

Cowpea [*Vigna unguiculata* (L.) Walp.] core collection defined by geographical, agronomical and botanical descriptors

V. Mahalakshmi¹, Q. Ng^{1†}, M. Lawson¹ and R. Ortiz^{2*}

¹International Institute of Tropical Agriculture (IITA), PMB 5320, Oyo Road, Ibadan, Nigeria and ²Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), Apdo. Postal 6-641, 06600 Mexico, D.F, Mexico

Received 24 May 2006; Accepted 6 October 2006

Abstract

Cowpea is a drought-tolerant food legume grown in the savannah regions of the tropics and subtropics. The International Institute for Tropical Agriculture (IITA) holds the world collection of 15,003 cultivated cowpea from 89 countries in its genebank. In excess of 12,000 accessions were characterized for 28 agrobotanical descriptors. The entire collection was first stratified by country of origin and biological status. Land race accessions (10,227) with information on origin and characterization data were grouped using clustering procedures. The clustering analysis was based on Euclidean distances between and among accessions. Accessions within each country were then grouped hierarchically, according to their similarity. The number of clusters selected for countries varied from 2 to 20, depending on the size of collection for that country. A percentage number of accessions (5–25%) was chosen from each country, based on the size of the collection and its proximity to the centre of diversity. The number of accessions from each cluster in a country was then chosen randomly, based on the predetermined percentage. In countries where the numbers of accessions were few (<10), the percentage chosen from those countries would be higher and may go up to 100% to ensure that at least one accession was chosen from every country. Accessions with no characterization information were treated as a group, and accessions were chosen randomly, based on their geographical distribution. In the process of selection, 200 accessions that are known to be resistant to pests and diseases and not originally chosen through the selection processes were, nevertheless, kept as part of the core collection. Following these procedures, a total of 1701 accessions of landraces were chosen from the entire collection. Using the same procedures, a subset of 225 accessions was chosen from 1422 advanced cultivars and breeding or research lines. A third subset of 130 accessions was chosen from 838 accessions that either lack information on origin or biological status, and six accessions of wild and weedy forms from the available 64 wild and weed accessions were added. Thus a core collection of 2062 accessions of cowpea was constituted. The diversity in the core collection was similar to that of the entire collection and correlated traits that may be linked were also preserved in the core collection. This core collection of cowpea provides an opportunity for further exploitation of the cowpea germplasm for improvement of this crop.

Keywords: core collection; cowpea; descriptors; diversity; sampling; *Vigna unguiculata*

* Corresponding author. E-mail: r.ortiz@cgiar.org

† Present address: Food and Agriculture Organization (FAO) Regional Office for Asia and Pacific, 39 Phrqa Atit Road, Bangkok 10200, Thailand.

Introduction

Cowpea [*Vigna unguiculata* (L.) Walp.] is a grain legume grown in savannah regions of the tropics and subtropics. It is adapted to very harsh environments and is considered a drought-tolerant legume. The major cowpea-growing countries are Nigeria, Niger, Mali, Senegal, Togo, Benin, Ghana, Chad in West Africa; Tanzania, Somalia, Kenya, Zambia, Zimbabwe, Botswana and Mozambique in eastern and southern Africa; India, Pakistan, Sri Lanka, the Philippines, Bangladesh, Indonesia and China in Asia; and Brazil, West Indies, Cuba and southern USA in America. The Food and Agriculture Organization (FAO) estimates that 3.7 million tonnes of cowpea dry grains were produced worldwide in 2003. Nigeria produced 2.1 million tonnes of this total, making this country the world's largest producer, followed by Niger (400,000 tonnes) and Mali (100,000 tonnes). The total area grown to cowpea worldwide was 9.8 million ha, with about 91% of this in West Africa. While the world average yield was 378 kg ha⁻¹, it was 440 kg ha⁻¹ in Nigeria and 114 kg ha⁻¹ in Niger.

The value of cowpea lies in its high protein content, its ability to tolerate drought, and the fact that it fixes atmospheric nitrogen which allows it to grow on, and improve, poor soils. The major pests attacking cowpea plants are flower thrips (*Megalurothrips sjostedti*), legume pod borer (*Maruca vitrata*) and a complex of pod-sucking bugs. Bruchid (*Callosobruchus maculatus*) causes damage to stored dry cowpea grains. Fungal diseases affecting cowpea include stem and root rots and leaf spot diseases. Viruses cause a mosaic of diseases and mottle symptoms in cowpea. The parasitic weed *Striga gesneroides* can severely damage cowpea plants. Losses due to pest attack or disease can be as high as 90% (Singh, 2002).

