

Managing self-pollinated germplasm collections to maximize utilization

Randall L. Nelson*

USDA – Agricultural Research Service, Soybean/Maize Germplasm, Pathology, and Genetics Research Unit, Department of Crop Sciences, 1101 W. Peabody Dr, University of Illinois, Urbana, IL 61801, USA

Abstract

The primary mission of germplasm collections is to preserve genetic diversity, but germplasm is preserved so that it can be used. Historically, the standard practice is that all germplasm accessions should be maintained as collected, so that even self-pollinated accessions are maintained as heterogeneous seed lots. In theory, this seems like an ideal strategy for preserving genetic diversity, but in practice, it is simply not workable. Heterogeneous accessions are in constant risk of change and loss. It is possible to mitigate the risk factors, but they can only be lessened and not eliminated. Maintaining pure-lined accessions for self-pollinated species not only eliminates the problems associated with genetic drift and natural selection, but also enhances the accuracy of the evaluations and fosters effective germplasm utilization. Neither the current potential to characterize entire germplasm collections with tens of thousands of DNA markers nor the future potential of whole genome sequencing to completely characterize the diversity of all accessions in collections can be fully realized for self-pollinated species unless accessions are homogeneous and homozygous. In this manuscript, the case is made for pure-lining self-pollinated germplasm accessions using the USDA Soybean Germplasm Collection, which has maintained pure-lined accessions for over 50 years, as an example. There is also an analysis of the extensive seed distribution from this collection to indicate the value of a diverse collection of genotypes.

Keywords: Glycine; germplasm utilization; germplasm management

Introduction

Recently, Carter *et al.* (2004) provided a thorough review of soybean genetic diversity, and in this issue, Qiu *et al.* (2011) have provided updated and detailed information about research on soybean germplasm. The purpose of this article is not to duplicate those reviews but to use the history of soybean germplasm in the United States to help understand the various transitions that have occurred in the management of genetic resources and the changes that will be needed if we are to meet the challenges of the future. The focus will be on the USDA Soybean Germplasm Collection, but the

concepts are applicable to any self-pollinated species. In a time of rapidly evolving genetic and genomic technologies, germplasm collections can become an even more important resource for plant breeders to develop improved cultivars that will produce the food, feed and fibre that will sustain a growing world population, and have a much greater role in the research of other plant scientists to understand the basic biology of critical plant functions. However, to reach this potential will require changes in how germplasm collections are managed and how those resources are utilized by those that can most benefit from their effective use. Germplasm collections have often been managed based on theoretical principles for maintaining maximum genetic diversity without sufficient consideration as to how that diversity will actually be maintained, evaluated and utilized.

*Corresponding author. E-mail: randall.nelson@ars.usda.gov

The beginning of the USDA Soybean Germplasm Collection

The soybean first arrived in the United States in 1765 (Hymowitz and Harlan, 1983), but neither that introduction nor others that occurred over the next 100 years were preserved or permanently established soybean as a crop in the U.S. The oldest extant accession in the USDA Soybean Germplasm Collection was introduced into the U.S. sometime prior to 1880, but the exact origin and circumstances of that introduction have been lost (Bernard *et al.*, 1987). Formal record keeping of plant introductions began in the U.S. in 1898. This function was assigned to U.S. Department of Agriculture, Division of Botany. The first plant introduction number (PI 1) was assigned to a *Brassica oleracea* (cabbage) accession from Russia. The first soybean recorded in the system (PI 480) came from Siberia in the same year (USDA, 1898). Both of these accessions were collected by Prof. Niels E. Hansen of the Agricultural College of South Dakota during a visit to Russia, central Asia and Siberia. Unfortunately, that soybean accession like most that were introduced through the first half of the 20th century was not preserved. At this time, soybean was not a crop but a scientific curiosity. New introductions were saved if they were perceived to have value, and the others were discarded. Between 1898 and 1949 when the USDA Soybean Germplasm collection (hereafter referred to as the USDA Collection) was established, 8225 soybean introductions were brought into the U.S. More than 5000 of these lines came from the plant collecting expeditions of P. H. Dorsett and W. J. Morse in China, Korea and Japan between 1924 and 1932 (Dorsett, 1927; Bernard *et al.*, 1987; Piper and Morse, 1910, 1923, 1925). As important as these collections were, only a very few of the major ancestral lines that formed the genetic base of current U.S. soybean production came from these expeditions. Based on the ancestral contributions calculated by Gizlice *et al.* (1994), the Dorsett and Morse expeditions provide about 13% of the current genetic base, and most of that (8%) was contributed by the major ancestral line Richland collected in Jilin, China in 1926. In 1949, only 1677 of these soybean accessions and other lines introduced prior to 1898 were maintained in individual U.S. research programmes (Bernard *et al.*, 1987), and those accessions became the initial USDA Collection. This advanced the concept that germplasm should be preserved even when the immediate value was unknown. Beginning in 1949, soybean germplasm introduced into the U.S. was carefully preserved.

Rationale for pure-lining self-pollinated accessions

In 1954, Dr. Richard Bernard was appointed curator of the USDA Northern Soybean Germplasm Collection. He made

