesting discoveries. Before the QTL era of quantitative genetics, it was assumed that complex traits were determined by a large number of genes of relatively small and equal effect (*14*). However, QTL analysis has revealed that while most complex traits are controlled by a number of loci, the effects of those loci are not equal. Often a substantial portion of the genetic variation in a population can be explained by a few QTLs of moderately large effects. The second significant discovery was that the phenotype of a plant is, at best, only a modest predictor of its genetic potential. For example, if one line of rice is high yielding and another low yielding, one might assume that the high-yielding type possesses most, if not all, of the genes for high yield and that the low-yielding parent has little or nothing to offer in this regard. However, when populations derived from such crosses are examined with molecular markers and the loci controlling yield are identified, a much different picture emerges. While the high-yielding line often does contain a great number of positive alleles at the loci associated with yield, there are almost always some loci for which the inferior parent contributes a superior allele (*11*, *15*).

The implications of these findings are profound with respect to germplasm utilization. They suggest that using phenotypic evaluation to determine the breeding value of an accession is likely to be misleading, especially with respect to quantitative traits. Thus, we have been screening germplasm in a way that fails to expose its potential. A corollary is that exotic germplasm is a likely source of new and valuable genes capable of increasing yield and other complex traits important to agriculture and that molecular linkage maps will enable us to find them. The paradigm needs to shift away from selecting potential parents on the basis of phenotype to evaluating them directly for the presence of useful genes. The tools that make such an analysis possible are molecular maps and the integrative power of QTL analysis.

Exploitation of Genetic Resources Based on Molecular Maps

Rice and tomato are representatives of the two major classifications of flowering plants. A monocot and member of the grass family, rice is one of a large number of staple cereal crops including wheat, barley, oat, maize, sorghum, and millet. Tomato is a dicot, a classification it shares with a varied assortment of other crops such as potato, cassava, soybeans, sugarbeet, cotton, lettuce, and sunflower.

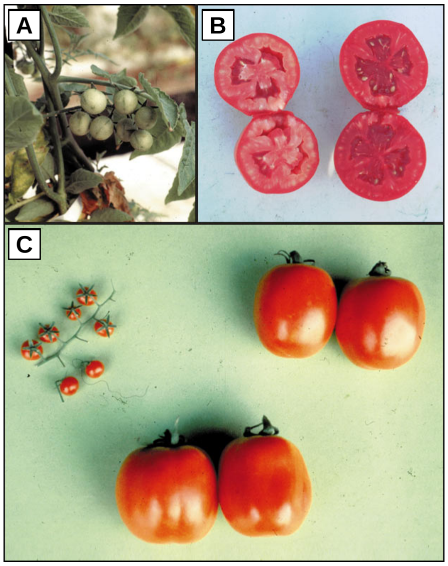

Fig. 5. (**A**) Wild tomato species *L. hirsutum* from Peru that produces small, inedible fruit that does not turn red upon ripening. (**B**) (Left) Fruit from modern processing tomato cultivar E6203. (Right) Fruit from nearly isogenic line (NIL) into which QTL for increased red pigment has been transferred from *L. hirsutum* by the advanced backcross QTL method (*17*). (**C**) (Top left) Fruit from *L. pimpinellifolium* from Peru that produces small berries typical of most fruit-bearing wild species. (Top right) Fruit from modern processing tomato cultivar E6203. (Bottom center) Fruit from NIL into which QTL for increased fruit size has been transferred from *L. pimpinellifolium* by the advanced backcross QTL method. Fruit of this NIL are significantly larger (~10%) than the original E6203 variety (*19*).

A gene-based approach to screening exotic germplasm has been tested in these two species. The strategy involves the use of molecular linkage maps and a breeding technique, referred as to the advanced backcross QTL method, that allows a subset of alleles from the wild or exotic plant to be examined in the genetic background of an elite cultivar (*16*). The molecular linkage map is used to identify the chromosomal position of "wild" alleles that have been transmitted into the progeny, to determine which of the wild species introgressions are associated with superior performance of lines, and to purify lines so that they contain only a specific "wild QTL" in an elite genetic background. The prediction is that the modified elite lines will perform even better than the original elite cultivar.