Cowpea belongs to the family *Leguminosae*, subfamily *Papilionoideae*, tribe *Phaseoleae* and genus *Vigna*. *Vigna* is a large and immensely variable genus consisting of more than 85 species divided into seven subgenera. Cowpea falls under the subgenus *Ceratotropis* under the species *unguiculata*. Cowpea is a diploid with chromosome number $2n = 2x = 22$ and with many cross-compatible or incompatible wild species. For example, *V. rhomboidea* (Ng, 1995) is a wild perennial creeping or climbing herb, which develops a large woody root stock and shows incompatibility with all other taxa within *V. unguiculata*. The centre of diversity of wild *V. unguiculata* and *V. rhomboidea* is in southern and south-eastern Africa, whereas the centre of diversity of cultivated cowpea *V. unguiculata* is in West Africa in the area comprising savannah regions of northern Nigeria, southern Niger, Burkina Faso, northern Benin, Togo and northern Cameroon (Padulosi, 1993).

The International Institute for Tropical Agriculture (IITA) genebank holds over 15,000 accessions of cultivated cowpea collected from 89 countries. At present, over 12,000 accessions have been characterized for 28 agrobotanical descriptors. Germplasm collections were originally set up to preserve the genetic diversity of crop species and their wild relatives. Given that such genetic diversity of crops has an economic value, conservation for use has been the driving force behind many genebanks. The sheer number of accessions making up germplasm collections could be an obstacle for their full exploitation, evaluation and utilization to impact the crop improvement or breeding programmes. In this regard, genetic diversity of such a large collection may not have been adequately evaluated for various biotic and abiotic stresses, due to resource and time constraints. It is impractical to evaluate such large collections in detail as it would be expensive and time consuming. This task could be more easily fulfilled by developing subsets of the whole collection, called active working collections by Harlan (1972) and core collections by Frankel and Brown (1984). A core collection should include a maximum of the genetic variation contained in the whole collection with minimal repetitiveness, ideally conserving at least 70% of the alleles in the whole collection (Brown, 1989a, b). The ninth activity of the Global Plan of Action addresses the issue of promoting the use of germplasm through expanding the characterization and evaluation, and developing a number of core collections (FAO <http://www.fao.org/WAICENT/FaoInfo/Agricult/AGP/AGPS/GpaEN/gpaact9.htm>).

Core collections have been established for many crop species, using morphological and genetic marker variation; e.g. bean *Phaseolus vulgaris* L. (Tohme *et al.*, 1995); barley *Hordeum vulgare* (Knüpfper and van Hintum, 1995); chickpea *Cicer arietinum* L. (Hannan *et al.*, 1994; Upadhyaya *et al.*, 2001); peanut *Arachis hypogaea* L. (Holbrook *et al.*, 1993; Upadhyaya *et al.*, 2003); quinoa *Chenopodium quinoa* (Ortiz *et al.*, 1998); sweetpotato *Ipomoea batatas* L. (Huamán *et al.*, 1999); potato *Solanum tuberosum* L. (Huamán *et al.*, 2000a, b); lentil *Lens culinaris* Medic (Ers-kine and Muehlbauer, 1991); and cassava *Mannihot esculanta* Crantz (Cordeiro *et al.*, 1995). A core collection made up of 699 accessions from 7737 cowpea accessions is available from the GRIN-USA collection at Griffin (<http://www.ars-grin.gov/ars/NoPlains/FtCollins/plants.htm>). The objective of this study was to develop a core collection of cowpea using the 28 agrobotanical characters in the larger world cowpea collection held at IITA.

Materials and methods

The cowpea collection at IITA consists of 15,003 accessions from 89 countries (Table 1), with the majority of

Table 1. Number of accessions based on their biological status in entire and core collections of cowpea at IITA

Biological status	Number of accessions	
	Entire	Core
Advanced cultivars	14	2
Breeding/research material	1408	223
Others	48	2
Traditional cultivars/landrace	12 594	1701
Unknown	875	128
Weedy	38	3
Wild	26	3
Total	15 003	2062

Frequency distribution of accession between entire and core by biological status $\chi^2 = 8.80$; $P \leq 0.18$.

the accessions from Africa. Africa is the primary centre of diversity for cowpea and this collection should represent the vast genetic diversity of this crop. Of this, 12,594 accessions are landraces collected from farmer's fields and 10,277 of these have been characterized using the cowpea descriptors. Advanced cultivars and breeding or research lines constitute 1422 accessions and 1291 of these have been characterized. Of the 923 accessions of unknown biological status, 880 have been characterized. Sixty-four accessions are weedy or wild cowpea species.