two observations that produced significant changes in how the USDA Collection was managed (R. L. Bernard, personal communication). Original seed samples had been preserved at Beltsville, MD for early soybean introductions. These samples had not been stored in a controlled environment, and most were decades old so, by the 1950s, none were viable. However, it did allow for a physical comparison of seeds from the original sample with those currently in the USDA Collection. Based on seed coat and hilum colour, and seed size and shape, Dr. Bernard determined that many of the current seed lots were different from the original samples. The accessions that are different were recorded and are reported in Bernard *et al.* (1987). The second observation also involved changes in accessions. The names associated with some Japanese accessions in the USDA Collection were descriptive. Based on those names, some of the accessions should have been glabrous, but by the mid 1950s, most of them were pubescent. These observations highlight two critical aspects of managing a germplasm collection. There could have been many reasons why some of the germplasm accessions were different from the original sources, but a key factor in these errors is the lack of an accurate description for each accession. Without these descriptions, errors in the seed lots were not detected and corrected, and accessions were irrevocably changed. It is possible that could explain the change in the glabrous accessions as well, except that there is a more likely explanation. In the midwest U.S., the potato leafhopper (*Empoasca fabae*) causes severe damage to soybeans without pubescence and greatly reduces plant growth and seed yield. We demonstrated this effect by planting equal quantities of seeds of near-isogenic soybean lines that differed only at the *PI* locus (Nagai and Saito, 1923; Palmer *et al.*, 2004), which controls the glabrous trait, in a mixed population. The stunting effect of the leafhoppers was compounded by the shading of the more vigorous pubescent plant. Natural selection removed all of the glabrous plants after three cycles of bulk harvesting the plots and replanting that seed (unpublished results). Because of these experiences, Bernard concluded that the optimum strategy for maintaining self-pollinated species in a germplasm collection was to maintain all accessions as pure lines.

Historically, the standard practice is that all germplasm accessions should be maintained as collected so as not to lose any of the variation in the original sample. In theory, this seems like an ideal strategy for preserving genetic diversity, but in practice, it is simply not workable. Heterogeneous accessions are in constant risk of change both during storage (from unequal seed longevity) and during regeneration (from natural selection, genetic drift and mechanical contamination) and as was previously cited those changes can lead to permanent loss of

diversity. It is possible to mitigate each of these risk factors but doing so can greatly increase the costs of curation, and in reality, the risks can only be lessened and not eliminated. Maintaining separate samples of original seeds to use for future regeneration has been suggested as a way of reducing genetic drift and natural selection, but it is impractical because original samples are generally much too small to provide the needed reserve. Even if large samples were obtained that would increase storage costs, not guarantee accession integrity, and only be viable for a finite period of time until the original sample is exhausted or dies. Assuming that all genotypes will produce some seeds each time the accession is grown, the effects of natural selection and genetic drift could be eliminated by storing equal number of seeds from each plant in the regeneration plot. This also assumes that each time the accession is grown, the sample size is sufficiently large to include even the very low frequency genotypes that may exist in heterogeneous accessions. This could be a practical solution for very small germplasm collections, but the labour costs required to implement such a strategy would be prohibitively high for any major collection. None of these strategies addresses the concern of accidental seed contamination. It is not practically possible to accurately describe a heterogeneous accession. When seed contamination occurs, it is highly unlikely that it will be recognized and remedied, and, no matter what precautions are taken to ensure seed purity, accidental seed contamination or mislabelled seed lots will occur in large germplasm seed collections.

In pure-lined collections, each accession is descended from a single seed in the original seed lot. This requires careful classification of each plant grown from the original seed lot. At maturity, multiple single plants are harvested to represent each identified phenotype. The following year, each plant row is again characterized, and each row with a different phenotype is harvested and added as a new accession. Rows that are segregating are discarded based on the assumption that these rows are the product of a recent cross pollination and both parents would be available. There are certainly risks associated with pure-lining accessions. Practically, pure lining can only select on phenotypic variation, so it is always possible that genotypic variation could be overlooked and lost. Because it is possible to select on both qualitative traits that are generally easily classified and quantitative variation that can be distinguished even when it cannot be precisely described, the odds of genotypic differences not being expressed in some phenotypic variation are low. Although wild soybean (Fujita *et al.*, 1997) can have higher levels of cross pollination than cultivated soybean (Ahrent and Caviness, 1994), both species are highly self-pollinated, so achieving homozygosity is a natural process. For domesticated

species such as soybean, primitive varieties are commonly derived by pure line selection, and heterogeneity is as likely to have come from accidental contamination (perhaps in a previous germplasm collection) as from inherent variation. For wild species, populations should be extensively sampled to effectively capture the natural variation in multiple accessions (Zhu *et al.*, 2007). Approximately, 75% of the accessions in the USDA Collection were phenotypically homozygous and homogeneous when they were received. Although there is a risk that the pure lines will not preserve all of the variation in the original sample, this risk occurs once, and then the integrity of each accession can be economically and predictably maintained forever. Genetic drift and natural selection are not factors in compromising the integrity of the accession, and accidental seed contamination can be detected and removed because each accession has a precisely known description.

Growth of the USDA Soybean Germplasm Collection

Beginning in 1949, all soybean germplasm introduced into the U.S. was carefully preserved, but for the next 25 years, there was little active effort to expand the collection. A major turning point for germplasm collections, in general, in the U.S. was the southern leaf blight epidemic (caused by *Helminthosporium maydis*, Nisikado & Miyake) (Sprague, 1971) of 1970 in maize. This publicly highlighted the perils of genetic uniformity of crops in modern agriculture and increased the visibility of and support for germplasm programmes (Committee on Genetic Vulnerability, National Academy of Science, 1972; Committee on Germplasm Resources, 1978). This period also coincided with an impetus to intentionally increase the holdings of the USDA Collection with a particular focus on Asia. From 1949 to 1974, an average of 72 accessions was added to the USDA Collection each year. From 1975 to 2000, this number increased to over 500 annually. Some of these accessions were collected by U.S. scientists in cooperation with host countries, but nearly 90% of the new additions were obtained through exchanges with other collections. Worldwide, the era of soybean collection is coming to an end. There are only a few places in Asia where primitive varieties, not currently in any germplasm collection, are being grown by farmers, but these isolated pockets are rapidly disappearing, as cultivars from new and expanding scientific breeding programmes are being widely distributed and grown. This loss of genetic diversity that has been previously managed by farmers increases the responsibility of *ex situ* collections and the importance of germplasm exchange among these collections.

