5-2003

Analysis of germplasm distribution patterns for collections held at the North Central Regional Plant Introduction Station, Ames, Iowa, USA

Mark P. Widrlechner
United States Department of Agriculture, isumw@iastate.edu

Lisa A. Burke
United States Department of Agriculture

Analysis of germplasm distribution patterns for collections held at the North Central Regional Plant Introduction Station, Ames, Iowa, USA

Abstract
Understanding the patterns of distribution of plant genetic resources, especially the extent and contextual bases of distributions, may be critical in setting appropriate targets for seed multiplication, packaging, storage space, and other technical operations. We analyzed germplasm distribution patterns over a 12-year period for 10 crop collections conserved by the North Central Regional Plant Introduction Station in Ames, Iowa, to determine if distribution rates over a given time interval help predict future distributions and to document how distribution patterns vary among accessions within collections. We demonstrated that, with an appropriate tracking system and commonly available statistical software, germplasm distribution patterns can be easily analyzed and plotted over time. Data measured over periods of up to 3 years had little predictive value, while a 6-year period gave relatively accurate projections of future distributions. Patterns of distributions within collections varied between those that are approximately normally distributed and those best described by an exponential function, with larger collections tending to be non-normally distributed. Means and standard deviations of standardized, long-term distribution rates, calculated from samples of 200–700 accessions, accurately described the distributional rates of 90–95% of all accessions. The documentation of changes in usage patterns within and among collections as they mature is also discussed. Analysis of average shipment size suggests that germplasm distributions became more focused over time for 8 of the 10 collections analyzed. This may result when users request germplasm based upon knowledge about specific accessions gained through personal experience and by examining evaluation and characterization data.

Disciplines
Agricultural Science | Agriculture | Plant Sciences

Analysis of germplasm distribution patterns for collections held at the North Central Regional Plant Introduction Station, Ames, Iowa, USA

Mark P. Widrlechner[*] and Lisa A. Burke
USDA-Agricultural Research Service, North Central Regional Plant Introduction Station, Department of Agronomy, Iowa State University, Ames, Iowa 50011-1170, USA; [*]Author for correspondence (e-mail: isumw@iastate.edu)

Abstract

Understanding the patterns of distribution of plant genetic resources, especially the extent and contextual bases of distributions, may be critical in setting appropriate targets for seed multiplication, packaging, storage space, and other technical operations. We analyzed germplasm distribution patterns over a 12-year period for 10 crop collections conserved by the North Central Regional Plant Introduction Station in Ames, Iowa, to determine if distribution rates over a given time interval help predict future distributions and to document how distribution patterns vary among accessions within collections. We demonstrated that, with an appropriate tracking system and commonly available statistical software, germplasm distribution patterns can be easily analyzed and plotted over time. Data measured over periods of up to 3 years had little predictive value, while a 6-year period gave relatively accurate projections of future distributions. Patterns of distributions within collections varied between those that are approximately normally distributed and those best described by an exponential function, with larger collections tending to be non-normally distributed. Means and standard deviations of standardized, long-term distribution rates, calculated from samples of 200–700 accessions, accurately described the distributional rates of 90–95% of all accessions. The documentation of changes in usage patterns within and among collections as they mature is also discussed. Analysis of average shipment size suggests that germplasm distributions became more focused over time for 8 of the 10 collections analyzed. This may result when users request germplasm based upon knowledge about specific accessions gained through personal experience and by examining evaluation and characterization data.

Introduction

The primary mission of most active plant genebanks is to conserve and distribute germplasm in support of scientific research, new product development, and crop improvement. The value of the collections held by genebanks is closely connected to their past, present, and future uses. To enhance both realized and potential value, considerable resources have been, and continue to be, devoted to germplasm evaluation and characterization. The utility of data generated through such efforts and of tools for the electronic communication of these data is obvious, and great progress is being made by many genebanks on these fronts. Less obvious, perhaps, is the concept that the collection and analysis of distributional data may also contribute to more efficient and effective use and management of germplasm.

In an environment where many genebanks face serious resource limitations and backlogs in regeneration (FAO (Food and Agriculture Organization of the United Nations) 1998), investments in capturing and analyzing distributional data could prove particularly valuable when applied correctly by curators and managers to help allocate resources for regenerating, storing, and distributing germplasm (Widrlechner 1997). For example, knowledge of distribution patterns, especially of the extent and intended uses of

distributions, may be critical in determining appropriate targets for seed multiplication, prepackaging, and other operations. Sadly, however, there are very few publications that analyze patterns of germplasm distribution or make pragmatic recommendations on how to apply the analysis of distribution to resolve practical germplasm management issues.