Data on the 28 agronomical and botanical descriptors (see Table 3) were collected following the International Board for Plant Genetic Resources descriptor list (IBPGR, 1983). Although the characterization of the germplasm was done over years, as the majority of the descriptors are qualitative and considered highly heritable, the influence of environment is expected to be minimal, if any. For each accession the qualitative traits were converted to binary values, as either presence or absence in that trait class. The data recorded on the quantitative traits were standardized using the range of each variable to eliminate scale differences (Upadhyaya *et al.*, 2006).

The entire collection was stratified by biological status followed by the country of origin. For each country and biological status, accessions with complete information on the 28 descriptors were chosen for clustering into groups. In certain cases a full set of 28 descriptors was not available. In such cases, the minimum number of descriptors that could include a maximum number of accessions was determined. For example, there are 103 accessions from Togo, but only 73 had complete data for 17 descriptors for cluster analysis. From the remaining 30 accessions few were included in the core set. A set of 18 accessions, based on the relative size of clusters, was selected. A proportionate number of

accessions was also chosen randomly from those accessions that were not included for cluster analysis due to lack of data.

The data were subjected to a hierarchical cluster algorithm based on Euclidean distances between and among accessions, using Ward (1963), and the number of clusters retained were those at an R^2 (squared multiple correlation) in excess of 75% (SAS Institute, 1989). Ward's algorithm is often used for the agglomerative cluster hierarchy when a precise solution for a specified number of groups is not practical. Given N sets, this procedure reduces them to $N-1$ mutually exclusive sets by considering the union of all possible $N(N-1)/2$ pairs and selecting a union having a maximal value for the functional relation, or objective function, which reflects the criterion chosen. By repeating this process until only one group remains, the complete hierarchical structure and a quantitative estimate of the loss associated with each stage in the grouping can be obtained. For each country by biological group, the percentage of accessions to be chosen was determined according to the size of the collection, importance of cowpea in the country's agriculture and country's relevance to the centre of diversity for cowpea. The number of accessions chosen in each cluster group in the core collection was determined as a proportion of the number of accessions in the cluster group relative to the total number of accessions in the entire collection. At least one accession from each cluster group was chosen to ensure that all the cluster groups were represented. Accessions were chosen randomly within cluster members.

The diversity for each trait in the entire and core collections was measured using the Shannon–Weaver diversity index (Shannon and Weaver, 1946). Homogeneity of distribution of the entire and core collections for each trait was measured by comparing the frequency distribution of the trait using the Chi-squared (χ^2) test. For the quantitative traits the means and standard deviation of the entire and core collections were also used as a measure for comparison. Correlations between the descriptors in the core and entire collections were estimated independently, to assess whether any of the trait correlations, which may be under genetic control, were lost in the sampling of the core collection.

Results and discussion

A cowpea core collection consisting of 2062 accessions covering most of the diversity found in the entire collection of 15,003 accessions of cultivated cowpea held at IITA genebank was defined. The analysis of frequency distribution between the entire and this core collection showed

heterogeneity between their distributions ($\chi^2 = 478$; $P = 0.001$). This was expected, as the sampling procedure that ensures representation from all countries, including those with very small collections, would result in a higher proportion of accessions in the core from these countries. West Africa and East and Central Africa, which constitute the major centres of diversity for cultivated cowpea, are well represented in the core collection. Therefore, this core collection should cover most of the diversity of this crop.

The frequency distribution between the core and entire collections based on the biological status was homogeneous (Table 1). This suggests that the core collection shows a similar pattern of distribution as the entire collection for biological status. Although the number of accessions selected from the wild and weedy forms is rather small (6), a priori information suggests that this group might contain useful genes for certain traits. Hence, it may be useful based on the need to define a set containing these wild and weedy accessions as their number is small.

The 28 agronomic and botanical descriptors used for grouping of accessions cover a wide range of phenotypic variation (see Table 3). For most countries a major proportion (>70%) of accessions had information on the full set of 28 descriptors. The clustering was based on the accessions with full information. From the remaining accessions that were not included due to lack of descriptor data, a few accessions were selected randomly, representing the same percentage as those with information. Similarly, for countries where data for all the 28 descriptors were not available, the clustering was based on the number of descriptors that accounted for most accessions from that country. This process resulted in lower number of descriptors for 11 countries in the landrace biological status [e.g. in Yemen and Oman the descriptor number of nodes (IBPGR, 1983) was missing, while Brazil, Russia and Central African Republic had the highest number of

descriptors missing, i.e. 8]. As the stratification was within a country and biological status, this approach would not bias the grouping strategy. It would ensure maximal information available across a large number of accessions, and it would cover the diversity from country and biological group.

