Natural Rubber Research, 21 (1&2):67-75, 2008

ESTABLISHMENT OF A CORE COLLECTION OF WILD *HEVEA BRASILIENSIS* (WILLD.EX ADR.DE JUSS.) MUELL. ARG. GERMPLASM IN INDIA

Saji T. Abraham, Ramesh, B. Nair and Y. Annamma Varghese
Rubber Research Institute of India, Rubber Board P.O. Kottayam – 686 009, Kerala, India

Abraham, S.T., Nair, R.B. and Varghese, Y.A. (2008). Establishment of a core collection of wild *Hevea brasiliensis* (Willd.ex Adr.de Juss.) Muell. Arg. germplasm in India. *Natural Rubber Research*, 21(1&2):67-75.

The present study is the first attempt towards establishment of a core collection of *Hevea brasiliensis* germplasm. The method adopted makes use of both quantitative data identified as morphological markers along with a set of qualitative morphological characterization data for evaluating the diversity in the base collection using principal component scores and Shannon-Weaver Diversity Index. The functional relationship between the cumulative contribution of the wild *Hevea* accessions to the total sum of squares and the number of accessions, which was essentially a logistic regression model, was used to decide the appropriate size for the core set. The method selected 23 wild accessions of wild *H. brasiliensis* constituting 28.75% of the base population.

Key words: Core collection, Germplasm, *Hevea brasiliensis.*

INTRODUCTION

Germplasm collections form an invaluable reservoir of genetic diversity of agriculturally important crops and of native plant species, many of which are being lost worldwide due to habitat destruction, invasion of foreign species and reliance on fewer high yielding geno types. In view of the long-term needs of mankind, countries all over the world have set up facilities for conservation and management of large collections of germplasm of various crops for their use by breeders and research workers either directly or indirectly. The usefulness of the collections largely depends on the extent of genetic diversity present in such collections. Most gene banks include a large number of accessions posing serious problems in its management and accessibility for breeding purposes besides imposing severe limitations on land and resources (Frankel, 1984; Frankel and Brown, 1984).

Recognizing the problems associated with large germplasm collections, Frankel (1984) proposed the concept of core collection, which was further elaborated by Frankel and Brown (1984) and Brown (1989). A core collection is defined as a subset consisting of a limited set of accessions derived from the base collection, chosen to represent the genetic spectrum in the whole collection. It is a manageable-sized, structured sample representing the diversity of the collection. The degree to which a collection represents the genetic diversity available in the species is more

Correspondence: Saji T. Abraham (Email: sajiabraham@rubberboard.org.in)

important than its size. Nevertheless, capturing a wide range of diversity can require a relatively large number of accessions. The core is expected to include the maximum possible genetic diversity in the base population with practically no duplication of accessions. The overall objective of establishment of core collection is to get a collection conserved better and used more effectively, along with savings through economy of size.

The core set contains materials of highest priority for conservation and evaluation and is the suitable material for developing an adequate list of descriptors (Brown, 1995), while for breeders, it represents a logical first set of accessions to be searched for desirable alleles for finding donors of various characters for use in hybridization. Brown *et al.* (1987) developed the first core collection from the Australian collections of perennial *Glycine* spp. Since then a number of core collections have been established for different crops.

There is a large assemblage of wild germplasm of rubber (*Hevea brasiliensis*) collected during the 1981 International Rubber Research and Development Board (IRRDB) expedition to the center of origin in Brazil and distributed among the member countries. India is maintaining around 4500 of the above collection. The unmanageable size of the collection of this perennial tree crop often poses limitations for a quick and effective evaluation and identification of the desired genetic variability. At present the wild accessions are being evaluated for various agronomic characters in field evaluation trials, in small batches taken directly from the base gene pool. However, it will take considerable time and resources to evaluate the entire collection for secondary characters to identify genetically divergent accessions. On the other hand, if a core collection is established, it will serve as an easily accessible set of clones that represent the genetic variability available in the base collection.