A major remaining challenge is the collection of wild soybean, *Glycine soja*. Phenotypically, the wild soybean has less variation than the cultivated soybean in plant and seed pigments, seed size, pubescence characteristics and plant type. Although there are fewer variations, there are significant differences among wild soybean accessions. Significant differences for mean seed size (1.9 vs. 2.8 g/100 seeds) were found among populations in southwestern Hokkaido (Ohara *et al.*, 1989). Morphological differences among *G. soja* populations may be related to adaptation to specific environments as two distinct types, twining and branching, were found in contrasting ecological niches on the Saru River in Hokkaido (Ohara and Shimamoto, 1994). There are some phenotypic traits for which the wild soybean does have more variation. Chen and Nelson (2004a) identified much greater diversity in leaflet size and shape within wild soybean accessions from Russia, China, Japan and Korea, and they found that variation was associated with geographical origin. They also found large differences in early plant growth not observed in soybean (Chen and Nelson, 2006). Although there are mixed results for measures of phenotypic diversity, all reported research clearly shows that the wild soybean is genotypically much more diverse than soybean. Both Li and Nelson (2002) and Chen and Nelson (2004b) using Random Amplified Polymorphic DNA markers identified much greater genetic diversity in wild soybean than in soybean. Using variation within gene sequences, Hyten *et al.* (2006) concluded that approximately half of the diversity in wild soybean was not transferred through the genetic bottleneck of domestication. Certainly, some of that loss is responsible for the improved characteristics that make soybean an agronomic crop, but it is highly likely that valuable alleles for soybean improvement exist within this species. There is ample evidence that there is great genetic diversity within and among wild soybean populations (Kiang *et al.*, 1992; Dong *et al.*, 2001; Lee *et al.*, 2008; Wen *et al.*, 2009; Li *et al.*, 2010). The total *ex situ* collection of wild soybean is only a small percentage of the worldwide soybean germplasm holdings and should be increased. With the current development of high throughput systems for single-nucleotide polymorphism (SNP) markers and the potential of extensive accession sequencing in the future, extraction of useful alleles from wild soybean will very likely become much easier if not routine in the future.

Germplasm evaluation

Germplasm evaluation is the critical first step in effective germplasm utilization. Without evaluation, germplasm collections have limited utility and unless they are used

or have the potential to be used, they have no practical value. Pure-lined accessions have an enormous advantage for evaluation. It is not possible to accurately evaluate heterogeneous accessions because of the uncertainty in separating genetic from environmental variation. As a very simple case, we find that hilum colour expressivity in some germplasm accessions is much more variable than in standard soybean cultivars. For some accessions, there can be a wide range of hilum colour among seeds within a seed lot. If the extremes are selected and planted as two seed lots, the same variation will be found within each harvested seed lot. This same variation is very likely to be observed for economically useful traits such as disease resistance. Unless one is certain that each seed lot is homozygous and homogeneous, this type of variation would most likely be classified as genetic differences. If useful variation is identified within a heterogeneous accession, it must be preserved as a pure line if it is to be used. It is very inefficient to have to locate the needed genotype within a mixed seed lot each time it is needed. In reality, only pure-lined accessions, either intentionally or by default, can be used because germplasm is introgressed one ovule or pollen grain at a time.

An argument against pure lining is that rare alleles are likely to be lost in the initial selection. This is a valid argument; however, maintaining and finding rare alleles in heterogeneous accessions is also problematic. We have already discussed the negative impact that genetic drift and natural selection can have on heterogeneous populations and particularly on rare alleles. If rare alleles are maintained within heterogeneous populations, there still exists the major challenge of finding them. If there were an allele with a frequency of 5% in a heterogeneous accession, one would have to evaluate 90 plants to have a 99% chance of finding that allele, assuming it could be definitively identified on a single plant. With an allelic frequency of 1%, 490 plants would need to be evaluated. If nothing is known of the allelic frequencies within accessions or whether or not an accession is actually heterogeneous, as is most likely to be the case with most self-pollinated accessions, it would not be cost effective to evaluate 100–500 plants/accessions in an attempt to find a potential rare allele. This would be the best case scenario for qualitative traits that could be reasonably identified on a single plant. The problem would become much greater for quantitative traits where replicated trials would be needed. Rare alleles are not likely ever to be identified in heterogeneous accessions, because even if they survive natural selection and genetic drift, the cost of finding them is prohibitive.

All of the accessions in the USDA Collection are evaluated for basic agronomic traits, major seed composition components and descriptive characters.

These data are critically important in providing an accurate description of each accession and helping to select accessions that will be most useful in a research or breeding programme. These data are published in technical bulletins (Nelson *et al.*, 1987; Nelson *et al.*, 1988; Coble *et al.*, 1991; Bernard *et al.*, 1998; Hill *et al.*, 2001; Hill *et al.*, 2005; Hill *et al.*, 2008; Peregrine *et al.*, 2008). Many accessions are also evaluated for important disease and insect resistance traits, amino-acid composition, and some abiotic stresses. All of the data collected are available through the Internet at <http://www.ars-grin.gov/npgs/>. There are more than 700,000 data points currently in our database. A core collection was recently selected for the USDA Collection (Oliveira *et al.*, 2010), and this can be used as a guide for selectively testing additional soybean accessions.

Not many years ago, evaluation data were considered a major limiting factor in utilizing germplasm collections, but for many collections, additional evaluation data are no longer the limiting factor in making the collection useful. For example, we have evaluation data for over 6800 accessions for resistance to *Fusarium virguliforme*, which causes sudden death syndrome (SDS). Even though there are fewer than 100 accessions with moderate resistance, it is questionable whether additional resistance sources would be useful at this time unless accessions with immunity or near immunity could be found. For plant breeders, this is a difficult disease for which to breed host plant resistance. Resistance is quantitative (Meksem *et al.*, 1999; Iqbal *et al.*, 2001; Njiti *et al.*, 2002; Farias Neto *et al.*, 2007), producing consistent symptoms in the field is very difficult (Scherf and Yang, 1996), and although more consistent results can be obtained in the greenhouse, those data can be different from those collected in field trials (Njiti *et al.*, 2001; Farias Neto *et al.*, 2008). Most breeders are going to be very hesitant to cross with a new source of SDS resistance that is most likely to come from an agronomically inferior line unless there is good evidence that this line can provide new alleles for resistance. Without that knowledge, the breeder may spend years developing a new resistant, high-yielding line that in the end will have the same resistance alleles as are already available in elite cultivars or improved germplasm. Characterizing genotypic differences among phenotypically similar accessions is a much more challenging activity than identifying additional sources of resistance, but it should be a high priority activity because it is critical for effective utilization of germplasm collections. An excellent example of what can be done is represented by the work of Tian *et al.* (2010). Using a Basic Local Alignment Search Tool (BLAST) search with a known

