Selection of a Core Collection from the U.S. Germplasm Collection of Peanut

C. Corley Holbrook,* William F. Anderson, and Roy N. Pittman

ABSTRACT

The U.S. germplasm collection for peanut, *Arachis hypogaea* L., consists of 7432 accessions and contains a great amount of genetic diversity. Information on economically important traits does not exist for most accessions due to the time and labor required for evaluation. The development of a core collection for peanut would provide a subset of accessions that are representative of the entire collection and that could be extensively examined. The objective of this research was to select a core collection for peanut. Data for the U.S. peanut germplasm collection were obtained from the Germplasm Resources Information Network (GRIN). The entire germplasm collection was then stratified by country of origin and by the amount of available morphological data. When information was available for at least four of the same morphological variables for at least 16 accessions from the same country of origin, then the data for these accessions were analyzed using multivariate statistical analysis. Results allowed the accessions to be clustered into groups which, theoretically, are genetically similar. Random sampling was then used to select ≈10% from each group. Accessions with inadequate data for multivariate analysis were selected using a 10% random sample from each country of origin. Accessions from countries having few (≤5) entries in the collection were pooled and a 10% random sample was selected. The resulting 831 accessions form a core collection for peanut. Examination of data for six phenotypic traits indicated that the genetic variation expressed for each trait in the entire collection has been preserved in this core collection. This peanut core collection has many potential uses and should increase the utilization of peanut germplasm resources.

THE U.S. GERMPLASM COLLECTION for peanut consists of 7432 accessions and contains a great amount of genetic diversity. However, this diversity has not been adequately evaluated (Banks, 1976) or extensively used in cultivar development (Knauft and Gorbet, 1989). Utilization of the collection could be enhanced by the development of more efficient evaluation techniques.

Frankel (1984) proposed that a germplasm collection could be reduced to what he termed a core collection. The core collection would be designed to minimize repetitiveness within the collection and should represent the genetic diversity of a crop species. The core collection could serve as a working collection that could be extensively examined, and the accessions excluded from the core collection would be retained as the reserve collection (Frankel, 1984). This proposal was further developed by Frankel and Brown (1984) and Brown (1988, 1989), who described methods to select a core collection using information on the origin and characteristics of the accessions.

The development of a core collection in peanut would provide a working collection which could be extensively evaluated. This information would allow estimation of genetic variation for economically important traits. These data would also provide an indication of clusters in the collection that should be extensively examined for particular traits. The objective of this research was to develop a core collection for peanut.

C.C. Holbrook and W.F. Anderson, USDA-ARS, Tifton, GA 31793; and R.N. Pittman, USDA-ARS, Griffin, GA 30223. Received 14 Sept. 1992. *Corresponding author.

Published in Crop Sci. 33:859–861 (1993).

MATERIALS AND METHODS

Data on peanut in the GRIN included country of origin and measurements on plant type, pod type, seed size, testa color, number of seed per pod, and average seed weight. Available information on accessions varies from just country of origin to information on all seven variables. The U.S. peanut germplasm collection was first stratified by country of origin and then divided into nine sets based on the amount of additional information available for accessions and on the number of accessions per country of origin. The procedures used to select accessions for the core collection from each set were as follows:

SET 1 — Consists of accessions from various countries with five or fewer accessions per country. Set 1 contained a total of 99 accessions and 10 of these were randomly selected for this core without regard to country of origin.

SET 2 — Consists of accessions from 14 countries for which morphological data was unavailable. Accessions were sorted by country of origin and then 10% were randomly selected from each country. Set 2 contained 1597 accessions and 160 were selected.

SET 3 — Consists of accessions from various countries with 15 or fewer accessions per country. One accession was randomly selected from each country represented by 14 or fewer accessions, and two accessions were randomly selected from each country represented by 15 accessions. Set 3 contained 139 accessions and 13 were selected.

SET 4 — Consists of accessions with data for all six morphological variables. Data were first sorted by country of origin and then the cluster procedure (SAS Institute, 1985) was used to divide accessions into an appropriate number of clusters. The tree procedure (SAS Institute, 1985) was used to print the clusters and 10% of the accessions in each cluster were randomly selected. Set 4 contained 1776 accessions and 210 were selected.

SET 5 — Consists of >300 accessions with data for all six morphological variables from each of two countries, Brazil and Zambia. The cluster procedure is not practical for data sets that are this large (SAS Institute, 1985). The data were sorted by country of origin and then the fastclus procedure (SAS Institute, 1985) was used to divide the accessions into clusters. Random sampling was used to select 10% of the accessions from each cluster. Set 5 contained 827 accessions and 94 were selected.