MATERIALS AND METHODS

Data collected from a set of 80 wild accessions of *H. brasiliensis* from a preliminary field evaluation trial at the Central Experiment Station of Rubber Research Institute of India, Chethackal, Kerala, was used for this study. The material belonged to the 1981 IRRDB wild Germplasm collection and represented three provenances in Brazil *viz.*, Acre (25 accessions), Rondonia (27 accessions) and Mato Grosso (28 accessions). They were planted in a field trial employing a simple lattice design with four replications.

Factor analysis carried out in an earlier study in the same set of accessions (Abraham, 2001) using a set of 33 quantitative characters comprising growth, leaf and bark anatomical characters recorded in the third year of growth, identified 12 marker characters *viz.*, girth, number of leaves, length of the petiole, number of stomata and epidermal cells per unit area of leaf surface, thickness of lamina, number of cells per unit length of spongy layer, bark thickness, total number of latex vessel rows, height of phloic rays, diameter of latex vessels and density of latex vessels (Table 1). Apart from these quantitative characters, a set of 18 qualitative characters that described the morphological characteristics *viz.*, nature of nodes, shape and size of leaf storey and leaves along with other foliar characters in the same accessions using relevant descriptors (Abraham, 2001) was used for computing the standardized

Table 1. **Quantitative characters with their range and mean used for principal component analysis**

Character	Minimum (Accession)	Maximum (Accession)	Mean
Girth (cm)	4.72 cm (MT 929)	10.41 (RO 322)	7.38
Total number of leaves	41.2 (MT 929)	119.8 (RO 322)	69.8
Petiole length (cm)	13.03 (MT 1064)	32.47 (RO 886)	21.86
Number of stomata per mm^2 area	281.16 (MT 1028)	612.67 (MT 928)	433.87
Number of epidermal cells per mm^2 area	1132.84 (MT 1028)	2741.23 (RO 369)	1711.5
Thickness of leaf blade (µm)	0.1107 (RO 380)	0.176 (AC953)	0.1371
Number of cells /mm spongy layer	189.24 (AC 706)	417.06 (AC 650)	273.95
Total bark thickness (mm)	2 (RO 886)	4 (RO 395)	2.86
Total number of latex vessel rows	2.99 (MT 1031)	11.01 (RO 399)	5.81
Height of phloic rays (µm)	0.18 (MT 1024)	0.41 (AC 644)	0.29
Diameter of latex vessels (µm)	13.44 (MT 906)	34 (MT 899)	21.46
Density of latex vessels/row/mm	11.5 (RO 399)	25 (RO 894)	17.15

Shannon-Weaver diversity index (SDI) to assess the phenotypic diversity (Table 2). The method used in this study made use of the functional relationship between the cumulative contribution of the wild *Hevea* accessions to the total sum of squares (TSS) and the number of accessions (Balakrishnan *et al.*, 2000). The size of the core set is decided by a cut-off value when the rate of increase of the cumulative contribution of the accessions to TSS reaches a maximum value, even though the sampling is continued until the cumulative contribution to the TSS reaches a maximum of 100.

Statistical analysis

Principal component analysis was carried for the 12 quantitative characters. The principal component scores for all the 12 components were computed for each of the 80 accessions. The contribution of the i^{th} accession to the total variability (or Total Sum of Squares) of the system was computed based on Noirot *et al.* (1996) as

Table 2. **Descriptor states for each character and Shannon- Weaver Diversity Index (SDI) for the base population and the core set**