The major interest after developing a core collection was to examine the extent of diversity retained in the core collection that was present in the entire collection. For any quantitative trait this was measured by comparing the mean and the standard deviation of the trait in the core and entire collection (Table 2). For all the quantitative traits the core collection retained the diversity that was present in the entire collection, suggesting that a non-biased procedure was used for defining the core and also that the diversity was fully represented in the core for these traits.

For qualitative traits with distinct classes, the comparison of the frequency distribution showed differences in the distribution for some traits (Table 3). Often the interest in a trait is in the diversity of the trait and how it is retained in the core. For all the descriptors, the diversity in the core and in the entire collection was similar. Cowpea descriptors are very diverse in the entire collection held at IITA genebank, and the ensuing core collection retains this diversity across all the descriptors.

Traits are often correlated with each other, and while developing a core collection these associations could be lost. The correlations between the descriptors in the core and entire collection showed no significant differences for the qualitative traits. Only pod length and number of branches, which showed no relationship in the entire collections, showed a significant relationship in the core collection (Table 4). Early flowering was associated with small leaflets, fewer pods and fewer seeds per pod, and this association was retained in the core but the magnitude was lower.

Table 2. Number of accessions, mean and standard deviation for the quantitative traits in the entire collection and in the core collection of cowpea at IITA

Variable	Entire collection			Core collection		
	Number	Mean	Std Dev.	Number	Mean	Std Dev.
Days to first flowering	12 215	52.45	12.24	1953	52.04	12.73
Days to first ripe pod	11 011	74.84	13.56	1759	74.41	13.43
Number of main branches	12 262	5.43	1.67	1966	5.56	1.70
Number of main stem nodes	11 094	10.77	4.91	1760	11.00	5.07
Terminal leaflet length (mm)	12 268	116.51	22.62	1966	117.07	23.83
Terminal leaflet width (mm)	12 261	76.23	16.70	1964	75.39	17.17
Pod length (mm)	10 125	152.45	33.92	1700	156.16	38.83
Pod width (mm)	9723	7.08	1.32	1628	7.17	1.4
Number of locules per pod	10 102	14.15	2.89	1697	14.51	3.04

Table 3. Shannon–Weaver diversity index and comparison of frequency distribution (χ^2) for qualitative (*) and quantitative descriptors between the entire and core collections of cowpea held at IITA genebank

Descriptor	Shannon–Weaver diversity index		Frequency distribution	
	Entire	Core	χ^2	Probability
Number of main branches	1.21	1.24	13.96	0.016
Seed crowding in pod*	1.08	1.07	2.43	0.488
Determinacy*	0.30	0.24	7.93	0.005
Days to first flowering	1.23	1.19	12.07	0.038
Days to first ripe pod	1.29	1.27	5.31	0.379
Pigmentation pattern on flower*	1.57	1.49	26.54	0.001
Plant growth habit*	1.67	1.64	12.19	0.058
Terminal leaflet length (mm)	1.42	1.46	10.92	0.053
Terminal leaflet shape*	0.77	0.81	8.35	0.039
Terminal leaflet width (mm)	1.42	1.45	8.37	0.137
Number of locules per pod	1.40	1.44	35.84	<0.001
Number of main stem nodes	1.34	1.38	11.25	0.047
Pattern of pigmentation on plant*	0.99	0.99	3.55	0.314
Pod attachment to peduncle*	0.77	0.75	4.62	0.099
Pod length (mm)	1.34	1.39	27.57	<0.001
Green pod pigmentation pattern*	1.13	1.10	7.79	0.168
Number of pods per peduncle	0.52	0.55	13.80	0.001
Pod shape*	0.55	0.52	1.81	0.404
Pod width (mm)	1.62	1.62	4.58	0.469
Raceme position*	0.92	0.92	16.00	0.001
Seed eye colour*	1.63	1.65	20.44	0.025
Seed eye pattern*	1.56	1.50	16.60	0.058
Seed length coded*	1.60	1.77	62.34	<0.001
Seed size*	1.59	1.60	20.16	0.010
Seed thickness*	1.34	1.43	33.12	<0.001
Testa texture*	1.10	1.04	16.66	0.010
Twining tendency*	0.83	0.89	14.38	0.001
Growth vigour*	1.18	1.20	6.10	0.107

This cowpea core collection, constituted to cover the entire diversity in the collection held at IITA, will open up new opportunities for assessing and evaluating germplasm for various biotic and abiotic stresses. Such evaluation should facilitate discovery of genes for resistance to traits of interest, and link them to both descriptor information and geographical information to probe the collections further in search of these genes in different genetic and agronomic backgrounds.