The North Central Regional Plant Introduction Station (NCRPIS) in Ames, Iowa, is a large, active genebank, with more than 46,000 accessions, holding a large majority of its collections as seeds in medium-term storage. The present study is a retrospective analysis of patterns of germplasm distribution from the NCRPIS between 1988 and 1999 for 10 of its most distributed crop collections (Table 1). The initial goals of this study were: 1) to determine if distribution levels for a given period of time could help predict future distributions; and 2) to document how patterns of distribution vary among accessions within collections. Initial results were then used to develop recommendations on how such analyses can help plan future resource allocation and select appropriate targets for seed multiplication.

Materials and methods

Distributional data for the period 1988–1999 for 10 major seed collections (Table 1) conserved by the NCRPIS were obtained from the Germplasm Resources Information Network (GRIN) database (Bird 1994). Distributional data included all shipments fulfilling external requests and internal research projects, but excluded distributions made for routine, internal management purposes, such as germination testing, regeneration, and security back-up. Distribu-

tions were counted as the total number of seed packets, not as the total number of accessions, distributed, as we wished to obtain data on the overall frequencies of distribution of the various collections. Each year's distributional data were standardized for the number of accessions available for distribution in that year to obtain a standardized measure of distribution rate, D, distributions $\times$ (available accessions)$^{-1}$ $\times$ year^{-1} (Widrlechner 1995). A standardized measure of shipment size, S, distributions $\times$ shipments^{-1}, was also calculated for each year in the study period. Standardized data were organized and plotted with Excel and statistically analyzed with linear regression by using MSTAT-C (MSTAT Development Team 1994).

To determine whether past distribution rates could be a useful predictor of future distribution, values of D for 1, 2, and 3-year periods were correlated with values of D for the subsequent 1, 2, and 3-year periods, and 6-year means and standard deviations for D for the period 1988-1993 were compared with values of D for the period 1994-1999.

In addition to these standard statistical evaluations applied to entire collections, distributional data for between 196 and 696 available accessions for each of the 10 major collections (Table 1) were extracted from the GRIN database to examine variation in the extent of distribution within collections. Data on available accessions for each collection were obtained from the Inventory Area of the GRIN database during November and December 2000. Selections were made by beginning at a random starting point and choosing those spaced at a regular interval thereafter, the spacing determined by the size of the collection and the number of accessions desired for analysis. The number of accessions examined for distribution was based on the size of each collection (Table 1). For the seven smallest crop collections, ca. 200 accessions were selected from each crop, giving sampling proportions of 13 to 65% of the available accessions. For two larger collections, *Amaranthus* L. and *Cucumis melo* L. 12% of the available accessions were examined. For *Zea mays* L. the largest of the 10 collections, a 6% sample was chosen by selecting two independent subsets, each including 3% of the collection. These two subsamples were not statistically different for values for D from the entire collection and were combined for analysis.

Collections often contain subsets, which because of prior research or their specific histories, can be expected to have much higher distribution rates than

Table 1. Characteristics of NCRPIS Collections Studied.

Collection	Available Accessions		Accessions Sampled
	1988	1999	
Amaranthus L. spp.	721	2712	331
Brassica L. spp. (oilseeds)	1139	1559	196
Cucumis melo L.	1300	2095	264
Cucumis sativus L.	707	1194	200
Cucurbita pepo L.	511	761	200
Daucus L. spp.	387	567	200
Helianthus annuus L. (cultivated)	720	1313	200
Helianthus L. spp. (wild)	199	959	200
Spinacia L. spp.	254	317	206
Zea mays L.	5003	10582	696

would be typical. This phenomenon was noted by Wilson and Mihm (1997) for *Zea mays* accessions known to be resistant to insect pests. To examine this aspect in more detail, in addition to the 6% *Zea* L. sample described earlier, we chose to examine distributional histories for a set of 32 popcorn lines discussed by Wilson and Mihm (1997). These lines were developed by J.C. Eldredge and included known sources of resistance to corn earworm, *Helicoverpa zea* (Boddie), European corn borer, *Ostrinia nubilalis* (Hübner), and fall armyworm, *Spodoptera frugiperda* (J.E. Smith) (Wilson et al. 1991).

The distributional histories of the selected accessions were examined, and all internal distributions related to the management of the accessions, i.e., regeneration, viability testing, and back-up, were omitted from further analysis. The remaining data represent germplasm distributions made to meet evaluation, crop improvement, direct introduction, and basic research needs. These distributional data were standardized for each accession by dividing the number of distributions by the number of years for which that accession was available during the period 1988–2000. For graphical analysis, values of **D** were categorized based on intervals of 0.5; e.g., all values of **D** falling between 0 and 0.5 were grouped together for analysis and were plotted in figures at a mean value of 0.25.