Qualitative trait	No of descriptor states	SDI values	
		Base population	Core set
Axillary buds	3	0.9900	0.9800
Leaf scars margin	2	0.9873	0.9807
Leaf storey shape	4	0.9973	0.9594
Leaf storey separation	3	0.9904	0.9884
Leaf storey appearance	2	0.9979	0.9810
Pulvinus	2	0.9918	0.9823
Petiole shape	4	0.8729	0.8543
Petiole size	3	0.9956	0.9907
Petiole angle	3	0.9907	0.9850
Petiole orientation	3	0.9724	0.9662
Leaflet color	2	0.9865	0.9811
Leaflet luster	2	0.9963	0.9916
Leaflet shape	3	0.9789	0.9610
Leaflet margin	3	0.9855	0.9812
Leaflet size	2	0.9936	0.9935
Leaflet cross sectional appearance	4	0.9464	0.9268
Leaf tip	4	0.5497	0.5197
Leaflet surface	3	0.9875	0.9577
Average		**0.9562**	**0.9434**

$$P_i = \Sigma y_{ij}^2$$

where, y_{ij} is the component score of the i^{th} accession for the j^{th} principal component and t is the number of principal components.

The relative contribution of the i^{th} accession was computed as

$CR_i = (P_i / \text{Total sum of squares}) * 100$

where, total sum of squares is equal to ΣP_i.

The constitution of the core collection was validated by comparing the phenotypic diversity in the two sets. To evaluate the phenotypic diversity in the wild accessions in terms of each of the 18 qualitative descripors, the Shannon-Weaver diversity index (SDI) was computed using the formula:

$$SDI = -\Sigma p_i \log_e (p_i)$$

where, *s* is the number of phenotypic classes or descriptor states for a given qualitative descriptor and p_i is the proportion of the total number of accessions in the i^{th} class (Shannon and Weaver, 1963; Jain *et al.*, 1975). The index was standardized to keep its value in the range 0 to 1, by dividing the value by $\log_e s$ (Li *et al.*, 1996).

To identify pairs of accessions which have a high degree of similarity and to cull out probable duplicates from the final core set, a coefficient of similarity V_{ij} (Gower metric) was computed for each pair of accessions *i* and *j* by the formula as suggested by Harch *et al.* (1996).

$$V_{ij} = 1 - \frac{1}{m} \sum_{p-1}^{m} \frac{|x_{ip} - x_{jp}|}{R_p}$$

where, *m* is the number of descriptors (which includes qualitative, binary and quantitative descriptors considered for the analysis), R_p stands for the range in case of quantitative descriptor or 1 for qualitative trait; x_{ip} and x_{jp} stand respectively for the values for the p^{th} descriptor for the i^{th} and j^{th} accessions. In the case of qualitative or binary descriptor, the quantity in parentheses equals 0 if the states are identical or 1 if they are not. The value of the similarity coefficient always lies between 0 and 1 and a value of 1 indicates an exact match or duplicate.

For further validation of the core collection, various measures for retention of diversity and tests of significance in the core collection were also carried out, as given below.

i. Measures of retention of diversity in the core collection

a. Retention(%) of Range = $\frac{1}{m}\sum_{n}\left[\frac{Range_{(CC)}}{Range_{(WC)}}\right]100$ (1)

b. Retention (%) of CV = $\frac{1}{m}\sum_{n}\left[\frac{CV_{(CC)}}{CV_{(WC)}}\right]100$ (2)

c. Retention (%) of SDI = $\frac{1}{m}\sum_{n}\left[\frac{SDI_{(CC)}}{SDI_{(WC)}}\right]100$ (3)

where, the subscripts (CC) and (WC) indicate the respective values in the core collection and the base collection respectively, and indicates the summation over m qualitative attributes. An SDI retention percent of more than 100 implies higher diversity in the core subset (ie, a better and evenly distributed frequency patterns of the attribute states) with respect to the descriptors under consideration.

ii. Tests of significance for differences in mean and variance of a quantitative trait in the core subset from the base collection values.

The deviation of the mean of the core subset from that of the base collection for each individual trait can be tested for statistical significance by using the standard normal Z-test. For testing the deviation of the core collection SD (s) from that of the base collection SD (ó) for each individual trait, a standard normal test can be carried out using the property that for large n (the core collection size), $\log_e(s^2)$ follows a normal distribution with mean $\log_e(ó^2)$ and variance (2/n-1).