SET 6 and 7 — Consists of accessions with data for five variables (no plant type). Accessions were selected using the procedure outlined for Set 4. Set 6 and 7 contained 756 accessions and 94 were selected.

SET 8 — Consists of accessions with data for four variables (no plant type and no seed weight). Accessions were selected using the procedure outlined for Set 4. Set 8 contained 1845 accessions and 210 were selected.

SET 9 — Consists of accessions that did not fit into any of the above sets. These accessions were grouped by country of origin and 10% of the accessions were randomly selected from each country. Set 9 contained 393 accessions and 40 were selected.

Abbreviations: GRIN, germplasm resources information network; PI, plant inventory number.

RESULTS AND DISCUSSION

The selection procedure used for developing a peanut core collection resulted in the selection of 831 accessions from the U.S. germplasm collection. The means for the six plant and seed characteristics in the GRIN for this core collection and for the entire collection were similar (Table 1). In addition, the range for the variables in this core collection is as large as the range in the entire collection for five of the six variables. This suggests that this core collection is a representative sample of the entire collection and that the genetic variation expressed for these traits in the entire collection has been preserved in this core collection.

To maximize the usefulness of the peanut core collection, the relationship between individual accessions and the clustering procedure used to develop the core collection must be presented in two table formats. The first format (Table 2) lists the accessions in numerical order so that the cluster designation for individual accessions can be rapidly determined. The second format (Table 3) lists accessions by clusters so that all accessions within a cluster can be rapidly determined. Table 3 also indicates the accessions that have been included in the core collection. Tables 2 and 3 are parts of tables that are much too large to include in a journal article. The complete tables are available as ASCII files on diskette, free of charge, from the corresponding author.

There are several ways a peanut core collection may be useful. A core collection can serve as a logical and efficient starting point for projects involving screening of the germplasm collection for sources of desirable alleles. This information could then be used to determine which clusters should be more extensively examined.

A core collection also provides a logical subset of germplasm to examine when it is not feasible to examine the entire collection. For example, this peanut core collection can be evaluated for resistance to preharvest aflatoxin contamination which is the most serious challenge facing the U.S. peanut industry. At present, there is very little information on the amount of variation that may exist in *A. hypogaea* for resistance to preharvest aflatoxin contamination. Using current available techniques, it would take at least 10 yr to examine the entire germplasm collection. Using a core collection, we can estimate the amount of genetic variation in the entire collection and possibly identify sources of resistant alleles in about one-tenth of the time.

The development of this peanut core collection has also resulted in a subset of accessions that is representative of the entire collection and for which a larger amount of seed is readily available. This will be valuable in rapidly addressing problems that may arise in the future. For example, if a new disease arises, this core collection can serve as a working collection that can be rapidly examined for sources of resistance.

The core collection concept can also be used with known sources of desirable traits to identify other accessions in the entire collection that should be examined. For example, PI 203396 is the resistant parent of 'Southern Runner', the only peanut cultivar with resistance to late leafspot [*Phaeoisariopsis personata* (Berk. & M.A. Curtis) Arx syn. *Cercosporidium personatum* (Berk. & M.A. Curtis)] disease. Table 2 shows that PI 203396 is included in Cluster 1 from the Brazilian accessions in Set 5. Table 3 lists accessions by clusters and shows that there are 13 accessions included in the cluster containing PI 203396. These accessions are theoretically genetically similar to PI 203396 and may also be valuable sources of resistance.

This core collection may also be useful in decisions on parental selection for peanut breeding. Sources of resistance from different clusters are more likely to have different, and possibly complementary, genes for resistance than sources of resistance from the same cluster.

Resources available for evaluation of peanut germplasm are very limited. This precludes an extensive evaluation of the entire U.S. peanut germplasm collection. This core collection will provide a working collection which can be extensively examined for all economically important traits. The resulting data should provide much needed information on genetic variability in peanut and could be used to examine relationships among traits. The data could also be used to make inferences about appropriate management of the entire collection. For example, if evaluation of this core collection shows little or no variability for a given trait, a decision on future acquisition can be made with the given trait in mind.

We selected this peanut core collection using all currently available information in the GRIN for country of origin and morphology of peanut accessions. The use of

Table 1. Means for six morphological variables for the entire peanut germplasm collection and for the core collection and the range of these variables in the core collection expressed as a percentage of the range in the entire collection.