The standardized data were plotted with Excel and initially compared with normal distributions and other descriptive mathematical functions with Table Curve™ 2D (SPSS 1997) and Sigma Plot® 2000 (SPSS 2000). Data sets with significant deviations from normality were manually compared with truncated normal distributions lacking negative values and to Poisson distributions with Chi-square tests.

Results and discussion

Aggregate distributional analysis

Linear regression analyses indicated that, for 8 of 10 collections, **D** declined over the period 1988–1999 (Table 2). Four of the declines and one increasing trend (for *Spinacia* L.) were significant at the 5% level. Unfortunately, these trends were not strong enough to serve as an effective management tool except to give a general overview, because most of the R^2 values were low (range: 0.022–0.624), explaining only a small proportion of the annual variation in

Table 2. Linear Regressions of the Standardized Distribution Rate, **D**, over Years for the Period 1988–1999.

Collection	Slope	R^2
Amaranthus spp.	0.049	0.199
Brassica spp. (oilseeds)	−0.185	0.461*
Cucumis melo	−0.130	0.194
Cucumis sativus	−0.166	0.294
Cucurbita pepo	−0.065	0.052
Daucus spp.	−0.031	0.022
Helianthus annuus (cultivated)	−0.256	0.531**
Helianthus spp. (wild)	−0.409	0.624**
Spinacia spp.	0.388	0.419*
Zea mays	−0.043	0.588**

*, ** - probability levels of 0.05 and 0.01, respectively.

D even when statistically significant. Data plots, however, did indicate one obvious change in distribution patterns over time. In the *Spinacia* collection (Table 3) the discovery of new, virulent races of downy mildew in the 1990s (Correll et al. 1994, 2000) coincident with the use of controlled pollination to regenerate the collection beginning in 1994, presumably led to a large increase in requests for *Spinacia* germplasm beginning late in 1995.

Weak declines in **D** over time may be expected in germplasm collections under at least four circumstances. First, **D** may decline, even while the overall number of distributions is increasing, whenever the number of available accessions is increasing at an even faster rate. We observed this in the *Daucus* L. and *Zea mays* collections, where the slope of non-standardized (total annual) distributions was positive (Table 4). Second, **D** may decline in periods after the completion of large-scale evaluation projects. We observed this situation in the *Cucumis sativus* L. and cultivated *Helianthus* L. collections, for which National Crop Germplasm Committees had been spon-

Table 3. Annual Standardized Distribution Rate, **D**, for the *Spinacia* Collection for the Period 1988–1999.

Year	D
1988	0.024
1989	0
1990	3.973
1991	0.619
1992	0
1993	0
1994	0
1995	1.429
1996	4.650
1997	4.963
1998	4.794
1999	3.347

soring extensive evaluations in the late 1980s and early 1990s. Third, as valuable germplasm from collections is incorporated into users' research and crop improvement programs and subsequently is made more widely available beyond the genebank, and when collections are relatively static, **D** may decline as collections are ''mined'' of their most useful genes. Fourth, as increasing amounts of evaluation data are made available to users through databases and publications, it becomes easier for users to target their germplasm requests to specific accessions, which should, in turn, reduce **D** when measured over entire collections.

An assessment of changes in targeting was made by analyzing trends in **S**, the average shipment size, over time (Table 5). All eight of the collections that experienced declines in **D** also showed declines in **S**. Three of these declines, for *Brassica* L., *Zea mays*, and cultivated *Helianthus*, were significant at the 5% level. For two of these three collections, declines occurred during periods when there were very large increases in the quantity of evaluation data available to users in the GRIN database, and while the GRIN database itself became more accessible to users via the Internet (Bird 1994). For example, the number of evaluation observations in GRIN for cultivated *Helianthus* increased more than threefold, from about 16,000 datapoints in 1988 to more than 70,000 in 2000, and the number of *Zea mays* observations in GRIN increased from about 128,000 datapoints in 1989 to more than 188,000 in 2000. Although the number of datapoints for *Brassica* was relatively stable during this period, there was a notable shift in germplasm requests toward a focus on phytoremediation, involving accessions with enhanced heavy-metal accumulation, following the publication of Nanda Kumar et al. (1995) pioneering study. Increases in

accessible evaluation data on GRIN and in scientific publications both likely contributed to users' ability to focus their germplasm requests, resulting in a declining number of samples per individual shipment over time.

Conversely, the two collections for which **D** increased over time, *Amaranthus* and *Spinacia* (Table 2), experienced concomitant, but non-significant increases in **S** (Table 5). For *Spinacia*, this resulted primarily from large-scale, disease evaluations noted earlier. For *Amaranthus*, it may have resulted from the very large change in the number of available accessions that occurred between 1988, when there were only 721 accessions available, and 1999, when the number had risen to 2712, and from concerted efforts by the curator to expand the available range of taxonomic and phenotypic diversity. Of the 10 collections studied, only one other, the wild *Helianthus*, experienced such a rapid increase in the scope of the available accessions.