RESULTS AND DISCUSSION

Many approaches in constructing core collections have been developed over the years. Several sampling methods to select entries for the core collections have been suggested ranging from random sampling (Brown, 1989; Charmet and Balfourier, 1995) to stratify sampling (Peeters and Martinelli, 1989; Yonezawa *et al.*, 1995; Balfourier *et al.*, 1999; Li *et al.*, 2004). Zewdi *et al.* (2004) have reported establishment of core collection in *Capsicum* germplasm by clustering and then by systematic selection of a number of accessions per cluster based on the presence of unique traits. Also, the criteria used to assess the degree of diversity in the core are most often based on phenotypic values such as mean, range, variance and Shannon-Weaver diversity index (Galwey, 1995). Alternative procedures such as the use of genetic markers (Hintum, 1994; Schoen and Brown, 1995) or the use of coefficient of parentage (Hintum and Haalman, 1994) have recently received much attention.

Core collections formed by simple random sampling, a procedure most widely reported in several crops, gives a satisfactory representation of the phenotypic variation of the original gene pool. However, this procedure has the disadvantage that it does not eliminate the possibility of duplication or doubling of entries and, yet, it is equally clear that any particular rare variant, found perhaps in only one entry of the whole collection, is likely to be absent from the core (Brown and Spillane, 1999).

The concept of core collection does not clearly define the size and content. Frankel and Brown (1984) and Brown (1989) had suggested a 10% size for the core collection in fairly large germplasm collections.

Principal component analysis was carried out on the 12 quantitative characters, by which the individual contribution of each of the 80

wild *H. brasiliensis* accessions to the total variability or TSS was estimated. Based on this, the accessions were sorted in the descending order with respect to their relative contributions. This procedure eliminates iterative selection and provides verification of the diversity in terms of SDI. It also ensures that the size of the core set depends on the diversity of the entire collection. The cumulative contribution to the TSS for the wild accessions was also worked out by adding the contribution of each of the accessions to the TSS. The cut-off value for inclusion of accessions in the core set was decided by the maximum curvature method (Fig. 1) as the

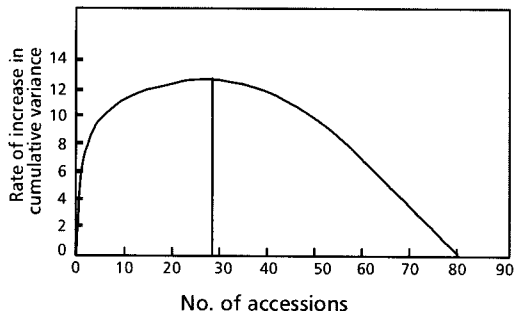

Fig.1. Rate of increase of cumulative contribution to the TSS by the accessions

number of accessions at the point when the rate of increase of the cumulative contribution to TSS reached a maximum value, which was 49.52 in this study. Accordingly 27 wild accessions with their cumulative contribution below 50.86 were selected in the preliminary subset (Fig. 1). These included 6 accessions from Acre, 8 from Rondonia and 13 from Mato Grosso.

The Gower metrices were worked out for all possible pairs of the 27 accessions in the preliminary core set, to identify the similar accessions and to eliminate one of them. Seven pairs of accessions recorded Gower metric values in the range of 0.8 to 0.9 (Table 3). From each pair of accessions thus identified as closely related, the accession that had contributed more to the TSS compared to its counterpart alone was retained in the preliminary core set.