One of the major challenges today for genebank curators is to ensure effective and economical ways of conserving diversity. Duplication and redundancy in collections has always been a major challenge faced by the curators. Core collections represent the entire diversity in the crop but have limited use to those users who are not interested in entire diversity but rather in accessions meeting a specific diversity of traits or domain. With the modern revolution in information and computing technology it is possible to sample diversity based on traits of interest, e.g. disease

resistance, using procedures such as core selector (van Hintum, 1999; Mahalakshmi *et al.*, 2003).

Core collections should not be considered static but dynamic to accommodate the various needs. If sufficient neutral molecular markers were available to cover the entire genome of cowpea, it would be possible to relate the diversity found outside the centres of diversity to those from the centre of diversity. The availability of molecular markers provides new avenues to probe these collections in search of new allelic forms of existing genes, and also in discovery of new genes. Seedborne viruses are one of the major challenges limiting the movement of cowpea germplasm worldwide. Many countries have stringent regulations for importing seeds. Virus testing and the production of pathogen- and pest-free seeds of cowpea are currently under way at IITA, but the cleaning of 15,003 accessions of cowpea needs a planned strategy, which will allow the movement of seeds and also cover users' needs. Development of this core collection has served this

Table 4. Correlation coefficients between quantitative traits in the entire collection (below diagonal) and core collection (above diagonal) of cowpea held at IITA genebank. Absolute values of 0.020 are highly significant at $P > 0.01$

	Days to first flowering	Days to first ripe pod	Number of main branches	Nodes of main stem (#)	Terminal leaflet length	Terminal leaflet width	Pod length	Pod width	Locule number per pod
Days to first flowering	0.811	0.776	0.116	-0.050	-0.171	-0.119	-0.145	0.188	-0.196
Days to first ripe pod	0.101	0.083	0.095	-0.050	-0.262	-0.194	-0.141	0.144	-0.278
Number of main branches	-0.066	-0.085	0.362	0.419	0.078	0.047	-0.058	0.122	0.090
Nodes of main stem (#)	-0.223	-0.309	0.081	0.403	0.370	0.194	0.091	0.148	0.223
Terminal leaflet length	-0.154	-0.212	0.026	0.249	0.668	0.657	0.368	0.298	0.293
Terminal leaflet width	-0.151	-0.175	-0.017 ^{ns}	0.101	0.371	0.253	0.227	0.193	0.271
Pod length	0.156	0.095	0.118	0.156	0.266	0.182	0.434	0.378	0.541
Pod width	-0.225	-0.323	0.098	0.186	0.291	0.260	0.555	0.150	0.152
Locule number per pod									

purpose and need. The seeds of this cowpea core collection are virus tested and available for distribution from the IITA genebank (iita@cgiar.org).