sequence from an *Arabidopsis thaliana* gene that affects inflorescence type (*TFL1*) against the soybean whole genome sequence (Schmutz *et al.*, 2010), four gene models were identified that were homologous to *TFL1*. Using the known map position of *Dt1* soybean locus, Glyma19g37890.1 was proposed as the candidate gene for the *Dt1*. Inserting that allele into *Arabidopsis* duplicated the function of *TFL1*. Sequencing that gene from a diverse selection of soybean and wild soybean germplasm identified four single base changes. Three of the changes corresponded to known phenotypic differences, but knowing these genotypes allows for classification of stem termination types that could not be separated phenotypically because of genetic background effects. As more agronomically important genes are identified and sequenced and technological advances reduce the cost of this research, finding useful, new diversity through sequence variation will become a powerful tool for germplasm utilization.

DNA markers are an essential tool for germplasm management. Not many years ago, it would not have been financially feasible to consider characterizing entire germplasm collections at a significant number of marker loci, but that has changed. The United Soybean Board in the U.S. is currently investing approximately 3.5 million U.S. dollars to characterize all of the annual accessions in the USDA Collection with more than 40,000 SNP markers. This is a cost of about \$175/accession, and the more than 800 million data points that this research will generate will provide an unprecedented resource for germplasm research. Re-sequencing several soybean germplasm accessions is already being discussed, and it is very likely that some time in the not too distant future, the cost of a complete genome sequence for a plant accession may not be much more expensive than the current characterization with SNP markers. Genotypic characterization is enormously useful for understanding and utilizing germplasm, but for self-pollinated species, unambiguous results can only be obtained from pure-lined accessions.

Germplasm distribution

The ultimate value of any germplasm collection is determined by how the accessions in that collection are used. The USDA Soybean Germplasm Collection is one of the most intensely used collections in the world. Of all of the collections in the USDA National Plant Germplasm System, it is the only collection that has an annual average distribution rate that is greater than the total number of accessions in the inventory. Figure 1 shows the growth of the USDA Collection and the annual sample distribution since 1991 when the portions of the

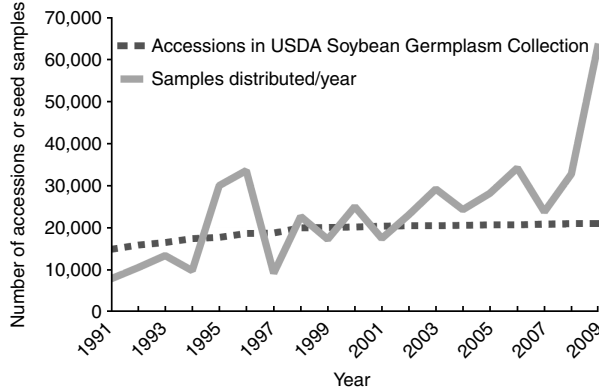


Fig. 1. Growth of the USDA Soybean Germplasm Collection since 1991 and the number of seed samples distributed/year.

USDA Soybean Germplasm Collection maintained in Stoneville, MS and in Urbana, IL were consolidated into a single collection. Over the past 15 years, the USDA Collection has an average distribution of 25,809 seed packets/year, while the average size of the Collection during that time is only 19,536 accessions. During that period, we have distributed seeds of 99% of the accessions in collection. More than 75% of the 211 accessions that have not been distributed are perennial *Glycine*, and the remaining are accessions recently added to the Collection. The most requested accession is the old cultivar Peking, the first widely used source of soybean cyst nematode (SCN) resistance. It has been requested 607 times. Only two other accessions have been requested more than 500 times: Williams 82, the line recently used to create the first soybean genomic sequence, and PI 88788, the currently most widely used source of SCN resistance in the U.S. There were 26 accessions requested more than 200 times, and 18 of those lines are U.S. cultivars. The remaining eight accessions include six sources of SCN resistance and the sources of *Rpp1* and *Rpp2*, Asian soybean rust resistance alleles. All of the highly requested cultivars were released more than 20 years ago, and some are 75 years old. The most often requested wild soybean line is PI 468916, which was a parent of the population that was used to create the first soybean linkage map using DNA markers (Keim *et al.*, 1990). Over 790 accessions have been requested an average of three times/year for the past 15 years. Fifty-nine percentage of the USDA Collection has been requested an average of once a year, and 97% of the USDA Collection has been requested more than once during the past 15 years.

To effectively manage a germplasm collection, it is important to understand the needs of the users, and to understand the needs of users, it is instructive to examine who is requesting germplasm and what types of germplasm are being requested. The USDA Collection is divided into eight sub-collections as follows: introduced