Table 3. **Frequency distribution of the accession pairs for the Gower metric distances**

Gower metric distance class	Frequency of accession pairs
< 0.6	62
0.6 – 0.7	187
0.7 – 0.8	122
0.8 – 0.9	7

From the seven pairs identified as closely related accessions, *viz,* (MT 928, MT 1011), (MT 928, MT 1008), (MT 928, MT 931), (MT 1011, MT 1008), (MT 1011, MT 931), (MT 948, AC 627) and (AC 627, RO 330) four accessions MT 1011, MT 1008, MT 931 and AC 627 were removed from the core

Table 4. **Final list of wild accessions included in the core set**

Serial number	Accession	Serial number	Accession
1	RO 369	13	MT 906
2	MT 928	14	RO 255
3	AC 453	15	RO 322
4	MT 944	16	RO 859
5	MT 1005	17	RO 330
6	RO 399	18	MT 920
7	MT 1031	19	MT 1029
8	RO 317	20	RO 319
9	AC 1043	21	MT 947
10	AC 754	22	MT 929
11	AC 657	23	AC 644
12	MT 948		

set as possible duplicates as their relative contribution was less than that of the counterpart accession in the combination. Hence the preliminary core set was finally limited to 23 accessions in the final core set (Table 4), which constituted 28.75% of the base population.

The choice of proportion of accessions to be included in the core is arbitrary, usually in the 5 to 20% range, and will depend on the purpose of the core collection. Malosetti *et al.* (2000) established a core collection constituting 19.5% of the base population in a spring barley collection of 231 accessions on two considerations *viz.* (a) manageable number of accessions, and (b) more representative of the small base collection. Yonezawa *et al.* (1995), in determining the sampling strategies utilizing the genetic diversity component, suggested a 20-30% core collection size as the best. Hence the 28.75% core size derived in this study is justifiable as a relatively small collection of 4967 accessions of wild *H. brasiliensis* germplasm in India is thus expected to have a core size of approximately 1400 accessions which is manageable and representing the true diversity of the base collection with the minimum chances of losing valuable representatives.

Establishing a core collection for clonal crops has several advantages in terms of maintenance, evaluation and utilization. One important factor is that field gene banks are very expensive to run and are exposed to damage and loss (Morales *et al.* 1995). In the case of *H. brasiliensis*, the maintenance of a base collection of wild germplasm cannot be avoided, irrespective of attempts to establish a core collection, since established protocols for alternative methods of conservation by *in vitro* methods or cryopreservation are lacking. However, identifying a core collection for the wild *H. brasiliensis* gene pool will enable concentrated evaluation of the core set under field conditions, thus saving time and resources.

The diversity index of Shannon and Weaver (1963) used as a measure of phenotypic diversity for morphological descriptors has also been used by many other workers (Polignano *et al.*, 1999; Balakrishnan *et al.*, 2000). The final core set and the base collection were separately tested for phenotypic diversity for 18 qualitative characters based on SDI, in order to verify the representative nature of the core set identified in terms of diversity. Such a comparison of SDI values between the two sets of accessions will ensure that the size of the core set depended on the diversity of the entire collection. It was found that the SDI values in both the sets were almost identical for all the 18 characters with an average SDI of 0.9562 for the base population and 0.9434 for the new core set established (Table 2). Close correspondence of SDI values between core collections and their base population was also reported by Balakrishnan *et al.* (2000) in sugarcane germplasm using the same methodology where the SDI values for the base collection and the core set established was 0.7400 and 0.7500 respectively. Estimates of SDI for descriptors across eight geographical regions in the whole and core subsets were reported for faba bean germplasm by Polignano *et al.* (2001) where the indices for the core subsets were found to be very close to that of the base population, with the mean diversity index ranging from 0.92 to 0.96 and from 0.90 to 0.92, respectively.

Table 5. **The retention percentages of range, CV and SDI**

	Retention percentage
Range	78.97
CV	99.07
SDI	98.56

Other validation procedures were also adopted in this study to confirm the representative nature of the core set identified. The degree of retention of the range, CV and SDI of the base population in the core set was computed. The core was successful in retaining 79.97% of the range of the base population, while the retention percentages of CV and SDI were close to 100 (Table 5), implying that the diversity of the core set and the base collection were same. In the validation study using computed Z values for each of the quantitative characters (Table 6), all Z values were found to be non-significant indicating no significant differences between the means of the base collection and the core for each of the quantitative trait. All the validation procedures carried out clearly indicate that the core set is truly representative of the base population, and hence the method is highly useful in the establishment of a core collection in wild *Hevea* germplasm. Accordingly, the 23 wild accessions identified were selected for inclusion in the final core set collection.

