

Development of the core collection through advanced maximization strategy with heuristic approach in cashew (*Anacardium occidentale* L.)

G. S. Mohana* and M. G. Nayak

ICAR-Directorate of Cashew Research, Puttur, Dakshina Kannada district, Karnataka State, India

Received 17 October 2017; Accepted 26 January 2018 – First published online 6 March 2018

Abstract

ICAR-Directorate of Cashew Research is the nodal agency for conducting cashew research and manages the largest field gene bank in India. Cashew is a perennial tree and needs more land and other resources to maintain accessions. Conservation through seeds is not feasible because of cross-pollination. Tissue culture efforts to regenerate plants from mature explants are not successful. Therefore, efficient management of the filed gene bank particularly utilization requires designation of the core collection representing the spectrum of diversity present in the entire collection. In this study, a relatively new technique, the advanced M strategy with heuristic approach was deployed to develop the core collection. Sixty-eight morphometric characters of 478 accessions were subjected to analysis resulting in the core collection of 49 accessions. Further, another core collection of same number was constituted by K-Means clustering to compare the efficiency of two approaches. The validation parameters like mean difference, variance difference, coincidence rate, variable rate and class coverage among others were employed for comparative analysis. The results of these parameters revealed that the core collection designated by heuristic approach was better able to efficiently represent and retain the diversity of the entire collection compared with the core identified by clustering approach. Future conservation and breeding efforts will be focused on establishing a separate block in the field gene bank having 49 accessions of cashew core collection.

Keywords: cashew, core collection, germplasm, heuristic approach

Introduction

Cashew is an economically important crop introduced to India by Portuguese from Brazil during 16th century. ICAR-Directorate of Cashew Research was established in 1986 at Puttur, Dakshina Kannada district of Karnataka state, India. This is the nodal centre for cashew research in the country and is conducting systematic surveys since its inception for germplasm collections across India. So far, 539 germplasm accessions have been conserved in the National Cashew Field Gene Bank (NCFGB), and out of these, 478 accessions have been evaluated and

characterized (Nayak *et al.*, 2015). The success of crop improvement programme in any crop pivots around utility of germplasm accessions. Hence detailed evaluation and subsequent conservation of accessions is of paramount importance. Since cashew is a perennial tree crop, it needs considerable resources for management and conservation. However, conservation through seeds is not possible in cashew as it is highly cross-pollinated and each seed will be genetically dissimilar from its mother. Further, as in other *Anacardiaceae* members such as mango, regeneration of plants by tissue culture of mature explants has not been successful in cashew (Thimmappaiah and Shirly, 1999; Krishna and Singh, 2007). Hence, any effort to minimize the number of accessions to be conserved and hence space is more rewarding for saving resources.

*Corresponding author. E-mail: mohangs2007@gmail.com

Considerable redundancy in germplasm accessions of most crops including cashew has been observed in field gene banks. Therefore, it is crucial to concentrate on the set of very significant collection that maximizes allelic richness out of a big entire collection. This representative collection is called as core collection (Frankel, 1984, Brown, 1989), which can be easily managed, utilized and conserved. Since it represents the diversity present in the entire collection, the core collection serves as a reference for adding new germplasm accessions with unique characters. In other words, it helps in identifying gaps and enrich the existing collections. Moreover, crop improvement efforts need diverse germplasm accessions to enhance genetic base of existing cultivars and to develop novel populations, which eventually leads to identification of better varieties and hybrids. So, the core collections are expected to ensure continued genetic gain in breeding efforts. Further, specifically in cashew, cashew stem and root borer is a serious pest, and every year, there is a chance of losing some trees/accessions due to infestation of this pest in spite of best remedial measures. Hence, it is prudent to have representative collection in an additional conservation block.

Core collection – earlier efforts

Different methodologies have been deployed for making core collection in many crops. Initially, stratified random sampling (Peeters and Martinelli, 1989; Chandra *et al.*, 2002; Franco *et al.*, 2003) and genetic distance sampling (Jansen and van Hintum, 2007) were most widely used methods. However, these methods have led to inequality in diversity estimates by skewing the quantitative trait variances in the core collections. In order to address this problem, approach like maximization strategy (Bataillon *et al.*, 1996; Marita *et al.*, 2002; Schoen and Brown, 1993, which selects specific combinations of accessions to ensure retention of maximum diversity, had been developed. Further, MSTRAT programme (Gouesnard and Bataillon, 2001) that uses the criterion of richness and highest sum of squares of active variables based on maximization strategy to select subsets has been found powerful in making core collection. However, in recent years, heuristic approach with advanced maximization strategy (Kim *et al.*, 2007) is gaining significance as it develops core collection representing all observation classes with least allelic redundancy ensuring a highly reproducible list of entries. This method has been successfully deployed in crops like rice (Chung *et al.*, 2009), mung bean (Moe *et al.*, 2012), dolichos bean (Vaijayanthi *et al.*, 2015a, b), foxtail millet (Gowda *et al.*, 2013) and soybean (Gireesh *et al.*, 2015). Keeping this in view, in the present study, the heuristic approach with advanced maximization strategy was followed for making core collection in cashew.