Values of **D** calculated for 1, 2, or 3-year periods generally have insufficient predictive power for projecting distribution rates into the subsequent 1, 2, or 3-year periods. The mean value of R^2 among the 10 collections for the 1-year period was only 0.146; for the 2-year period, it was 0.319; and for the 3-year period, it was 0.294. For three of the collections in the 1-year analyses, four in the 2-year analyses, and three in the 3-year analyses, the slopes of the regressions were actually negative, suggesting that distributions may follow a sawtooth pattern from year to year (periods of low distribution followed by high distribution and vice versa), but only 2 (*Daucus* and *Amaranthus*) of the 10 negatively-sloped regressions were statistically significant at the 5% level.

To make these models practically useful to managers, we propose that they should explain at least half

Table 4. Linear Regressions of the Non-standardized (Total) Number of Annual Distributions over Years for the Period 1988–1999.

Collection	Slope	R^2
Amaranthus spp.	208	0.487**
Brassica spp. (oilseeds)	−185	0.348*
Cucumis melo	−130	0.117
Cucumis sativus	−72	0.091
Cucurbita pepo	−34	0.040
Daucus spp.	5.5	0.004
Helianthus annuus (cultivated)	−93	0.194
Helianthus spp. (wild)	−30	0.122
Spinacia spp.	107	0.461*
Zea mays	33	0.022

*, ** - probability levels of 0.05 and 0.01, respectively.

Table 5. Linear Regressions of Average Shipment Size, **S**, over Years for the Period 1988–1999.

Collection	Slope	R^2
Amaranthus spp.	2.368	0.315
Brassica spp. (oilseeds)	−8.670	0.379*
Cucumis melo	−4.988	0.141
Cucumis sativus	−1.171	0.012
Cucurbita pepo	−1.893	0.059
Daucus spp.	−1.877	0.083
Helianthus annuus (cultivated)	−2.840	0.420*
Helianthus spp. (wild)	−0.968	0.170
Spinacia spp.	7.594	0.144
Zea mays	−2.534	0.666***

*, *** - probability levels of 0.05 and 0.001, respectively.

of the total variation in **D**. But none of the 1-year regressions and only two, 2-year regressions and three, 3-year regressions, had R^2 values >0.5. It is notable that both the mean R^2 values and the number of R^2 values above 0.5 were higher for the 2 and 3-year models than for the 1-year model, suggesting that applications of historical, distributional data as a management tool should always be based on analyses of more than a single year to reduce short-term fluctuations.

The mean values for **D**, based on the period 1988–1993, ranged between 0.609 and 3.41 (Table 6), which are somewhat higher than those reported for various NPGS collections by Widrlechner (1995). Of 10 standard deviations, 8 were smaller than the mean values during 1988–1993 (Table 6), but two collections, *Cucurbita pepo* and *Spinacia*, experienced extremely high among-year variation in **D** during this period. There was little activity for *Spinacia* except during 1990 (see Table 3). And for *Cucurbita pepo*, 1991 was an outlier from the other years in the study period. In that year, there were two *Cucurbita* L. germplasm requests resulting in very large distributions, one to evaluate our collection for a physiological disorder, squash silverleaf (Cohen et al. 1991; Zitter et al. 1996), and the other to conduct a comprehensive taxonomic verification.

We then compared annual values of **D** for the period 1994–1999 in relation to the mean values and standard deviations (SD) for **D** based on the period 1988–1993. During the period 1994–1999, 54 of 60 (90%) values of **D** were ≤**D** (1988–1993) + 1 SD and 56 of 60 (93%) were ≤**D** (1988–1993) + 2 SD. These data compare fairly well to statistical expectations for normal distributions generally, wherein 84% of values should be ≤1 SD above the mean and 98%

should be ≤2 SD above the mean (Gibbons 1976). Most of the exceptionally high values were for *Spinacia*, which, as described earlier, experienced a significant change in the collection during the period 1994–1999. For collections that are relatively stable, we would suggest that one could use a 6-year period to estimate a mean and standard deviation for **D** and use **D** + 1 SD as an upper estimate of distributions to help set general targets for seed regeneration, especially for newly acquired accessions and those where the manager has no *a priori* information about potential demand. For collections that are experiencing large increases in distributions, it may be necessary to use a higher estimate, such as **D** + 2 SD.

These estimates can then be applied to empirical models for determining appropriate regeneration size, such as the model recommended by Sackville-Hamilton and Chorlton (1997) or could be used to help determine optimal numbers of samples to prepackage to meet future requests. In addition, for genebanks that conserve a number of different collections, these data, in conjunction with specific information about the collections themselves, may serve as a rough gauge of relative user activity and interest, which can be factored into plans for future resource allocation and regeneration efforts. But it is important to remember that these analyses were aggregated among all the available accessions within a crop. Additional insights can be gained by examining within-collection variation for distribution rates to clarify usage patterns and identify accessions of special interest.