Table 6. **Z values computed for each of the quantitative characters studied**

Character	Z test
Girth (cm)	1.05 NS
Total number of leaves	0.87 NS
Petiole length (cm)	-1.10 NS
Number of stomata per mm^2 area	0.77 NS
Number of epidermal cells per mm^2 area	-1.39 NS
Thickness of leaf blade (µm)	0.58 NS
Number of cells /mm spongy layer	0.12 NS
Total bark thickness (mm)	-0.52 NS
Total number of latex vessel rows	0.66 NS
Height of phloic rays (µm)	0.80 NS
Diameter of latex vessels (µm)	0.01 NS
Density of latex vessels/row/mm	-0.37 NS

ACKNOWLEDGEMENT

Authors acknowledge Dr. N.M. Mathew, Director, Rubber Research Institute of India, for his valuable support and encouragement. We also thank Dr. N.Vijayan Nair, Director, Sugarcane Breeding Institute, Coimbatore, Tamil Nadu, India for his advice and support given during the preparation of this manuscript.

REFERENCES

Abraham, S.T. (2001). *Genetic parameters and divergence in certain wild genotypes of Hevea brasiliensis* (Willd.ex Adr.de Juss.) Muell. Arg. Ph.D. Thesis, Mahatma Gandhi University, Kerala, India, 203p.

Balakrishnan, R., Nair, N.V. and Sreenivasan, T. V. (2000). A method for establishing a core collection of *Saccharum officianarum* L. germplasm based on quantitative- morphological data. *Genetic Resources and Crop Evolution*, 47 (1): 1-9.

Balfourier, F., Prosperi, J. M., Charmet, G., Goulard, M. and Monestiez, P. (1999). Using spatial patterns of diversity to develop core collections. In: *Core Collections for Today and Tomorrow* (Eds. R.C. Johnson and T. Hodgkin). International Plant Genetic Resources Institute, Italy, pp. 37-48.

Brown, A.H.D., Grace, J.P. and Speer, S.S. (1987). Designation of a "core" collection of perennial *Glycine*. *Soybean Genetics Newsletter*, 14: 59-70.

Brown, A.H.D. (1989). Core collections: a practical approach to genetic resources management. *Genome*, 31(2): 818-824.

Brown, A.H.D. (1995). The core collection at the cross roads. In: *Core Collections of Plant Genetic Resources*. (Eds. T. Hodgkin, A.H.D. Brown, T.J.L. Hintum and E.A.V. Morales). International Plant Genetic

Resources Institute, John Wiley & Sons, Chichester, UK, pp. 3-19.

Brown, A.H.D. and Spillane, C. (1999). Implementing core collections- principles, procedures, progress, problems and promise. In: *Core Collections for Today and Tomorrow* (Eds. R.C. Johnson and T. Hodgkin). International Plant Genetic Resources Institute, Rome, Italy, pp.1-10.

Charmet, G. and Balfourier, F. (1995). The use of geostatistics for sampling a core collection of perennial ryegrass populations. *Genetic Resources and Crop Evolution,* **42** (4): 303-309.

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

Frankel, O.H. and Brown, A.H.D. (1984). Plant genetic resources today: A critical appraisal. In: *Crop Genetic Resources: Conservation and Evaluation.* (Eds. J.H.W. Holden and J.T. Williams). George Allen and Unwin, London, UK, pp. 249-257.

Galwey, N. W. (1995). Verifying and validating the representativeness of a core collection. In: *Core Collections of Plant Genetic Resources.* (Eds. T. Hodgkin, A.H.D. Brown, T.J.L. Hintum and E.A.V. Morales). International Plant Genetic Resources Institute, John Wiley & Sons, Chichester, UK, pp. 187-198.