Material and methods

The gene bank at the ICAR-Directorate of Cashew Research houses the largest germplasm repository of the country. In the NCFGB, 539 clonal accessions are collected from different parts of India including Andaman and Nicobar Islands and other countries during 1986–2016. The characterization and evaluation of 478 accessions (up to 2003 planting) with six softwood grafts per accession at spacing of 6 m × 6 m was done after six annual harvests by referring ‘Cashew Descriptors’ published by IBPGR (presently the Bioversity International) (IBPGR, 1986). A total of 68 characters, both quantitative (27) and qualitative characters (41), were recorded on these accessions. Colour of mature cashew apple was recorded by referring RHS colour chart (Anonymous, 1995). The evaluated accessions are maintained as active collections in the field gene bank with four softwood grafts per accession at spacing of 4 m × 4 m under rainfed conditions by adopting recommended package of practices. The characterization data of these accessions were compiled in ‘Catalogue of Minimum Descriptors of Cashew (*Anacardium occidentale* L.) Germplasm Accessions-I, II and III’ and were published (Swamy *et al.*, 1997; 1998, 2000). The Catalogues-IV and V were published during the years 2014 and 2015 (Nayak *et al.*, 2014, 2015).

The data on 478 accessions with 68 characters were subjected to heuristic analysis with advanced maximization strategy using PowerCore v.1.0 software (Kim *et al.*, 2007) developed at the Genetic Resource Division, National Institute of Agricultural Biotechnology, Rural Development Administration, Republic of Korea. This software basically classifies continuous variables into different categories based on Sturges rule ($K=1+\log_2 n$ where n = number of observed accessions, Sturges, 1926). Upon classification, the software considers all classes for selection of representative accessions from each class with least redundancy and the list of entries is highly reproducible. Using this, a core collection of 49 accessions (10.25% of the accessions) was generated which will be referred to as power core.

To compare the efficiency of heuristic approach with other methods of establishing core collection, *K*-Means clustering was also done using IBM-SPSS Statistics V. 20 to generate 10 clusters. The number of clusters in this method was arrived after testing the criterion of minimum variance within cluster and maximum variance between clusters by analysis of variance, and the variance of at least 50% of the characters is significantly different across the clusters. From each cluster, 10% of the accessions proportional to cluster size (22, 98, 4, 29, 67, 87, 20, 5, 76, 70 accessions, respectively, in 1–10 clusters) were randomly selected and pooled to form the core collection. Through this method also, there were 49 accessions forming the core and this core will be

referred to as cluster core. In fact, Brown (1995) suggested 10% sample size from each cluster so that 70% of the extant alleles could be drawn with 95% certainty.

In order to validate the efficiency of the core collection in representing the entire collection, parameters like mean difference (MD%), variance difference (VD%), coincidence rate (CR%), variable rate (VR%) and class coverage (CC%) for quantitative characters (formula of each given below) and Shannon–Weaver diversity and Nei's diversity indices generated by Powercore software for qualitative characters were considered (Shannon and Weaver, 1949; Nei, 1978).

$$\text{Mean difference} - \text{MD} (\%) = \frac{1}{m} \sum_{j=1}^m \frac{M_e - M_c}{M_c} \times 100$$

(M_e = mean of each trait in the entire collection; M_c = mean of each trait in the core collection; m = number of the variables).

$$\text{Variance difference} - \text{VD} (\%) = \frac{1}{m} \sum_{j=1}^m \frac{V_e - V_c}{V_c} \times 100$$

(V_e = variance of each trait in the entire collection; V_c = variance of each trait in the core collection).

$$\text{Coincidence rate} - \text{CR} (\%) = \frac{1}{m} \sum_{j=1}^m \frac{R_c}{R_e} \times 100$$

(R_c = range of each trait for the entire collection; R_e = range of each trait for the core collection).

$$\text{Variable rate} - \text{VR} (\%) = \frac{1}{m} \sum_{j=1}^m \frac{CV_c}{CV_e} \times 100$$

(CV_c = coefficient of variation of trait for the core collection and CV_e = coefficient of variation of trait for the entire collection).

$$\text{Class coverage} (\%) = \frac{1}{m} \sum_{j=1}^m \frac{D_c}{D_e} \times 100$$

(D_c = number of classes occupied in the core collection and D_e = number of classes occupied in the entire accession in each character).

To confirm the representativeness of the core collection, it is prescribed that MD% should be smaller than 20% and CR% should be more than 80%. Larger VD% and VR% should also be there to reflect the representativeness of the core collection with respect to the entire collection (Hu *et al.*, 2000; Kim *et al.*, 2007). Further, the CC% based on qualitative traits in the core collection with reference to the entire collection should be more than 80%.

The χ^2 test is used as an alternative to see whether qualitative traits are adequately represented in the core. Here the expected number of accessions is derived with reference to the core size and the number of accessions in different categories of traits in the entire collection. Principal component analysis (PCA) deploying 27 quantitative characters

was also carried out to assess the segregation patterns of the entire, power core and cluster core collections. Further, variance, mean and frequency distribution of quantitative traits in both the entire and core collections were arrived using Descriptive Statistics option of IBM SPSS Statistics version 20. The significance of variance of different characters between the entire and core collections was tested using *F*-test. Similarly, the significance of mean was tested deploying two-sample *t*-test assuming unequal variances. In addition, skewness, kurtosis and coefficient of variation of characters were derived using descriptive statistics option of Microsoft Excel.

Results

(a) Validation parameters in the power core and cluster core collections

The results revealed that the MD (%) between the entire and core collections was 4.94 in the power core and -0.87 in the cluster core (Table 1). It was observed that the power core has more CR% (95.37) compared with the cluster core (69.09). The VD% and VR% values of the power core were 41.01 and 129.54, and cluster core were 6.16 and 97.35, respectively. Further, accessions from respective classes of all quantitative traits are represented in the power core collection. This is indicated by the CC% of quantitative characters and it was 100% in the power core collections with reference to the entire collection. As far as qualitative characters are concerned, in the cluster core, only 29 out of 41 characters retained the categories similar to that of the entire collection and the CC% worked out to be 88%. However, in the power core, the accessions could be grouped into all character categories similar to that of the entire collection and the CC% worked out to be 100%.