In tomato, where the most extensive experiments have been conducted, lines have been created that contain specific QTLs from the wild species *Lycopersicon hirsutum* and that outperform the original elite variety by 48, 22, and 33% for yield, soluble solids content, and fruit color, respectively (*17*). Moreover, the performance of these lines has been confirmed under different growing conditions around the world (*17*). The magnitude of these improvements is substantial if we consider that the normal yearly improvement for these traits achieved through traditional breeding is less than 1% per year (*18*). The dramatic improvement in red fruit color (attributable to the pigment lycopene) is especially remarkable considering that the wild tomato lacks an active enzyme for the last step in the pathway and cannot synthesize lycopene (fruits remain green even when ripe) (Fig. 5A). Apparently, the wild tomato contains genes (alleles) that can enhance earlier steps in the biosynthetic pathway leading to lycopene which, when combined with an active form of the gene for lycopene synthesis from cultivated tomato, leads to even higher levels of pigment production in the interspecific offspring (Fig. 5B). Similarly, fruit size has been increased in cultivated tomato lines by the introduction of genes, identified through molecular mapping from the small-fruited ancestor *L. pimpinellifolium* (*19*) (Fig. 5C). These results serve to underscore the point that exotic germplasm often contains genes that are capable of improving traits important to humans. This could not have been predicted from phenotypic evaluation alone, nor could the useful genes be readily discovered without the aid of QTL mapping.

The superior lines described above were created in a two-step process. The QTLs were discovered in an advanced backcross generation (BC2) and, on the basis of that information, new lines were selected containing specific QTL alleles from the wild species. The odds of such favorable genetic combinations occurring by chance and without marker-assisted selection are extremely low.

Thus far in tomato, the genomes of four wild species have been screened by means of the advanced backcross QTL method. While some putatively allelic QTLs have been identified in comparisons among these species, more than 50% of the beneficial QTLs discovered in each experiment appear to be unique to the species tested (*20*). This suggests that continued sampling of wild germplasm is likely to be rewarded with new gene discoveries. It also raises the possibility of creating new crop varieties with a mixture of beneficial genes coming from several different wild ancestors—genotypes that can be efficiently created through molecular breeding.

In rice, conventional hybrid varieties developed at the National Hybrid Rice Research Center (NHRRC) in China outyield the best Chinese inbreds and are among the most productive rice varieties in the world. Experiments were recently conducted to determine whether genes from the low-yielding wild ancestor *O. rufipogon* could significantly enhance yields of one of China's most productive rice hybrids. When the advanced backcross method was used to examine alleles from the wild species in the genetic background of the elite Chinese hybrid, two QTLs were identified that each increased yield ~17% compared with the original hybrid. Moreover, the wild alleles identified in this study appear to be free of many of the negative effects often associated with yield enhancement in cultivated germplasm (*21*). Similar results have been observed in three other experiments conducted independently in Korea and Colombia (*21*).

The results from both tomato and rice indicate that exotic germplasm does contain many new and useful genes that can significantly enhance agricultural production, even for complex traits like yield. They also demonstrate that the phenotype of exotic germplasm is a poor predictor of its genetic potential and that, until we begin applying molecular mapping techniques like the advanced backcross QTL method that allow us to identify genes, it is unlikely that we will realize the benefits to agriculture of most genetic resources.

Strategies for Sampling Exotic Germplasm

Considering the large number of exotic accessions in seed banks and the limited amount of time and resources available, how do we decide which accessions to sample in order to maximize the chance of finding new and useful genes? Realistically, this question can be answered only after further studies have been conducted involving molecular maps and exotic germplasm. However, one principle seems likely to emerge: The sampling of exotic germplasm should emphasize the genetic composition rather than appearance of exotic accessions. With the use of DNA profiles, the genetic uniqueness of each accession in a seed bank, relative to all other accessions, can be determined and quantified (*22*). Accessions with DNA profiles most distinct from that of modern germplasm are likely to contain the greatest number of novel alleles. It is in these accessions that one is most likely to uncover the largest number of unique and potentially agronomically useful alleles. The DNA profiling necessary to make such sampling decisions has already been accomplished or is underway in most crops (*22*).

In tomato and rice, exotic germplasm accessions for use with the advanced backcross QTL method have been selected on the basis of genetic uniqueness with the following two criteria: (i) the exotic accession had to possess a significant number of unique DNA polymorphisms (throughout the genome) relative to modern cultivars and (ii) each new exotic had to be genetically dissimilar (on the basis of DNA profiling) to previously sampled exotics. This strategy has resulted in the sampling of accessions representing the broadest possible spectrum of wild species and races present in the seed banks. With each sampling of a new exotic accession, a high proportion (~50%) of new and useful QTL alleles has been identified (*20*).