G. max, *G. soja*, germplasm releases, old cultivars, modern cultivars, private cultivars, isolines, genetic types and perennials. The introduced *G. max* and *G. soja* sub-collections are self-explanatory. The old cultivars are a group of introduced *G. max* that were given names in the U.S. or were selected in the U.S. but have unknown parentage. Some, but not all, were used as early commercial cultivars. The modern cultivars are U.S. cultivars with known pedigrees released by public institutions beginning in the 1940s. This collection of 523 cultivars includes almost all publicly released U.S. soybean cultivars with the exception of some cultivars that cannot be distributed without restrictions. The private cultivar collection was initially formed with cultivars that were farmer selections from the late 1940s through the 1960s. About 20 years ago, the decision was made to preserve cultivars developed by private industry that met the following criteria: (1) has been among the leading varieties within a maturity group for total acreage for several years, has some unique feature that merits preservation or has been a significant part of published research, (2) has a publicly available pedigree, (3) can be distributed and used without restrictions, (4) has the approval of the organization that developed the cultivar and (5) is no longer sold as a commercial variety. Approximately, 60 cultivars have been added to this sub-collection in less than 20 years. Germplasm releases that are registered in Crop Science and can be distributed without restriction are maintained in the collection. Since many of the releases have a limited time of usefulness, each release is evaluated every 10 years. Those that have not been requested are no longer retained in the active collection but are still maintained in the National Center for Genetic Resources Preservation at Fort Collins, Colorado. The isoline sub-collection is largely a collection of near-isogenic lines that were developed through backcrossing using principally three recurrent parents (Harosoy, Clark and Williams). These lines have been used extensively in the genetic study of qualitative traits and essentially discovering the first Quantitative trait loci (QTL) in soybean by isolating major genes affecting quantitative traits through backcrossing. The discovery of major genes affecting the time of flowering and maturity is a notable example (Bernard, 1971). These lines have also been very useful in determining the impact of allelic substitutions especially for disease resistance and seed composition traits. The genetic type collection contains mutants not known to occur in other lines within the collection. Approximately, 80% of the entries in this sub-collection have been genetically characterized.

Examining the distribution patterns of each of the sub-collections provides important information about the needs of germplasm users. There are a variety of statistics that can be used in this process. To compare

the usage of the various types of germplasm, the average number of times each accession with the sub-collection had been requested, and the average time that the accessions in the sub-collection had been available during the past 15 years was calculated (Table 1). The latter is needed since accessions are continually being added to the USDA Collection and germplasm releases are being dropped from the active collection when they are no longer requested. The introduced soybean and wild soybean sub-collections averaged more than one request/accession per year. The much smaller wild soybean collection is likely a factor in the slightly higher demand/accession for wild soybean than for soybean accessions. The modern and old cultivars are being requested a rate more than double the next most requested category of germplasm with each cultivar being requested three times/year on average. All of the other specialty collections have very similar rates of requests and given that many of the entries in these collections would have only very specific uses, the high rate of usage is somewhat surprising. The least requested sub-collection is that of the perennial species.

Over the past 15 years, the number of seed lots distributed from each sub-collection is generally similar to the number of accessions in each sub-collection (Table 1). The introduction of soybean collection is 81% of the whole collection and 81.7% of the seed lots distributed. Most of the smaller sub-collections (isolines, private, germplasm releases and genetic types) make up slightly smaller percentages of the distributions than percentages of the USDA Collection. The percentage of perennial accessions distributed is much lower than the percentage in the USDA Collection. This lack of activity is not surprising since few research programmes are actively working with perennial *Glycine* species. These diverse species will very likely be an important source of genetic diversity for soybean breeders in the future. Wild soybean accessions make

up a slightly higher percentage of the requests than of the total accessions in the USDA Collection. As was demonstrated with the average distribution rate, the two small collections that are exceptions are the old and modern varieties, which have distribution percentages more than twice the size of the sub-collections. These distribution numbers indicate the importance of having a source of seeds for these types of accessions for research purposes. Gene bank should not be an appropriate alternative name for germplasm collection, because in many cases, we are preserving genotypes and not just genes.

A final way of analyzing distribution records is to determine who is requesting germplasm (Table 2). Although we keep more detailed records, for this manuscript, the requestors are divided into six categories: foreign commercial companies, foreign public institutions, domestic commercial companies, domestic institutions, U.S. government agencies and unaffiliated individuals. Over the past 15 years, 25% of the accessions distributed go outside the United States. Of those foreign distributions, 5% went to 58 private companies and 95% of the seed lots went to 239 public institutions. Within the U.S., 76% of the seed lots distributed went to public institutions and 24% to private industry. Slightly more than half of the public distributions go to scientists within the Agricultural Research Service (ARS) of the United State Department of Agriculture and the remainder to 156 colleges and universities, and 33 other public institutions. Twenty-four percentage of the domestic distributions went to 145 commercial companies within the U.S. We had 130 individuals not associated with public institutions or private research companies that requested seeds, but the total number of seed lots sent to these people was only 0.3% of the total.

Each of these categories of users has a different need for germplasm. We can assess differences in requesting

Table 1. Number of accessions in each sub-collection of the USDA Soybean Germplasm Collection, the average and total number of requests and the average number of years accessions in each sub-collection were available between 1995 and 2009

Sub-collection	No. of accessions	Collection (%)	Samples distributed	Requested samples (%)	Average request/accession	Average number of years accessions were available
Introduced <i>G. max</i>	17,037	81.0	316,301	81.7	19.3	14.4
<i>G. soja</i>	1178	5.6	25,110	6.5	21.7	14.4
Modern cultivars	523	2.5	22,140	5.7	45.1	13.6
Old cultivars	208	1.0	9053	2.3	45.4	15.0
Private cultivars	75	0.4	773	0.2	11.2	14.5
Genetic types	197	0.9	2008	0.5	10.8	13.9
Germplasm releases	185	0.9	1820	0.5	11.2	9.0
Isolines	640	3.0	6480	1.7	11.8	15.0
Perennial species	1003	4.8	3453	0.9	3.8	13.6
Total	21,046		387,138			

Table 2. Categories of requestors, and types and numbers of germplasm accessions requested between 1995 and 2009