(b) Pattern of segregation in the core collections

In the PCA graphs (Fig. 1), spatial representation of the power core accessions showed better segregation and the first two axes accounted for 43.73% variability. Whereas, the

Table 1. Validation parameters

Particulars	Power core	Cluster core
Number of accessions	49	49
Variables	68	68
MD%	4.94	-0.87
VD%	41.01	6.16
CR%	95.37	69.09
VR%	129.54	97.35

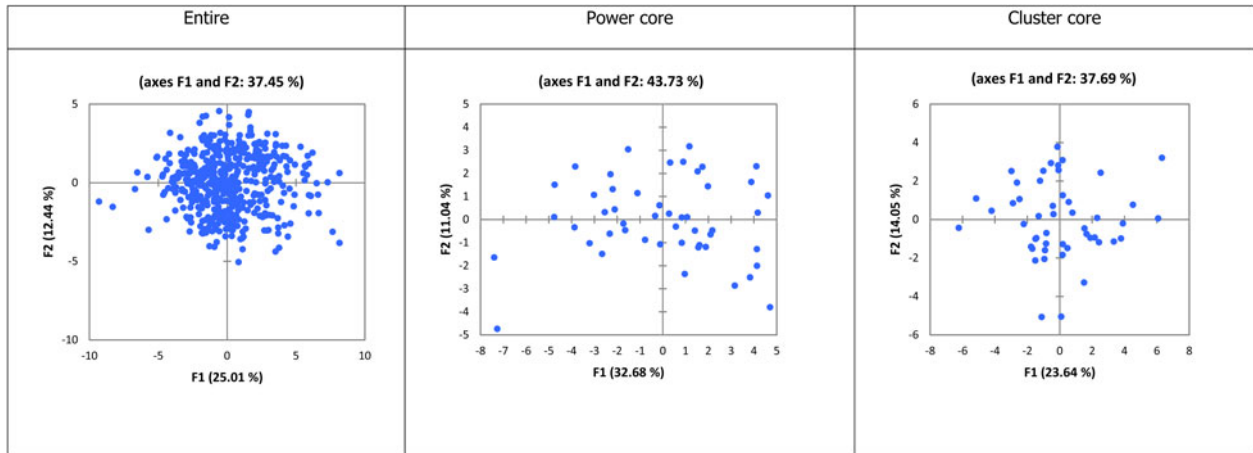


Fig. 1. PCA graphs depicting spread of the entire and core collections.

segregation of accessions in the cluster core is weaker compared with the power core and the first two axes accounted for 37.69% variability. In case of the entire collection, only 37.45% of the variability was accounted by first two axes.

(c) Variance, mean and frequency distribution of traits in the entire and core collections

Out of 27 quantitative characters, the variances of 22 characters in the power core collections were significantly different from respective variances in the entire collection (Table 2). However, when the cluster core and entire collection variances were compared, only one was found to be significant. Whereas, between the power core and cluster core collections, the variances of 18 characters were significant.

The mean values of 27 quantitative characters between the power core and entire collections did not significantly differ except for the character sex ratio (Table 3). At the same time, the mean values of these quantitative characters also did not differ significantly between the cluster core and entire collections. However, when the mean values of these characters between the power core and cluster core were compared, significant difference was found for sex ratio and flowering duration, the mean values of both characters being higher in the power core collection. For qualitative characters, higher or equal Shannon–Weaver and Nei's diversity values were observed for 39 out of 41 qualitative characters in the power core collection compared with the entire collection. When the power core and cluster core were compared, 38 in Shannon–Weaver and 36 in Nei's diversity indices of qualitative characters in the power core collections were higher or equal to the cluster core collections (Table 4).

The *t*-test revealed that means of Shannon–Weaver and Nei's diversity indices of the entire, power core and cluster

core collections for these qualitative characters were not significantly different from each other. The results of χ^2 test for observed and expected frequencies of accessions for traits revealed that it was significant for only four characters (10.81%) in the power core collection and non-significant for all the characters in the cluster core collections.

Frequency distribution patterns of quantitative traits (online Supplementary Table S5 and Fig. 2) showed that skewness values did not differ for 17 out of 27 characters in the power core and cluster core collections when compared with the values of entire collection. The kurtosis values for 23 characters were unaltered in the power core and the values for 11 characters were unaltered in the cluster core when compared with the kurtosis values of characters in the entire collection.

(d) Geographic representation of the entire and core collections

The geographic origin of the entire and core collections is depicted in online Supplementary Table S6. In the entire collection, maximum number of accessions were from Karnataka (24.69%) followed by Andhra Pradesh (21.34%) and Kerala (14.44%). Whereas in the power core, it is in the order of Kerala and Maharashtra (Both 20.41%), Karnataka (16.33%) and Andhra Pradesh (14.29%), and in the cluster core, it is Andhra Pradesh (32.65%) followed by Karnataka (22.45%) and Kerala (14.29%). However, West Bengal collections that are represented in the power core are not represented in the cluster core collection.