	Entire collection	Core collection	
Variable†	Mean	Mean	Range of entire collection
			%
Plant type	2.16	2.17	100
Pod type	3.96	3.87	100
Seed size	5.13	5.14	100
Testa color	2.95	3.05	100
Seed per pod	3.23	3.22	80
Seed weight	50.81	51.46	100

† GRIN variables with the following specifications: Plant type: 1 = bunch; 2 = runner; 3 = Valencia; 4 = bunch-runner; 5 = runner-bunch; 6 = Virginia; 7 = Spanish; 8 = mixed; 9 = Chinese. Pod type: 1 = Spanish; 2 = Spanish-Valencia; 3 = Valencia-Spanish; 4 = Valencia; 5 = Virginia; 6 = Chinese; 7 = mixed. Seed size: 1 = small (71–90 seed per 28 g); 5 = medium (51–70 seed per 28 g); 8 = large (36–50 seed per 28 g); 9 = jumbo (20–35 seed per 28 g). Testa color: 1 = tan; 2 = pink; 3 = red; 4 = purple; 5 = gaspeado (tan or red with purple fleck); 6 = dark purple; 7 = white; 8 = variegated; 9 = other colors/or mixed. Seed per pod: 1 = one; 2 = two; 3 = three; 4 = more than three; 5 = two and three; 6 = three and four. Seed weight: g per 100 seed.

Table 2. Part of a 51-page table that shows the relationships for all peanut plant introductions to the clustering procedure used to select this core collection.

PI number	Set	Country†	Cluster within country within set	PI number	Set	Country†	Cluster within country within set
200441	4	J2	4	203396	5	B4	1
200443	4	J2	3	203397	5	B4	1
200444	4	J2	2	207530	4	Z1	4
203394	5	B4	13	207531	4	Z1	1
203395	5	B4	1	207532	4	Z1	1

† B4 = Brazil; J2 = Japan; Z1 = Zaire.

Table 3. Part of a 39-page table that lists all plant introductions within each of the clusters used to select this peanut core collection.

Core number	PI number
Set 5–Brazil Cluster 1	
	152113
	337306
	337297
	203397
	152121
	203395
394†	203396
	149633
	337300
	221067
	298838
	336977
	152131

† This accession was randomly selected from Cluster 1 of the Brazilian accessions in Set 5 for inclusion in this core collection and was assigned the core number 394.

this data may have introduced some bias into the selection of this core collection. The country of origin sometimes indicates only the country donating the sample to the National Plant Germplasm System, not its ultimate origin. Previous to the selection of this core collection, much effort was devoted to improving the accuracy for country of origin information included in the GRIN. This work resulted in numerous changes in country of origin information in comparison to the last published peanut germplasm catalog (U.S. Department of Agriculture, 1985). Although some discrepancies probably still exist with country of origin information, this is the best currently available information that can be used to categorize peanut germplasm based on geographic origin.

This core collection may also be biased because the morphological data included in the GRIN for the peanut germplasm collection were not collected in a single environment. An additional bias may have arisen because of correlations between morphological variables. Additional research will be needed to determine how serious these biases are.

In summary, this research has resulted in the development of a core collection of peanut germplasm. This core collection should result in more effective and efficient utilization of peanut germplasm. This core collection can be revised periodically as additional information and accessions are obtained.

ACKNOWLEDGMENTS

This work was supported in part by a grant from the National Peanut Foundation.

REFERENCES

Banks, D.J. 1976. Peanuts: Germplasm resources. Crop Sci. 16:499–502.

Brown, A.H.D. 1988. The case for core collections. p. 136–156. *In* A.H.D. Brown et al. (ed.) The use of plant genetic resources. Cambridge Univ. Press, Cambridge.

Brown, A.H.D. 1989. Core collections: A practical approach to genetic resources management. Genome 31:818–824.

Frankel, O.H. 1984. Genetic perspectives of germplasm conservation. p. 161–170. *In* W.K. Arber et al. (ed.) Genetic Manipulation: Impact on Man and Society. Cambridge Univ. Press, Cambridge.

Frankel, O.H., and A.H.D. Brown. 1984. Current plant genetic resources — a critical appraisal. p. 1–11. *In* Genetics: New Frontiers Vol. IV. Oxford & IBH Publ. Co., New Delhi.

Knauft, D.A., and D.W. Gorbet. 1989. Genetic diversity among peanut cultivars. Crop Sci. 29:1417–1422.

SAS Institute. 1985. SAS user's guide. Statistics. Version 5 ed. SAS Inst., Cary, NC.

U.S. Department of Agriculture. 1985. Catalog of seed and vegetative stock available from the southern regional plant introduction station: Peanut (*Arachis* species). U.S. Gov. Print. Office, Washington, DC.