Category	Requestors		Accessions distributed in each sub-collection (n and %)										
	Individuals (n and %)	Institutions or agencies (n and %)	Isolines	G. max	Modern cultivars	Old cultivars	Perennial species	Private cultivars	Germplasm releases	G. soja	Genetic types	Total	
Foreign commercial	82 (5%)	58 (9%)	118	4437	702	95	94	7	80	158	15	5706	
Total distribution			2%	1%	3%	1%	3%	1%	4%	1%	1%	1%	
Category distribution			2%	78%	12%	2%	2%	0%	1%	3%	0%	100%	
Foreign public	354 (20%)	239 (39%)	2203	63,387	5028	1118	832	260	625	4996	656	79,105	
Total distribution			34%	20%	23%	12%	24%	34%	34%	20%	33%	20%	
Category distribution			3%	80%	6%	1%	1%	0%	1%	6%	1%	100%	
U.S. government	165 (9%)	4 (0.5%)	1554	94,402	5239	2643	1062	139	421	9538	577	115,575	
Total distribution			24%	30%	24%	29%	31%	18%	23%	38%	29%	30%	
Category distribution			1%	82%	5%	2%	1%	0%	0%	8%	0%	100%	
Other U.S. public	717 (40%)	190 (29%)	1571	91,660	7051	3364	1100	138	271	7466	491	113,112	
Total distribution			24%	29%	32%	37%	32%	18%	15%	30%	24%	29%	
Category distribution			1%	81%	6%	3%	1%	0%	0%	7%	0%	100%	
U.S. commercial	343 (19%)	145 (22%)	1025	61,854	3709	1757	346	223	406	2936	237	72,493	
Total distribution			16%	20%	17%	19%	10%	29%	22%	12%	12%	19%	
Category distribution			1%	85%	5%	2%	0%	0%	1%	4%	0%	100%	
U.S. individuals	130 (7%)	0 (0%)	9	561	411	76	19	6	17	16	32	1147	
Total (%)			0%	0%	2%	1%	1%	1%	1%	0%	2%	0.3%	
Category (%)			1%	49%	36%	7%	2%	1%	1%	1%	3%	100%	
Total	1791	636	6480	31,6301	22,140	9053	3453	773	1820	25,110	2008	387,138	

patterns, by comparing the percentage of distributions of each class of accessions with the percentage of total distributions that were sent to each class of requestors. Those instances where requests for specific categories of accessions were much larger or smaller than the overall percentage of requests made are highlighted. The foreign commercial companies requested only 1% of the seed lots distributed, but 3% of all modern cultivar requests came from this group and it represented 12% of the seed lots sent to these companies (Table 2). Approximately, 20% of our distributions went to foreign public institutions, but one-third of the requests for isolines, privately development cultivars, germplasm releases and genetic type accessions came from these institutions. ARS scientists and other public institution scientists in the U.S. each account from about 30% of the seed lot distributions. ARS scientists requested 38% of all wild soybean seed lots distributed but only 18% of the privately developed cultivars. University and other public institution scientists requested 37% of the old cultivars that were distributed but only 18 and 15% of the privately developed cultivars and germplasm releases, respectively. Because developers of germplasm releases provide most of the domestic distribution for the first 5 years after release, this statistic most certainly underestimates the demand for this type of germplasm. Private industry in the U.S. received 19% of all distributions. They requested only 10% of the perennial accessions, but 29% of the privately developed cultivars distributed. Germplasm collections meet different needs for different users. For many users having access to specific genotypes that are not available from any other source is critically important. Maintaining all aspects of the genetic resources of a crop is needed to fully meet the expectations of the great variety of users of a modern germplasm collection.

The users of the USDA Collection have changed dramatically over the past 40 years. Research on soybeans in private industry was essentially non-existent in 1970, so users were primarily the few soybean breeders, pathologists and occasionally physiologists at public institutions. As recently as 1991, in the first year, when the USDA Southern and Northern Soybean Germplasm Collections were consolidated, only 7803 seed lots were distributed (Fig. 1). Over the past 15 years, we have distributed seeds to 1791 individuals (Table 2). Although we try to keep our records to a single contact/laboratory, our efforts are not perfect, and some of these individuals are graduate students, technicians and post-docs that are working in the same laboratory. Reporting multiple persons/laboratory is most likely to occur in the statistics of domestic public institutions. However, these nearly 1800 individuals do represent over 200 private companies, and 429 public research organizations, government

agencies and universities. In 2008, there were only 70 countries with soybean production of greater than 1000 metric tons (http://www.geohive.com/charts/ag_soybean.aspx, verified 13 September 2010), and in 2009, there were 31 states in the United States with reported soybean production (http://www.soystats.com/2009/page_15.htm, verified 13 September 2010). We have distributed soybean germplasm to 69 foreign countries and all 50 states within the United States. The users of soybean germplasm now go far beyond the traditional agricultural scientists.

Conclusions

Maintaining heterogeneous germplasm accession for self-pollinated species has long been the standard for germplasm collections. In theory, this seems like a good strategy, but in practice, it has many shortcomings. Genetic drift and natural selection will alter the composition of these accessions and inevitably lead to loss of diversity unless expensive and time-consuming sampling measures are instituted each time the accession is grown. Without an accurate description of each accession, accidental seed or pollen contamination is unlikely to be found and removed. Rare alleles, if they survive, will only be found with testing of large numbers of individual plants from each accession, and evaluation for quantitative traits will be an even greater challenge. Collections of heterogeneous accessions will be excluded for taking advantage of the current capacity to extensively use DNA markers and eventually whole genome sequencing. Heterogeneous accessions of self-pollinated species may have been the standard of the past, but pure-lined accessions will be the standard for the future.

The USDA Soybean Germplasm Collection has been maintaining pure-lined accessions for over 50 years. It is the most intensively used collection in the USDA National Plant Germplasm System with an average distribution of 132% of the accessions in inventory/year over the past 15 years. Analysis of these distribution records shows the great diversity of clientele that utilize this collection and the varied needs of these germplasm users. Users of the USDA Soybean Germplasm Collection are no longer limited to traditional agricultural scientists. All users expect accessions to be consistent from year to year, so that research results can be repeated even with procedures that can be done on a single seed. To meet the needs of a very diverse research community, it is important to maintain not only extensive collections of land races and wild species but also commercial cultivars, genetic stocks and improved germplasm

releases. All of these accessions play a vital role in the research to understand and use genetic diversity.