Discussion

The MD (%) between the entire and core collections in both the power core and cluster core were less than the

Table 2. Variance of the entire and core collections

Characters	Power core	Entire collection	'F' statistic	'F' critical	Cluster core	Entire collection	'F' statistic	'F' critical	Power core	Cluster core	'F' statistic	'F' critical
ILT	0.25	0.15	1.74	1.38	0.15	0.15	1.04	1.38	0.25	0.15	1.67	1.62
TRH	2.44	2.05	1.19	1.38	1.58	2.05	1.29	1.47	2.44	1.58	1.54	1.62
TS	4.00	2.46	1.62	1.38	2.13	2.46	1.15	1.47	4.00	2.13	1.88	1.62
EGT	36.10	26.16	1.38	1.38	17.63	26.16	1.48	1.47	36.10	17.63	2.05	1.62
TD	1.56	1.20	1.29	1.38	1.43	1.20	1.19	1.38	1.56	1.43	1.09	1.62
NLT	5.74	5.15	1.12	1.38	6.39	5.15	1.24	1.38	5.74	6.39	1.11	1.62
LS	959.32	590.60	1.62	1.38	678.19	590.60	1.15	1.38	959.32	678.19	1.41	1.62
NW	10.89	4.37	2.49	1.38	3.80	4.37	1.15	1.47	10.89	3.80	2.87	1.62
ISL	25.89	8.93	2.90	1.38	6.80	8.93	1.31	1.47	25.89	6.80	3.81	1.62
ISW	25.44	16.96	1.50	1.38	13.78	16.96	1.23	1.47	25.44	13.78	1.85	1.62
SR	0.00	0.00	1.44	1.38	0.00	0.00	1.14	1.47	0.00	0.00	1.64	1.62
SAL	2.58	1.37	1.89	1.38	1.38	1.37	1.01	1.38	2.58	1.38	1.87	1.62
SAW	1.40	0.57	2.46	1.38	0.45	0.57	1.25	1.47	1.40	0.45	3.09	1.62
AW	1485.78	546.36	2.72	1.38	605.20	546.36	1.11	1.38	1485.78	605.20	2.46	1.62
ST	0.50	0.27	1.90	1.38	0.24	0.27	1.09	1.47	0.50	0.24	2.06	1.62
NL	0.58	0.21	2.82	1.38	0.18	0.21	1.17	1.47	0.58	0.18	3.31	1.62
NDW	0.19	0.10	1.93	1.38	0.11	0.10	1.10	1.38	0.19	0.11	1.75	1.62
NT	0.15	0.07	2.21	1.38	0.06	0.07	1.14	1.47	0.15	0.06	2.53	1.62
FD	343.97	325.21	1.06	1.38	250.56	325.21	1.30	1.47	343.97	250.56	1.37	1.62
FI	397.39	257.95	1.54	1.38	324.46	257.95	1.26	1.38	397.39	324.46	1.22	1.62
ANR	18.43	9.83	1.87	1.38	10.52	9.83	1.07	1.38	18.43	10.52	1.75	1.62
SP	26.83	18.64	1.44	1.38	16.46	18.64	1.13	1.47	26.83	16.46	1.63	1.62
KWT	0.68	0.30	2.25	1.38	0.30	0.30	1.01	1.47	0.68	0.30	2.27	1.62
KL	0.19	0.10	1.93	1.38	0.12	0.10	1.26	1.38	0.19	0.12	1.54	1.62
KW	0.19	0.12	1.54	1.38	0.12	0.12	1.00	1.47	0.19	0.12	1.54	1.62
KT	0.05	0.02	1.85	1.38	0.03	0.02	1.25	1.38	0.05	0.03	1.48	1.62
CY	37.00	26.67	1.39	1.38	22.14	26.67	1.20	1.47	37.00	22.14	1.67	1.62

Table 3. Mean of the entire and core collections

	Entire colleciton	Power core	't' statistic	't' critical	Entire collection	Cluster core	't' statistic	't' critical	Power core	Cluster core	't' statistic	't' critical
ILT	1.411	1.419	-0.120	2.005	1.411	1.492	-1.393	2.002	1.419	1.492	-0.797	1.987
TRH	5.001	4.848	0.656	2.002	5.001	5.060	-0.309	2.000	4.848	5.060	-0.739	1.986
TS	6.200	5.948	0.857	2.005	6.200	6.364	-0.745	2.000	5.948	6.364	-1.179	1.987
EGT	14.482	14.532	-0.056	2.004	14.482	15.500	-1.581	1.998	14.532	15.500	-0.925	1.988
TD	5.676	5.877	-1.088	2.003	5.676	5.634	0.239	2.002	5.877	5.634	0.987	1.985
NLT	10.280	10.844	-1.576	2.002	10.280	10.888	-1.618	2.003	10.844	10.888	-0.089	1.985
LS	87.334	95.606	-1.813	2.005	87.334	86.705	0.162	2.002	95.606	86.705	1.540	1.986
NW	6.876	7.583	-1.469	2.007	6.876	6.733	0.486	2.000	7.583	6.733	1.552	1.991
ISL	16.145	16.858	-0.963	2.008	16.145	16.742	-1.504	1.999	16.858	16.742	0.141	1.993
ISW	19.331	20.089	-1.018	2.004	19.331	19.733	-0.715	2.000	20.089	19.733	0.398	1.987
SR	0.092	0.114	-2.656	2.004	0.092	0.093	-0.234	2.000	0.114	0.093	2.014	1.986
SAL	5.820	6.274	-1.923	2.006	5.820	5.825	-0.025	2.002	6.274	5.825	1.577	1.987
SAW	4.339	4.403	-0.366	2.007	4.339	4.365	-0.253	2.000	4.403	4.365	0.192	1.992
AW	61.817	72.967	-1.988	2.007	61.817	61.500	0.086	2.002	72.967	61.500	1.755	1.989
ST	3.064	3.143	-0.760	2.006	3.064	3.066	-0.032	2.001	3.143	3.066	0.621	1.988
NL	3.305	3.348	-0.386	2.007	3.305	3.294	0.185	2.000	3.348	3.294	0.439	1.992
NDW	2.418	2.466	-0.746	2.006	2.418	2.431	-0.251	2.002	2.466	2.431	0.453	1.987
NT	1.749	1.745	0.059	2.006	1.749	1.752	-0.094	2.000	1.745	1.752	-0.104	1.990
FD	82.538	87.588	-1.820	2.002	82.538	79.837	1.122	1.999	87.588	79.837	2.225	1.986
FI	65.778	60.861	1.672	2.004	65.778	64.372	0.525	2.003	60.861	64.372	-0.915	1.985
ANR	9.321	9.831	-0.810	2.006	9.321	9.142	0.369	2.002	9.831	9.142	0.897	1.987
SP	28.499	27.409	1.424	2.004	28.499	28.856	-0.582	2.000	27.409	28.856	-1.539	1.986
KWT	1.938	2.045	-0.890	2.007	1.938	1.940	-0.031	2.002	2.045	1.940	0.740	1.989
KL	2.585	2.670	-1.332	2.006	2.585	2.614	-0.568	2.003	2.670	2.614	0.694	1.986
KW	1.458	1.513	-0.860	2.004	1.458	1.475	-0.329	2.002	1.513	1.475	0.475	1.986
KT	1.190	1.174	0.495	2.006	1.190	1.210	-0.776	2.003	1.174	1.210	-0.905	1.986
CY	10.496	10.036	0.510	2.004	10.496	10.512	-0.022	2.000	10.036	10.512	-0.433	1.987