References

- Brown AHD (1989a) A case for core collections. In: Brown AHD, Frankel OH, Marshall DR and Williams JT (eds) *The Use of Plant Genetic Resources*. Cambridge: Cambridge University Press, pp. 136–156.
- Brown AHD (1989b) Core collections: a practical approach to genetic resources management. *Genome* 31: 818–824.
- Cordeiro CMT, Morales EAV, Ferreira P, Rocha DMS, Costa IRS and Valios ACC (1995) Towards a Brazilian core collection of cassava. In: Hodgkin T, Brown AHD, van Hintum ThJL and Morales BAV (eds) *Core Collections of Plant Genetic Resources*. New York: International Plant Genetic Resources Institute (IPGRI), John Wiley & Sons, pp. 155–168.
- Erskin W and Muehlbauer FJ (1991) Allozyme and morphological variability, outcrossing rate and core collection formation in lentil germplasm. *Theoretical and Applied Genetics* 83: 119–125.
- Frankel OH and Brown AHD (1984) Plant genetic resources today: a critical appraisal. In: Holden JHW and Williams JT (eds) *Crop Genetic Resources: Conservation and Evaluation*. London: George Allen & Unwin, pp. 249–257.
- Hannan RM, Kaiser WJ and Muehlbauer FJ (1994) *Development and Utilization of the USDA Chickpea Germplasm Core Collection*. Agronomy Abstracts 1994. Madison, Wisconsin: American Society of Agronomy, p. 217.
- Harlan JR (1972) Genetics of disaster. *Journal of Environmental Quality* 1: 212–215.
- Holbrook CC, Anderson WF and Pittman RN (1993) Selection of core collection from the U.S. germplasm collection of peanut. *Crop Science* 33: 859–861.
- Huamán Z, Aguilar C and Ortiz R (1999) Selecting a Peruvian sweet potato core collection on the basis of morphological, ecogeographical and disease reaction data. *Theoretical and Applied Genetics* 98: 840–844.
- Huamán Z, Ortiz R and Gómez R (2000a) Selecting a *Solanum tuberosum* subsp. *Andigena* core collection according to morphological, geographical, disease and pest descriptors. *American Journal of Potato Research* 77: 183–190.
- Huamán Z, Ortiz R, Zhang DP and Rodríguez F (2000b) Isozyme analysis of entire and core collections of *Solanum tuberosum* spp. *andigena* potato cultivars. *Crop Science* 40: 273–276.
- IBPGR (1983) *Cowpea Descriptors*. Rome: International Board of Plant Genetic Resources.
- Knüpffer H and van Hintum ThJL (1995) The barley core collection: an international effort. In: Hodgkin T, Brown AHD, van Hintum ThJL and Morales BAV (eds) *Core Collections of Plant Genetic Resources*. New York: International Plant Genetic Resources Institute (IPGRI), John Wiley & Sons, pp. 171–178.
- Mahalakshmi V, van Hintum ThJL and Ortiz R (2003) Enhancing germplasm utilization to meet specific users' needs through interactive stratified core selections. *Plant Genetic Resources Newsletter* 136: 14–22.

- Ng NQ (1995) Cowpea. In: Smart J and Simmonds NW (eds) *Evolution of Crop Plants*. 2nd edn. Burnt Mill, Harlow, UK: Longman Scientific and Technical, pp. 326–332.
- Ortiz R, Ruiz-Tapia EN and Mújica-Sánchez A (1998) Sampling strategy for a core collection of Peruvian quinoa germplasm. *Theoretical and Applied Genetics* 96: 475–482.
- Padulosi S (1993) Genetic diversity, taxonomy, and ecogeographic survey of the wild relatives of cowpea (*V. unguiculata*). PhD Thesis, University of Louvain-La Neuve, Belgium.
- SAS Institute (1989) *SAS/SAT User Guide version 6*. 4th edn. Cary, North Carolina: SAS Institute.
- Shannon CE and Weaver W (1946) *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.
- Singh BB (2002) Breeding cowpea varieties for resistance to *Striga gesnerioides* and *Alectra vogelii*. In: Fatokun CA, Tarawali SA, Singh BB, Kormawa PM and Tamo M (eds) *Challenges and Opportunities for Enhancing Sustainable Cowpea Production*. Ibadan, Nigeria: International Institute of Tropical Agriculture, pp. 154–163.
- Tohme J, Jones P, Beebe S and Iwanaga M (1995) The combined use of agroecological and characterization data to establish the CIAT *Phaseolus vulgaris* core collection. In: Hodgkin T, Brown AHD, van Hintum ThJL and Morales BAV (eds) *Core Collections of Plant Genetic Resources*. New York: International Plant Genetic Resources Institute (IPGRI), John Wiley & Sons, pp. 95–108.
- Upadhyaya HD, Bramel PJ and Singh S (2001) Development of a chickpea core subset using geographic distribution and quantitative traits. *Crop Science* 41: 206–210.
- Upadhyaya HD, Gowda CLL, Pundir RPS, Goppal Reddy V and Singh S (2006) Development of core subset of finger millet germplasm using geographical origin and data on 14 quantitative traits. *Genetic Resources and Crop Evolution* 53: 679–685.
- Upadhyaya HD, Ortiz R, Bramel PJ and Singh S (2003) Development of a groundnut core collection using taxonomical geographical descriptors. *Genetic Resources and Crop Evolution* 50: 139–148.
- van Hintum ThJL (1999) The core selector, a system to generate representative selections of germplasm collections. *Plant Genetic Resources Newsletter* 118: 64–67.
- Ward JH (1963) Hierarchical grouping to optimize an objective function. *Journal of American Statistical Association* 58: 236–244.