Distributional analysis among accessions within collections

Of the 10 crop collections studied, frequency distributions for 7 of them, oilseed *Brassica*, *Cucumis melo*, *Cucumis sativus*, *Cucurbita*, *Daucus*, *Spinacia*, and cultivated *Helianthus*, passed tests for normality at the 5% level in the Sigma Plot® 2000 program (SPSS 2000). A typical example is shown in Figure 1 for accessions of *C. sativus*.

The other three collections, *Amaranthus*, wild *Helianthus*, and *Zea* L., exhibited distribution patterns significantly deviating from normality. The *Amaranthus* and *Zea* data sets displayed obviously asymmetric, non-normal frequency distributions (Figure 2,3). These two collections also had low, overall mean values for **D**, **D** < 0.7. Since values for **D** cannot be negative, data sets that otherwise would be normally

Table 6. Six-year Means and Standard Deviations for the Standardized Distribution Rate, **D**, for the Period 1988–1993.

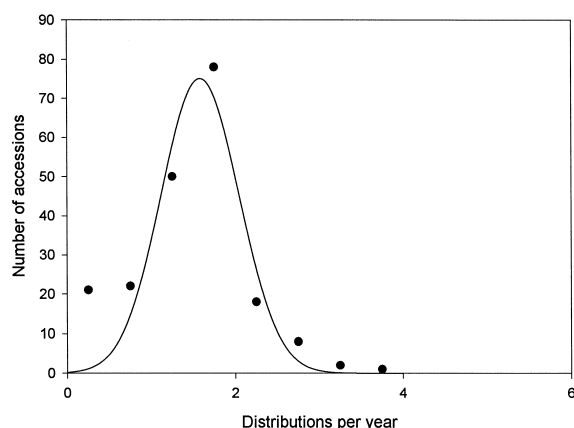| Collection	Mean **D**	SD **D**
Amaranthus spp.	0.609	0.343
Brassica spp. (oilseeds)	2.448	1.098
Cucumis melo	1.312	1.466
Cucumis sativus	2.375	1.058
Cucurbita pepo	0.883	1.431
Daucus spp.	1.273	0.981
Helianthus annuus (cultivated)	3.288	1.422
Helianthus spp. (wild)	3.410	1.893
Spinacia spp.	0.769	1.589
Zea mays	0.695	0.189

334

Figure 1. Frequency of Distributions $\times$ Year^{-1} for 200 Accessions of *Cucumis sativus*, with the Best-fit, Normal Distribution (solid line); r^2 = 0.864.

Figure 2. Frequency of Distributions $\times$ Year^{-1} for 331 Accessions of *Amaranthus*, with the Corresponding Exponential Function, y = 474$^{-x/0.363}$ (solid line); r^2 = 0.998.

distributed might then present distributions with truncated left tails and an excess of values in the lowest positive data class. Thus, we subjected the *Amaranthus* and *Zea* distributions to Chi-square analyses to determine whether they conformed to truncated normal distributions. However, these distributions differed significantly from truncated normal distributions at the 1% probability level. In addition, all three non-normal distributions were tested for their degree of conformity to the Poisson distribution, a typical pattern describing relatively rare, discrete events in large populations (Sokal and Rohlf 1995), but again they were significantly different at the 1% level.

The overall patterns of *Amaranthus* and *Zea* data sets (Figure 2,3) can be described by the exponential function, y = a$^{-x/b}$, with best-case R^2 values >0.997, based on analyses with Table Curve™ 2D (SPSS 1997). Generally, when a is large and b < 1, as it is for these data sets, this function describes situations where most accessions are rarely distributed, and the frequency of more commonly distributed accessions decreases at an exponential rate. It cannot, however, be extrapolated to predict accurately the number of accessions that are not distributed at all, but rather, it greatly overestimates them.

The "atypical" patterns shown by *Amaranthus* and *Zea* more closely conform to commonly held views that germplasm collections are underutilized (Morales et al. 1995; Spagnoletti Zeuli and Qualset 1995), with many accessions distributed rarely, if at all (Goodman 1990; Diwan et al. 1994), than do patterns of dis-
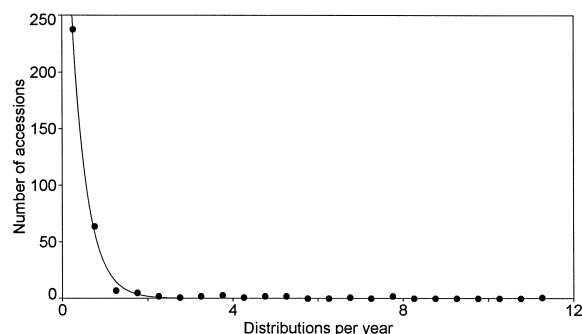
tribution for the other collections analyzed. It is notable that these two collections are also the two largest we studied. Extremely large collections can be more difficult for users to access, which is one of the primary rationales for the creation of core collections (Frankel 1984; Brown and Spillane 1999). Thus, they might be expected to display different patterns of usage than for smaller, more accessible collections where larger blocks or entire collections are often requested and studied at once.