Table 4. Diversity indices for the entire and core collections

Characters	Shanon–Weiver index			Nei's diversity index		
	Power core	Entire	Cluster core	Power core	Entire	Cluster core
TH	0.79	0.65	0.69	0.45	0.35	0.37
LS	0.91	0.82	0.81	0.56	0.51	0.51
CTB	0.00	0.00	0.00	0.00	0.00	0.00
CAM	0.00	0.00	0.00	0.00	0.00	0.00
EPB	0.51	0.49	0.51	0.33	0.31	0.33
BP	0.51	0.53	0.51	0.33	0.35	0.33
CYL	1.07	0.86	0.95	0.60	0.50	0.58
CML	0.63	0.37	0.38	0.32	0.18	0.19
OL	0.00	0.00	0.00	0.00	0.00	0.00
LM	0.17	0.22	0.10	0.08	0.11	0.04
LAS	0.93	0.82	0.79	0.57	0.49	0.47
BL	0.20	0.04	0.00	0.08	0.01	0.00
ALP	0.41	0.36	0.45	0.25	0.21	0.27
LCS	1.14	1.16	1.13	0.62	0.65	0.63
SF	1.04	0.87	0.92	0.63	0.52	0.56
IS	0.65	0.54	0.63	0.40	0.30	0.38
FC	0.45	0.36	0.23	0.27	0.20	0.12
MAC	1.02	0.93	0.88	0.62	0.56	0.54
CAS	1.15	1.09	1.02	0.65	0.61	0.54
CMN	0.20	0.06	0.00	0.08	0.02	0.00
NS	0.10	0.03	0.00	0.04	0.01	0.00
CBL	0.10	0.02	0.10	0.04	0.00	0.04
CI	0.28	0.18	0.00	0.15	0.08	0.00
TIB	0.00	0.00	0.00	0.00	0.00	0.00
SFL	0.28	0.23	0.17	0.15	0.11	0.08
AB	1.24	1.29	1.37	0.68	0.70	0.74
RA	0.80	0.73	0.77	0.50	0.48	0.51
AA	0.68	0.61	0.56	0.48	0.42	0.37
GA	0.61	0.49	0.45	0.34	0.28	0.27
CA	0.80	0.67	0.65	0.50	0.40	0.40
SA	0.10	0.09	0.10	0.04	0.03	0.04
NA	1.04	1.07	1.04	0.63	0.65	0.63
NB	1.16	1.02	1.03	0.63	0.55	0.57
SN	0.75	0.69	0.68	0.49	0.48	0.48
FN	0.92	0.83	0.81	0.56	0.50	0.47
SSN	0.71	0.55	0.51	0.45	0.35	0.33
SNA	1.31	1.07	0.82	0.69	0.60	0.52
PSA	1.00	0.91	0.83	0.60	0.56	0.53
SUT	0.63	0.60	0.65	0.44	0.41	0.45
PK	1.00	0.84	0.64	0.56	0.45	0.35
CG	0.77	0.68	0.66	0.51	0.48	0.47
Mean	0.63	0.55	0.53	0.37	0.33	0.32