There are increasing trends to restrict the exchange of genetic resources within the U.S. and internationally, but the USDA has maintained a policy of free exchange for all germplasm within the National Plant Germplasm System. As *ex situ* collections become the only source of soybean germplasm in future, exchange among collections will become more important. It is imperative that these collections are well maintained and accurately described if this critically important diversity is to be preserved and the integrity of each accession is retained. This high level of quality control is necessary for the maximum usefulness of germplasm exchange. It will allow for accurate data as well as correct genotypes to be moved between collections and from collections to soybean scientists. As new technology allows a greater understanding of the genetic diversity among germplasm accessions, the challenges created by the enormous diversity available will likewise increase. The great genetic challenge of the future will be to identify the allelic diversity within the species, and perhaps related species, for the loci affecting economically important traits of soybean. Accomplishing such a task will take more and not less cooperation. The future of soybean breeding and much of soybean research is in our germplasm collections, and new genetic technologies will continue to increase the value of these collections.

References

- Ahrent DK and Caviness CE (1994) Natural cross-pollination of twelve soybean cultivars in Arkansas. *Crop Science* 34: 376–378.
- Bernard RL (1971) Two major genes for time of flowering and maturity in soybeans. *Crop Science* 11: 242–242.
- Bernard RL, Juvik GA and Nelson RL (1987) USDA Soybean Germplasm Collection Inventory. Vol. 1. International Agricultural Publications. INTSOY Series number 30.
- Bernard RL, Cremeens CR, Cooper RL, Collins FI, Krober OA, Athow KL, Laviolette FA, Coble CJ and Nelson RL (1998) Evaluation of the USDA Soybean Germplasm Collection: Maturity Groups 000 to IV (FC 01.547 to PI 266.807). U.S. Department of Agriculture Technical Bulletin no. 1844.
- Carter TE, Nelson RL, Sneller CH and Cui Z (2004) Genetic Diversity in Soybean. In: Boerma HR and Specht JE (eds) *Soybeans: Improvement, Production, and Uses*, vol. Agronomy. Madison, WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, pp. 303–416.
- Chen Y and Nelson RL (2004a) Evaluation and classification of leaflet shape and size in wild soybean. *Crop Science* 44: 671–677.
- Chen Y and Nelson RL (2004b) Genetic variation and relationships among cultivated, wild, and semi-wild soybean. *Crop Science* 44: 316–325.
- Chen Yiwu and Nelson RL (2006) Variation in early plant height in wild soybean. *Crop Science* 46: 865–869.
- Coble CJ, Sprau GL, Nelson RL, Orf JL, Thomas DI and Cavins JF (1991) Evaluation of the USDA Soybean Germplasm Collection: Maturity Groups 000 to IV (PI 490.765 to PI 507.573). USDA Technical Bulletin no. 1802.
- Committee on Genetic Vulnerability of Major Crops (1972) *Genetic Vulnerability of Major Crops*. Washington, DC: National Academy of Science.
- Committee on Germplasm Resources (1978) *Conservation of Germplasm Resources, An Imperative*. Washington, DC: National Academy of Science.
- Dong YS, Zhuang BC, Zhao LM, Sun H and He MY (2001) The genetic diversity of annual wild soybeans grown in China. *Theoretical and Applied Genetics* 103: 98–103.
- Dorsett PH (1927) Soybeans in Manchuria. *Proceedings of American Soybean Association* 1: 173–176. Publ. 1928.
- Farias Neto AL, Hasmi R, Schmidt M, Carlson SR, Hartman GL, Li S, Nelson RL and Diers BW (2007) Mapping and confirmation of a new sudden death syndrome resistance QTL on linkage group D2 from the soybean genotypes PI 567374 and “Ripley”. *Molecular Breeding* 20: 53–62.
- Farias Neto AL, Schmidt M, Hartman GL, Li S and Diers BW (2008) Inoculation methods under greenhouse conditions for evaluating soybean resistance to sudden death syndrome. *Pesquisa Agropecuaria Brasileira* 43: 1475–1482.
- Fujita R, Ohara M, Okazaki O and Shimamoto Y (1997) The extent of natural cross-pollination in wild soybean (*Glycine soja*). *Journal of Heredity* 88: 124–128.
- Gizlice Z, Carter TE Jr and Burton JW (1994) Genetic base for North American public soybean cultivars released between 1947 and 1988. *Crop Science* 34: 1143–1151.
- Hill JL, Peregrine EK, Sprau GL, Cremeens CR, Nelson RL, Kenty MM, Kilen TC and Thomas DA (2001) Evaluation of the USDA Soybean Germplasm Collection: Maturity Groups VI (VIII (FC 03.659 to PI 567.235B)). U.S. Department of Agriculture Technical Bulletin no. 1894, p. 147.
- Hill JL, Peregrine EK, Sprau GL, Cremeens CR, Nelson RL, Orf JH and Thomas DA (2005) Evaluation of the USDA Soybean Germplasm Collection: Maturity Groups 000–IV (PI 507670–PI 574486). U.S. Department of Agriculture Technical Bulletin no. 1914, p. 131.
- Hill JL, Peregrine EK, Sprau GL, Cremeens CR, Nelson RL, Orf and Thomas DA (2008) Evaluation of the USDA Soybean Germplasm Collection: Maturity Groups 000–IV (PI 578371–PI 612761). U.S. Department of Agriculture Technical Bulletin no. 1919, p. 151.
- Hymowitz T and Harlan JR (1983) The introduction of the soybean to North America by Samuel Bowen in 1765. *Econ. Bot* 37: 371–379.
- Hyten David L, Qijian Song, Youlin Zhu, Ik-Young Choi, Nelson Randall L, Costa Jose M, Specht James E, Shoemaker Randy C and Cregan Perry B (2006) Impacts of genetic bottlenecks on soybean genome diversity. *PNAS* 103: 16666–16671.
- Iqbal MJ, Meksem K, Njiti VN, Kassem MA and Lightfoot DA (2001) Microsatellite markers identify three additional quantitative trait loci for resistance to soybean sudden-death syndrome (SDS) in Essex x Forrest RILs. *Theoretical and Applied Genetics* 102: 187–192.
- Keim Paul, Diers Brian W, Olson Terry C and Shoemaker Randy C (1990) RFLP mapping in soybean: association between marker loci and variation in quantitative traits. *Genetics* 126: 735–742.