The only collection that deviated from both a normal distribution and an exponential function was the wild *Helianthus* collection (Figure 4). The pattern for wild *Helianthus* accessions seems to fit somewhere between the normal and the exponential distributions, and can be approximated by a log-normal distribution, with a best-case R^2 value of approximately 0.94. An interesting subject for further in-
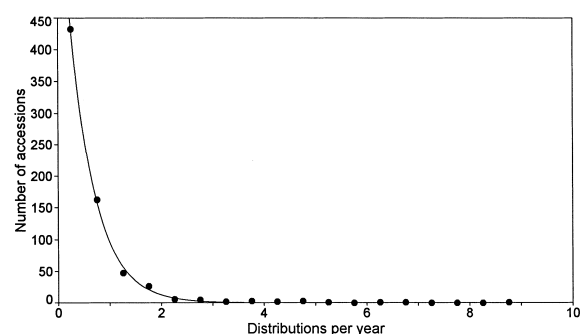

Figure 3. Frequency of Distributions $\times$ Year^{-1} for 696 Accessions of *Zea mays*, with the Corresponding Exponential Function, y = 718$^{-x/0.495}$ (solid line); r^2 = 0.999.

vestigation would be to determine whether this collection has been evolving from a normally distributed pattern to an exponential pattern as the collection expanded and more was learned about it.

Ideally, germplasm regeneration targets should be set high enough to produce enough seeds to meet requests for the vast majority of accessions, but low enough so that large quantities of seeds are not wasted. From graphical presentations of distribution rates among accessions (such as Figure 1–4) the right tails of the patterns of distributions are of particular interest, because they include those accessions most frequently distributed and thus most likely to be quickly depleted.

The proportions of accessions found in the right tails of the patterns of distribution for all 10 collections that are ≥ 1 and 2 SD beyond the mean values of **D** for the aggregate samples were thus examined. We found that between 78 and 94% of all sampled accessions have individual distribution rates (values of **D**) ≤ 1 SD above the overall mean. And between 95 and 100% of all sampled accessions have individual distribution rates (values of **D**) ≤ 2 SD above the overall mean. There is no significant difference between the proportions of accessions found in the right tails for the seven collections that generally conform to normal distributions and those for the three that do not. These data also compare well to statistical expectations for normal distributions generally (Gibbons 1976).

We would suggest that within-collection **D** values be calculated for samples of at least 200 accessions and that a measure between **D** + 1 SD and **D** + 2 SD be used to derive target values of **D** that should capture between 90 and 95% of all accessions. These

recommendations resemble those presented herein based upon annual variation in **D**, but may actually be more directly applicable to most management tasks because they are based upon within-collection variation and patterns among individual accessions, rather than variation in aggregated values over time.

These target values are based on patterns of variation in distributional histories for large numbers of accessions and should be particularly useful together with information about the longevity of seed samples under specific storage conditions to set targets for the optimal size of regeneration samples (Sackville-Hamilton and Chorlton 1997). For those genebanks that distribute set numbers of seeds to meet requests, pre-packaging of samples can increase the efficiency of inventory and order fulfillment. Target values can be used to estimate future distributions over a set time period, which can then be used as a guideline for the number of distribution units to be pre-packaged. Similarly, estimates of the quantities of seeds required to meet future distribution requests can be used in concert with seed size, shape, and viability characteristics to help determine the most appropriate types and sizes of containers and other packaging materials. Managers may also wish to store those accessions that are most frequently distributed in the most accessible locations.