Index for quantitative and qualitative characters

Quantitative characters		Qualitative characters			
ILT	Internode length of twig (cm)	TH	Tree habit	AA	Cashew apple apex
TRH	Tree height (m)	LS	Leaf shape	GA	Groves on cashew apple apex
TS	Tree spread (m)	CTB	Cracks on trunk bark	CA	Cavity at the cashew apple apex
EGT	Extension growth of twigs (cm)	CAM	Crotch angle of main branches	SA	Skin of cashew apple
TD	Twig diameter (mm)	EPB	Ease of peeling bark from twigs	NA	Attachment of nut to apple
NLT	Number of leaves per twig	BP	Branching pattern	NB	Shape of nut base
LS	Leaf size (cm ²)	CYL	Colour of young leaves	SN	Suture of nut
NW	Nut weight (gm)	CML	Colour of mature leaves	FN	Flanks of nut
ISL	Inflorescence size – length (cm)	OL	Odour of leaves	SSN	Stylar scar on nut
ISW	Inflorescence size – width (cm)	LM	Leaf margin	SNA	Shape of nut apex
SR	Sex ratio	LAS	Leaf apex shape		
SAL	Size of cashew apple – length (cm)	BL	Brittleness of leaf		
SAW	Size of cashew apple – width (cm)	ALP	Angle of leaf petiole		
AW	Cashew apple weight (g)	LCS	Leaf cross-section		
ST	Shell thickness (mm)	SF	Season of flowering		
NL	Cashew nut dimension – length (cm)	IS	Inflorescence shape		
NW	Cashew nut dimension – width (cm)	FC	Flower colour		
NT	Cashew nut dimension – thickness (cm)	MAC	Mature cashew apple colour		
FD	Flowering duration (days)	CAS	Cashew apple shape		
FI	Flowering intensity (%)	CMN	Colour of mature nut shell		
ANR	Apple to nut ratio	NS	Nut shape		
SP	Shelling percentage	CBL	Colour of boot leaf		
KWT	Kernel weight (g)	CI	Compactness of inflorescence		
KL	Kernel length (cm)	TIB	Type of inflorescence branching		
KW	Kernel width (cm)	SFL	Secondary flowering		
KT	Kernel thickness (cm)	AB	Shape of cashew apple base		
CY	Cumulative yield (kg/plant)	RA	Ridges on cashew apple		

prescribed limit of 20% to infer that the core collections derived are really efficient. The CR% represents the range of different characters in the core collection with reference to the entire collections and it should be more than 80% (Hu *et al.*, 2000; Kim *et al.*, 2007). Accordingly, it was observed that the power core has more CR% compared with the cluster core, and hence the power core is more efficient compared with the cluster core. As per the criteria, larger the values of VD% and VR%, the more the extent to which the core collection represents the entire collection. The VD% and VR% values indicated that the power core has

higher values for these parameters, and hence the power core is an efficient representative of the entire collection compared with the cluster core. The CC% for quantitative traits in the power core was maximum. The CC% is not applicable to the cluster core because it is derived out of random allocation of accessions from all clusters resulted due to *K*-Means clustering.

With respect to qualitative characters, the CC% in the cluster core was less than the power core with reference to the entire collection. So it is evident that the power core collection is efficient in terms of retaining classes and hence

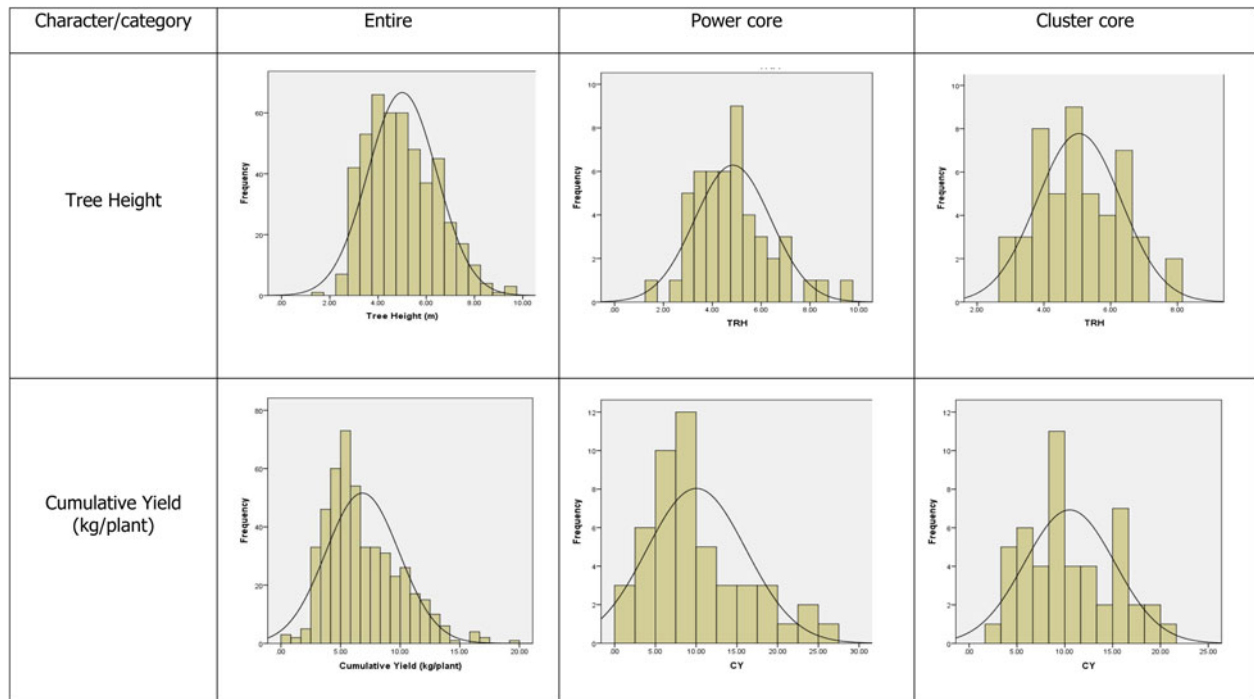


Fig. 2. Frequency distribution pattern of characters in the entire and core collections.

diversity present in the entire collection. The segregation of accessions in the cluster core is weaker compared with the power core as evident by the variability accounted through first two axes and their graphical representation. This indicates that the power core accessions are divergent from each other, and hence they retain maximum diversity with minimum accessions compared with the cluster core.