- Kiang YT, Chiang YC and Kaizuma N (1992) Genetic diversity in natural populations of wild soybean in Iwate Prefecture, Japan. *Journal of Heredity* 83: 325–329.
- Lee JD, Yu JK, Hwang YH, Blake S, So YS, Lee GJ, Nguyen HT and Shannon JG (2008) Genetic diversity of wild soybean (*Glycine soja* Sieb. and Zucc.) accessions from South Korea and other countries. *Crop Science* 48: 606–616.
- Li Z and Nelson RL (2002) RAPD marker diversity among soybean and wild soybean accessions from four Chinese provinces. *Crop Science* 42: 1737–1744.
- Li YH, Li W, Zhang C, Yang L, Chang RZ, Gaut BS and Qiu LJ (2010) Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for SSR and SNP loci. *New Phytologist* 188: 242–253.
- Meksem K, Doubler K, Chanchaoenchai K, Njiti VN, Chang SJ, Rao Arelli AP, Cregan PB, Gray LE, Gibson PT and Lightfoot DA (1999) Clustering among loci underlying soybean resistance to *Fusarium solani*, SDS and SCN in near-isogenic lines. *Theoretical and Applied Genetics* 99: 1131–1142.
- Nagai I and Saito S (1923) Linked factors in soybeans. *Japan Journal of Botany* 1: 121–136.
- Nelson RL, Amdor PJ, Orf JH and Cavins JF (1988) Evaluation of the USDA Soybean Germplasm Collection: Maturity Groups 000 to IV (PI 427.136 to PI 445.845). USDA Technical Bulletin no. 1726.
- Nelson RL, Amdor PJ, Orf JH, Lambert JW, Cavins JF, Kleiman R, Laviolette FA and Athow KA (1987) Evaluation of the USDA Soybean Germplasm Collection: Maturity Groups 000 to IV (PI 273.483 to PI 427.107). USDA Technical Bulletin no. 1718. p. 267.
- Njiti VN, Johnson JE, Torto TA, Gray LE and Lightfoot DA (2001) Inoculum rates influences selection for field resistance to soybean sudden death syndrome in the greenhouse. *Crop Science* 41: 1–6.
- Njiti VN, Meksem K, Iqbal MJ, Johnson JE, Kassem MA, Zobrist KF, Kilo VY and Lightfoot DA (2002) Common loci underlie field resistance to soybean sudden death syndrome in Forrest, Pyramid, Essex, and Douglas. *Theoretical and Applied Genetics* 104: 294–300.
- Ohara M and Shimamoto Y (1994) Some ecological and demographic characteristics of two growth forms of wild soybean (*Glycine soja*). *Canadian Journal of Botany* 72: 486–492.
- Ohara M, Shimamoto Y and Sanbuichi T (1989) Distribution and ecological features of wild soybeans (*Glycine soja*) in Hokkaido. *Journal of the faculty of Agriculture – Hokkaido University* 64: 43–50.
- Oliveira Marcelo F, Randall L Nelson, Isaias O Geraldi, Cosme D Cruz and José Francisco F de Toledo (2010) Establishing a soybean germplasm core collection. *Field Crops Res* 119: 277–289.
- Palmer RG, Pfeiffer TW, Buss GR and Kilen TC (2004) Soybeans: improvement, production, and uses. In: Boerma HR and Specht JE (eds), vol. Agronomy. Madison, WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, pp. 137–233.
- Peregrine EK, Sprau GL, Cremeens CR, Handly P, Kilen TC, Smith JR, Thomas DA, Sarins JD and Nelson RL (2008) Evaluation of the USDA Soybean Germplasm Collection: Maturity Group V (FC 30265–PI 612614). and Maturity Groups VI–VIII (PI 416758–PI 606432B). U.S. Department of Agriculture Technical Bulletin no. 1920, p. 369.
- Piper CV and Morse WJ (1910) The soybean: History, varieties, and field studies. *USDA Bureau of Plant Industry Bulletin* 197.
- Piper CV and Morse WJ (1923) *The Soybean*. New York: McGraw Hill.
- Piper CV and Morse WJ (1925) How we got our soybeans. *Proceedings of American Soybean Association* 1: 58 Publ. 1928.
- Qiu Li-Juan, Peng-Yin Chen, Zhang-Xiong Liu, Ying-Hui Li, Rong-Xia Guan, Li-Hui Wang and Ru-Zhen Chang (2011) The worldwide utilization of the Chinese soybean germplasm collection. *Plant Genetic Resources* 9: 109–122.
- Scherm H and Yang XB (1996) Development of sudden death syndrome of soybean in relation to soil temperature and soil water potential. *Phytopathology* 86: 642–649.
- Schmutz J, *et al.* (2010) Genome sequence of the paleopolyploid soybean (*Glycine max* (L.) Merr.). *Nature* 463: 178–183.
- Sprague GF (1971) Genetic vulnerability in corn and sorghum. *Proceedings of Annual Corn and Sorghum Research Conference* 26: 96–104.
- U.S. Department of Agriculture, Division of Botany (1898) Inventory no. 1. Foreign seeds and plants imported by the Section of Seed and Plant Introduction, nos. 1–1000 (Available at <http://ddr.nal.usda.gov/handle/10113/37063>).
- Wen ZX, Ding YL, Zhao TJ and Gai JY (2009) Genetic diversity and peculiarity of annual wild soybean (*G. soja* Sieb. et Zucc.) from various eco-regions in China. *Theoretical and Applied Genetics* 119: 371–381.
- Zhu W, Zhou T, Zhong M and Lu B (2007) Sampling strategy for wild soybean (*Glycine soja*) populations based on their genetic diversity and fine-scale spatial genetic structure. *Frontiers of Biology in China* 2: 397–402.