As noted by Widrlechner (1995) Wilson and Mihm (1997), groups of accessions, with special characteristics that make them more frequently distributed than is typical, may occur in collections. To demonstrate how striking these differences can be, we chose to examine distribution patterns for a set of 32 popcorn lines (Table 7) which were reported by Wilson et al. (1991) to include sources of insect resistance. Although the overall **D** value for the maize collection, as measured from both aggregated annual data and from a sample of nearly 700 accessions, was <0.7, the mean **D** value for these 32 lines was 3.6, more than 3.5 standard deviations above the overall collection mean. During the course of this study, we noted other groups of accessions with high distribution rates that were sometimes even more striking than those found for the 32 popcorn accessions, including a group of 44 inbred lines of maize that had been selected as check varieties for Plant Variety Protection with a mean **D** value of 8.2. Methods based on capturing information that describes 90 to 95% of all accessions do not apply well to such exceptional accessions. Thus it is extremely important that curators pay close attention to groups of special accessions with atypically high
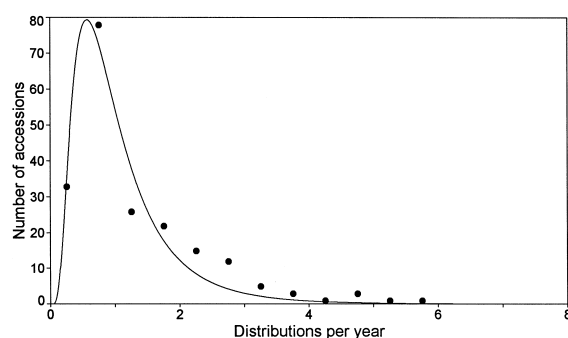

Figure 4. Frequency of Distributions $\times$ Year^{-1} for 200 Accessions of wild *Helianthus*, with the Best-fit, Log-normal Distribution (solid line); $r^2 = 0.940$.

Table 7. Distributions $\times$ Year^{-1} for 32 Popcorn Accessions for the Period 1988–1999[1].

Distributions $\times$ Year^{-1}	Number of Accessions
0–2	0
2–2.5	1
2.5–3	8
3–3.5	9
3.5–4	10
4–4.5	0
4.5–5	2
5–5.5	0
5.5–6	1
6–6.5	1

[1]In contrast, the standardized distribution rate, **D**, for the entire *Zea mays* collection from aggregated annual data was 0.55, and the mean value for the sample of 696 *Zea* accessions was 0.67 (see also Figure 3).

demand, likely accompanied by great value, by closely monitoring requests and through frequent communication with users.

Conclusions

In this paper, we have demonstrated that, with an appropriate tracking system (as found in the GRIN database) and with commonly available statistical software, germplasm distribution patterns can be easily analyzed on both an aggregate and a within-collection basis. We would suggest that these data be used together with other management tools, such as information about unfulfilled requests, new threats to crop production, and broader surveys of germplasm users (McFerson et al. 1996; Widrlechner 1997), to help plan the size of seed regenerations, the extent of or need for pre-packaging, and resource allocation for storage, as part of larger strategies to improve efficiency and customer service.

We found that the utility of distributional data increases over time. Aggregated data had insufficient predictive value as a management tool, even when examined over periods of up to 3 years. However, after a period of 6 years, the means and standard deviations of **D** based on aggregated data may help project future distributions, with relatively accurate projections resulting from a single test of two 6-year periods based on values between **D** + 1 SD and **D** + 2 SD.

Analyses of distributional data within collections should be even more useful than those based on aggregated data. Contrary to commonly held views, analyses of distributions within collections suggest that, for most of our collections, many accessions are distributed relatively frequently and few languish unrequested. We realize, however, that many genebanks may not yet have tools in place to easily track distributional histories for individual accessions. Thus, while aggregated measures of distributions over time may not be ideal, they may be easier for genebanks to determine than are analyses of within-collection variation. Our analyses of long-term, distributional data within collections show that patterns of distribution vary between those that are more or less normally distributed and those that are described by the exponential function, $y = a^{-x/b}$, with the larger collections sampled tending to be non-normally distributed. For both normally and exponentially distributed collections, means and standard deviations of **D** calculated from samples of between 200 and 700 accessions can be used to describe the distributional histories of 90 to 95% of all accessions, when based on values between **D** + 1 SD and **D** + 2 SD. These estimates should be particularly useful in setting targets for seed quantities for regeneration, for pre-packaging, and other storage-management issues.

Long-term analyses of distributional data can also help document how usage patterns within and among collections evolve as collections are developed and mature over time. Analysis of the statistic, **S**, average shipment size, suggested that germplasm distributions became more focused over time for 8 of the 10 collections analyzed. This can be expected as users request germplasm based upon knowledge gained about specific accessions through personal experience and by examining the collective results of past evaluation and characterization work.

We hope that this study motivates other genebanks to develop and use appropriate information technologies that will allow their staffs to analyze patterns of germplasm distribution over time. We suspect that the utility of such analyses in germplasm management is much greater than the suggested uses that we have proposed herein, and hope that our report inspires greater attention and ingenuity directed to this neglected topic.

Acknowledgements

References

Bird E.M. 1994. The USDA germplasm information management and access. Quarterly Bulletin of the International Association of Agricultural Information Specialists 39: 111–115.

Brown A.H.D. and Spillane C. 1999. Implementing core collections - principles, procedures, progress, problems, and promise. In: Johnson R.C. and Hodgkin T. (eds), Core Collections for Today and Tomorrow. International Plant Genetic Resources Institute, Rome, Italy, pp. 1–9.