The *F*-test clearly indicated the higher variance level in the power core collections compared with either cluster core or entire collection. This is expected because of reduction in redundancy of accessions and effective stratification in the power core collection. The mean values of 27 quantitative characters between the power core and entire collections did not significantly differ except for the character sex ratio. This shows the efficiency of heuristic approach in making core collections. As far as Shannon–Weaver and Nei’s diversity values are considered, several workers infer that the core collections with comparable diversity indices in relation to the entire collection can be considered as representatives (Upadhyaya *et al.*, 2003; Dwivedi *et al.*, 2008; Upadhyaya *et al.*, 2009; Vijayanthi *et al.*, 2015a, b). However, Odong *et al.* (2013) deduced that higher the diversity index, higher is the retention of diversity by the core collection. So it is evident that the power core collections with higher or equal values of these diversity indices in comparison to the entire collection reflect better diversity compared with the cluster core collections. However, *t*-test indicated that means of Shannon–Weaver and Nei’s diversity indices of the entire, power core and

cluster core collections for these qualitative characters were not significantly different from each other. This is also conformity with the inference of Vijayanthi *et al.* (2015a, b) that mean Shannon–Weaver and Nei’s diversity indices in the core collections should be comparable to that of the entire collection. The results of χ^2 test revealed that the percentage of significant characters in both collections is well below the prescribed 20% (Vijayanthi *et al.*, 2015a, b), and hence it can be concluded that both core collections represent the entire collection when qualitative characters are considered.

The results of frequency distribution patterns of quantitative traits indicated that the patterns of the entire collection are more or less retained in the power core collection. Further, coefficient of variation values were higher for the power core collections compared with the entire and cluster core collections, which clearly indicated that the power core collections possess more diversity compared with the entire or cluster core collections.

When the geographic representation of the entire and core collections are considered, it appears that the Kerala and Maharashtra collections are more diverse compared with the collections from other states. Karnataka and Andhra Pradesh accessions are also quite divergent as indicated by their share in the power core collection. Tamil Nadu, Orissa, Brazil and West Bengal collections seem to be less divergent as their representation is very less in the power core collection. It is worthwhile to note that the Pondicherry and Mozambique collections represented in

the entire collection are not represented both in the power core and cluster core collections. However, West Bengal collections that are represented in the power core are not represented in the cluster core collection.

Conclusion

From the validation parameters, patterns of segregation, and variance, mean and frequency distribution of traits in the core collections compared with the entire collection, it is evident that the power core collection generated through heuristic approach is very effectively capturing the diversity present in the entire collection. Hence this core, because of its small size and less number of duplicates, can be gainfully utilized as working collection for cashew germplasm conservation and crop improvement programmes. These accessions can be evaluated in different environment conditions to identify widely/specifically adapted stable accessions, which further can be utilized in cashew improvement. This is the first ever attempt of use of advanced M strategy with heuristic approach in cashew, which is one among the plantation crops of India and this can be profitably extended to other plantation crops as well.

Since morphometric characters are used in the study, there is an inherent phenomenon of genotype \times environment interaction on the expression of characters. Therefore, the accessions in the core collection may not be genetically representing the diversity present in the entire collection. This can be circumvented through deployment of molecular markers for the development of molecular core collection from the entire collection, which will throw light on the genuineness of the core collection designated using morphological characters.

Supplementary material

The supplementary material for this article can be found at <https://doi.org/10.1017/S1479262118000035>

Acknowledgement

The authors thank Dr S. Ramesh, Professor, Department of Genetics and Plant Breeding, University of Agricultural Sciences, Bangalore, India for his inputs in data analysis