As the results accumulate from molecular mapping studies in crop plants, patterns are likely to emerge concerning the chromosomal positions of key loci controlling yield and other important traits. Once this information becomes available, one can imagine sampling exotic germplasm for targeted areas of the genome through backcross breeding and the generation of allelic series in nearly isogenic lines that could then be evaluated agronomically. Such targeted studies can be conducted more rapidly and at less expense than whole-genome studies. Eventually, the genes underlying yield-enhancing QTLs will be cloned, and at that time we can imagine that methods like the polymerase chain reaction will allow allelic variants at these loci to be selectively amplified and sequenced from large numbers of exotic accessions. We can expect that such gene-based searches will ultimately be extended beyond the realm of sexually compatible species so that homologous genes from other genera and families can be tested for optimum performance by a transgenic approach.

Molecular Breeding and Germplasm Utilization in the Future

We are fortunate to be living at a time when genetic engineering holds much promise for modifying crop performance. However, most of the advances thus far in genetic engineering have been directed toward traits other than yield, largely because of the complexity of this trait. The wide repertoire of genetic variants created and selected by nature over hundreds of millions of years is contained in our germplasm banks in the form of exotic accessions. More than 50 years ago, Vavilov predicted the value to agriculture of collecting and maintaining the wild relatives of crop plants in gene banks. Owing to the advent of molecular mapping and the ability to scan the genomes of wild species for new and useful genes, we may now be in a position to unlock the genetic potential of these germplasm resources.

REFERENCES AND NOTES

1. L. R. Brown, *State of the World* (Norton, New York, 1994).
2. N. I. Vavilov, in *The New Systematics*, J. Huxley, Ed. (Clarendon, Oxford, 1940), pp. 549–566.
3. J. R. Harlan, *Crops and Man* (American Society of Agronomy, Madison, WI, 1975).
4. D. N. Duvick, *Ann. N.Y. Acad. Sci*. **287**, 86 (1977).
5. J. R. Harlan, in *Plant Genetic Resources: A Conservation Imperative*, C. W. Yeatman, D Kafton, G. Wilkes, Eds. (Westview, Boulder, CO, 1987), pp. 111–129.
6. A. J. Ullstrop, *Agric. Handb.* **199**, 1 (1978).
7. Anonymous, *Genetic Vulnerability of Major Crops* (National Academy of Sciences, Washington, DC, 1972).
8. D. L. Plunknett *et al*., *Gene Banks and the World's Food* (Princeton Univ. Press, Princeton, NJ, 1987).
9. Z. Y. Wang, G. Second, S. D. Tanksley, *Theor. Appl. Genet*. **83**, 565 (1992).
10. J. C. Miller and S. D. Tanksley, *ibid.* **80**, 437 (1990).
11. J. Xiao *et al*., *Nature* **384**, 223 (1996).
12. R. L. Phillips and I. K. Vasil, *DNA-based Markers in Plants* (Kluwer, Boston, 1994); A. H. Paterson, *Genome Mapping in Plants* (Academic Press, Austin, TX, 1996).
13. S. D. Tanksley, *Annu. Rev. Genet*. **27**, 205 (1993).
14. R. Falconer, *Introduction to Quantitative Genetics* (Ronald, New York, 1993).
15. M. C. DeVicente and S. D. Tanksley, *Genetics* **134**, 585 (1993); Y. Eshed and D. Zamir, *Theor. Appl. Genet*. **88**, 891 (1994).
16. S. D. Tanksley and J. C. Nelson, *Theor. Appl. Genet*. **92**, 191 (1996).
17. D. Bernacchi *et al*., in preparation.
18. S. Grandillo *et al*., in preparation.
19. S. D. Tanksley *et al*., *Theor. Appl. Genet*. **92**, 213 (1996).
20. T. F. Fulton *et al*., *ibid*., in press.
21. S. R. McCouch and J. Xiao, personal communication.
22. S. M. Brown and S. Kresovich, in *Genome Mapping in Plants*, A. H. Paterson, Ed. (Landes, New York, 1996), pp. 85–93; S. Smith and T. Helentjaris, *ibid*., pp. 95–110; A. Karp *et al*., *Nature Biotechnol.* **15**, 625 (1997).

23. We thank A. Frary and E. Earle for many helpful comments. Supported by grants from the U.S. Department of Agriculture (USDA) National Research Initiative–Plant Genome Program, the Binational Agricultural Research and Development Fund, and The Rockefeller Foundation.