Harch, B.D., Basford, K.E., DeLacy, I.H., Lawrence, P.K. and Cruickshank, A. (1996). Mixed data types and the use of pattern analysis on the Australian groundnut germplasm data. *Genetic Resources and Crop Evolution,* **43**(4): 363-376.

Hintum, T.J.L.V. (1994). Comparison of marker systems and construction of a core collection in a pedigree of European spring barley. *Theoretical and Applied. Genetics,* **89**(7-8): 991-997.

Hintum, T.J.L.V. and Haalman, D. (1994). Pedigree analysis for composing a core collection of modern cultivars, with examples from barley (*Hordeum vulgare* s. lat.). *Theoretical and Applied Genetics,* **88**(1):70-74.

Jain, S.K., Qualset, C.O., Bhat, G.M. and Wu, K.K. (1975). Geographical patterns of phenotypic diversity in a world collection of durum wheat. *Crop Science,* **15**(5): 700-704.

Li, Y., Cao, Y.S. and Zhung, X. (1996). A phenotype diversity analysis of foxtail millet (*Seturia italica* L.P.Beauv.) landraces of Chinese origin. *Genetic Resource and Crop Evolution,* **43**(4): 377-384.

Li, Y., Shi, Y., Cao, Y. and Wang, T. (2004). Establishment of a core collection for maize germplasm preserved in Chinese National Gene bank using geographic distribution and characterization data. *Genetic Resource and Crop Evolution,* **51**(8): 845-852.

Malosetti, M., Abadie, T. and German, S. (2000). Comparing strategies for selecting a core subset for the Uruguayan barley collection. *Plant Genetic Resource Newsletter,* **121**: 20-26.

Morales, E.A.V., Valois, A.C.C and Costa, I.R.C. (1995). Core collection for gene banks with limited resources. In: *Core Collections of Plant Genetic Resources.* (Eds. T. Hodgkin, A.H.D. Brown, T.J.L. Hintum and E.A.V. Morales). International Plant Genetic Resources Institute, John Wiley & Sons, Chichester, UK, pp. 241-250.

Noirot, M., Hamon, S. and Anthony, F. (1996). The principal component scoring: a new method of constituting a core collection using quantitative data. *Genetic Resource and Crop Evolution,* **43**(1):1-6.

Peeters, J. P. and Martinelli, J.A. (1989). Hierarchical cluster analysis as a tool to manage variation in germplasm collection. *Theoretical and Applied Genetics,* **78**(1):42-48.

Polignano, G.B., Alba, E., Uggenti, P. and Scippa, G. (1999). Geographical patterns of variation in Bari faba bean germplasm collection. *Genetics Resource and Crop Evolution,* **46**(2):183-192.

Polignano, G.B., Uggenti, P. and Scippa, G. (2001). Diversity analysis and core collection formation in Bari faba bean germplasm. *Plant Genetic Resource Newsletter,* **125**: 33-38.

Schoen, D. J. and Brown, A. H. D. (1995). Maximizing genetic diversity in core collections of wild relatives of crop species. In: *Core Collections of Plant Genetic Resources.* (Eds. T.Hodgkin, A.H.D. Brown, T.J.L. Hintum and E.A.V. Morales). International Plant Genetic Resources Institute, John Wiley & Sons, Chichester, UK, pp. 55-77.

Shannon, C. E. and Weaver, W. (1963). *The Mathematical Theory of Communication.* University of Illinois Press, Urbana, IL, USA, 144 p.

Yonezawa, K., Nomura, T. and Morishima, H. (1995). Sampling strategies for use in stratified germplasm collections. In: *Core collections of Plant Genetic Resources* (Eds. T. Hodgkin, A.H.D. Brown, T.J.L. Hintum and E.A.V. Morales). International Plant Genetic Resources Institute, John Wiley & Sons, Chichester, UK, pp. 35-53.

Zewdie, Y., Tong, N. and Bosland, P. (2004). Establishing a core collection of *Capsicum* using a cluster analysis with enlightened selection of accessions. *Genetic Resource and Crop Evolution,* **51**(2): 147-151.