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Anonymous (1995) *RHS Colour Chart*. London, UK: The Royal Horticultural Society.
- Bataillon TM, David JL and Schoen DJ (1996) Neutral genetic markers and conservation genetics: simulated germplasm collection. *Genetics* 144: 409–417.
- Brown AHD (1989) Core collections: a practical approach to genetic resources management. *Genome* 31: 818–824.
- Brown AHD (1995) The core collections at the cross roads. In: Hodgkin T, Brown AHD, Van Hintum TJJ and Morales EAV (eds) *Core Collections of Plant Genetic Resources*. West Sussex, UK: John Wiley and Sons and Rome, Italy: IPGRI.
- Chandra S, Huaman Z, Hari Krishna S and Ortiz R (2002) Optimal sampling strategy and core collection size of Andean tetraploid potato based on isozyme data—a simulation study. *Theoretical and Applied Genetics* 104: 325–334.
- Chung HK, Kim KW, Chung JW, Lee JR, Lee SY, Dixit A, Kang HK, Zhao W, McNally KL, Hamilton RS, Gwag JG and Park YJ (2009) Development of a core set from a large rice collection using a modified heuristic algorithm to retain maximum diversity. *Journal of Integrative Plant Biology* 51: 1116–1125.
- Dwivedi SL, Puppala N, Upadhyaya HD, Manivannan N and Singh S (2008) Developing a core collection of peanut specific to Valencia market type. *Crop Science* 48: 625–632.
- Franco J, Crossa J, Taba S and Shands H (2003) A multivariate method for classifying cultivars and studying group \times environment \times trait interaction. *Crop Science* 43: 1249–1258.
- Frankel OH (1984) Genetic perspective of germplasm conservation. In: Arber W, Llimensee K, Peacock NJ and Starlinger P (eds) *Genetic Manipulation – Impact of Man and Society*. Cambridge, UK: Cambridge University Press, pp. 161–170.
- Gireesh C, Husain SM, Shivakumar M, Satpute GK, Kumawat G, Arya M, Agarwal DK and Bhatia VS (2015) Integrating principal component score strategy with power core method for development of core collection in Indian soybean germplasm. *Plant Genetic Resources*, available on CJO2015. doi: 10.1017/S1479262115000556.
- Gouesnard B and Bataillon TM (2001) MSTRAT: an algorithm for building germplasm core collections by maximizing allelic or phenotypic richness. *Journal of Heredity* 92: 93–94.
- Gowda J, Krishanappa M, Pathak N, Mathur PN and Seetharam A (2013) Use of heuristic approach for the development of a core set from large germplasm collection of foxtail millet (*Setaria italica* L.) Indian. *Journal of Plant Genetic Resources* 26: 13–18.
- Hu J, Zhu J and Xu HM (2000) Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theoretical and Applied Genetics* 101: 264–268.
- IBPGR (1986) *Cashew Descriptors*. Rome, Italy: International Board for plant Genetic Resources (presently, International Plant Genetic Resources Institute), p. 33.
- Jansen J and van Hintum JL (2007) Genetic distance sampling: a novel sampling method for obtaining core collections using genetic distances with an application to cultivated lettuce. *Theoretical and Applied Genetics* 114: 421–428.
- Kim KW, Chung HK, Cho GT, Ma KH, Chandrabalan D, Gwag JG, Kim TS, Cho EG and Park YJ (2007) Powercore: a programme applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinformatics* 23: 2155–2162.
- Krishna H and Singh SK (2007) Biotechnological advances in mango (*Mangifera indica* L.) and their future implication in crop improvement – a review. *Biotechnological Advances* 25: 223–243.

- Marita JM, Rodriguez JM and Nienhuis J (2002) Development of an algorithm identifying maximally diverse core collections. *Genetic Resources Crop Evolution* 47: 515–526.
- Moe KT, Gwag JG and Yong JP (2012) Efficiency of power core in core Set development using amplified fragment length polymorphic markers in Mungbean. *Plant Breeding* 131: 110–117.
- Nayak MG, Mohana GS, Bhat PS, Saroj PL and Swamy KRM (2014) *Minimum Descriptors of Cashew Germplasm Accessions*. Catalogue-IV, Puttur: Directorate of Cashew Research, p. 32.
- Nayak MG, Mohana GS, Bhat PS, Saroj PL, Swamy KRM and Bhat MG (2015) *Minimum Descriptors of Cashew Germplasm Accessions*. Catalogue-V, Puttur: Directorate of Cashew Research, p. 44.
- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583–590.
- Odong TL, Jansen J, van Eeuwijk FA and van Hintum TJJ (2013) Quality of core collections for effective utilization of genetic resources review, discussion, and interpretation. *Theoretical and Applied Genetics* 126: 289–305.
- Peeters JP and Martinelli JA (1989) Hierarchical cluster analysis as a tool to manage variation in germplasm collections. *Theoretical and Applied Genetics* 78: 42–48.
- Schoen DJ and Brown AHD (1993) Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. *Proceedings of National Academy of Sciences, USA* 90: 10623–10627.
- Shannon CE and Weaver W (1949) *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.
- Sturges H (1926) The choice of a class-interval. *Journal of American Statistical Association* 21, 65–66.
- Swamy KRM, Bhaskara Rao EVV and Bhat MG (1997) *Catalogue of Minimum Descriptors of Cashew (Anacardium occidentale L.)*. Germplasm Accessions-I. Puttur, Karnataka: National Research Center for Cashew, p. 41.
- Swamy KRM, Bhaskara Rao EVV and Bhat MG (1998) *Catalogue of Minimum Descriptors of Cashew (Anacardium occidentale L.)*. Germplasm Accessions-II. Puttur, Karnataka: National Research Center for Cashew, p. 54.
- Swamy KRM, Bhaskara Rao EVV and Bhat MG (2000) *Catalogue of Minimum Descriptors of Cashew (Anacardium occidentale L.)*. Germplasm Accessions-III. Puttur, Karnataka: National Research Center for Cashew, p. 54.
- Thimmappaiah and Shirly RA (1999) In vitro regeneration of cashew (*Anacardium occidentale L.*). *Indian Journal of Experimental Biology* 37: 384–390.
- Upadhyaya HD, Ortiz R, Bramel PJ and Singh S (2003) Development of a groundnut core collection using taxonomical, geographical, and morphological descriptors. *Genetic Resources and Crop Evolution* 50: 139–148.
- Upadhyaya HD, Pundir RPS, Dwivedi SL, Gowda CLL, Reddy VG and Singh S (2009) Developing a mini core collection of sorghum for diversified utilization of germplasm. *Crop Science* 49: 1769–1780.
- Vaijayanthi PV, Ramesh S, Byre Gowda M, Mohan Rao A and Keerthi CM (2015a) Development of core sets of Dolichos bean (*Lablab purpureus L. Sweet*) germplasm. *Journal of Crop Improvement* 29: 405–419.
- Vaijayanthi PV, Ramesh S, Byre Gowda M, Mohan Rao A, Gowda J, Ramappa HK, Keerthi CM and Rajendra Prasad BS (2015b) Development and validation of a core set of Dolichos bean germplasm. *International Journal of Vegetable Science* 21: 419–428.