Cohen S., Duffus J.E., Liu H.Y. and Perry R. 1991. Induction of silverleaf by *Bemisia* whitefly from California desert whitefly populations. Plant Disease 75: 862.

Correll J.C., Irish B.M., Koike S.T. and Morelock T.E. 2000. Update on downy mildew (*Peronospora farinosa* f. sp. *spinaciae*) of spinach In: Summaries and Program: International Leafy Vegetable Conference, 25–28 September 2000, Seaside, California., pp. 19–20.

Correll J.C., Morelock T.E., Black M.C., Brandenberger L.P. and Dainello F.J. 1994. Economically important diseases of spinach. Plant Disease 78: 653–660.

Diwan N., Bauchan G.R. and McIntosh M.S. 1994. A core collection for the United States annual *Medicago* germplasm collection. Crop Science 34: 279–285.

FAO (Food and Agriculture Organization of the United Nations) 1998. The State of the World's Plant Genetic Resources for Food and Agriculture. Food and Agriculture Organization of the United Nations, Rome, Italy.

Frankel O.H. 1984. Genetic perspectives of germplasm conservation. In: Arber W.K., Llimensee K., Peacock W.J. and Starlinger P. (eds), Genetic Manipulation: Impact on Man and Society. Cambridge University Press, Cambridge, UK, pp. 161–170.

Gibbons J.D. 1976. Nonparametric Methods for Quantitative Analysis. American Sciences Press, Columbus, Ohio.

Goodman M.M. 1990. Genetic and germ plasm stocks worth conserving. Journal of Heredity 81: 11–16.

McFerson J.R., Lamboy W.F. and Kresovich S. 1996. Assessing user perceptions of genetic resource collections in crucifer crops. Crop Science 36: 831–838.

Morales E.A.V., Valois A.C.C. and Costa I.R.S. 1995. Core collections for gene banks with limited resources. In: Hodgkin T., Brown A.H.D., van Hintum T.J.L. and Morales E.A.V. (eds), Core Collections of Plant Genetic Resources. John Wiley, Chichester, UK, pp. 241–249.

MSTAT Development Team 1994. MSTAT-C Version 2.1: A Microcomputer Program for the Design, Management and Analysis of Agronomic Research Experiments. Michigan State University, East Lansing.

Nanda Kumar P.B.A., Dushenikov V., Motto H. and Raskin I. 1995. Phytoextraction: The use of plants to remove heavy metals from soils. Environmental Science and Technology 29: 1232–1238.

Sackville-Hamilton N.R. and Chorlton K.H. 1997. Regeneration of Accessions in Seed Collections: A Decision Guide. Handbook for Genebanks No. 5. International Plant Genetic Resources Institute, Rome, Italy.

Sokal R.R. and Rohlf F.J. 1995. Biometry: The principles and practices of statistics in biological research. 3rd edn. Freeman, New York.

Spagnoletti Zeuli P.L. and Qualset C.O. 1995. The durum wheat core collection and the plant breeder. In: Hodgkin T., Brown A.H.D., van Hintum T.J.L. and Morales E.A.V. (eds), Core Collections of Plant Genetic Resources. John Wiley, Chichester, UK, pp. 213–228.

SPSS 1997. Table Curve™ 2D User's Manual. SPSS, Inc., Chicago, Illinois.

SPSS 2000. Sigma Plot® 2000 User's Guide. SPSS, Inc., Chicago, Illinois.

Widrlechner M.P. 1995. A new look at prairie plant germplasm. In: Hartnett D.C. (ed.), Proceedings of the Fourteenth North American Prairie Conference. Prairie Biodiversity, Kansas State University, Manhattan., pp. 207–210.

Widrlechner M.P. 1997. Managerial tools for seed regeneration. Plant Varieties and Seeds 10: 185–193.

Wilson R.L. and Mihm J.A. 1997. Maintenance of, and requests for, maize germplasm having resistance to insect pests. In: Insect Resistant Maize: Recent Advances and Utilization. Proceedings of an International Symposium held at the International Maize and Wheat Improvement Center, 27 November - 3 December 1994, CIMMYT, Mexico City, D.F., pp. 230–233.

Wilson R.L., Wiseman B.R. and Reed G.L. 1991. Evaluation of J. C. Eldredge popcorn collection for resistance to corn earworm, fall armyworm (Lepidoptera: Noctuidae), and European corn borer (Lepidoptera: Pyralidae). Journal of Economic Entomology 86: 952–956.

Zitter T.A., Hopkins D.L. and Thomas C.E. (eds) 1996. Compendium of Cucurbit Diseases. APS Press, Saint Paul, Minnesota.