JOURNAL OF THE DRYLANDS 6(2): 519 - 530, 2016

ESTABLISHING A CORE COLLECTION OF FINGER MILLET (*Eleusine coracana* [L.] Gaertn.) *EX SITU* HOLDINGS OF THE ETHIOPIAN GENEBANK

Yemane Tsehaye

Yemane Tsehaye (2016): Establishing a core collection of finger millet (*Eleusine coracana* [L.] Gaertn.) *ex situ* holdings of the Ethiopian genebank. Journal of the Drylands, 6(2): 519 - 530, 2016.

The Ethiopian genebank, a pioneer genebank in Africa, has large number of collections of different crops species (including finger millet) collected from the country, donations as well as repatriations from other African countries such as Zimbabwe, Zambia as well as Eritrea. Nonetheless, the vast genetic stock conserved so far constrain genebank management and proper utilization. The main objective of this work was to establish a representative core sets. A total of 677 finger millet accessions (391, 10, 250 and 26 of Ethiopia, Eritrea, Zimbabwe and Zambia origin respectively) and altogether 13458 individual plants were characterized and evaluated for 11 quantitative and five qualitative morphological traits in augmented design. Using Best Linear Unbiased Estimator (BLUE) derived from the evaluation data, four core subsets were developed. The first three core sets (Core 1-3) were assembled using principal components scores strategy (PCSS). Accessions in core 4 (10% of the entire collections) were formulated using logarithmic sampling approach. To facilitate utilization and simplify advanced evaluation, accessions in Core 4 were further grouped into 'mini' cores using the Ward's clustering. A pooled Shannon-Weaver index (H') was calculated on the five qualitative traits in all established core sets and appeared to represent well the level of diversity existed in the whole *ex situ* holdings. Such a small sized core set is expected to enhance genebank management and facilitate germplasm evaluation against a number of abiotic and biotic stresses including blast and may also enhance the effort of screening accessions for quality traits particularly essential amino acids and malting quality.

Key words: Core collection, principal components scores, conservation, utilization

Mekelle University (Ethiopia), College of dryland agriculture and natural resources

E-mail: yemtse@yahoo.com

Received: March 15, 2016; Accepted: November 30, 2016.

INTRODUCTION

It has been long that certain eco-geographic areas were recognized to harbor unusually high diversity (cultivated as well as wild germplasm) and these gene repositories were perceived as important resources for crop improvement and conservation (Vavilov 1951). After the systematic collection mission in the 1920's and exploration and expedition effort of Vavilov, particularly to the important gene centers including the Abyssinian center (Ethiopian center), a worldwide efforts were launched to thoroughly collect the genetic diversity of major food crops. Nonetheless, the vulnerability of the gene repositories to genetic erosion and complete extinction possibly resulting from technology and economic changes (Harlan and Martini 1935) has raised concern. It was this concern that has later led to the establishment of genebanks in many parts of the world including the Ethiopian genebank. Apparently landraces (hereafter referred as farmers' varieties) and wild relatives of crop germplasm were intensively collected and conserved in several genebanks around the world.

Despite the extensive collection and expedition worldwide efforts, the *ex situ* conservation strategy has resulted in immense holdings. Thus, genebank collections in many parts of the world (including the Ethiopian genebank) have grown as large as to hamper the very purposes, not only conservation but also sustainable utilization of the genetic diversity they seized. The ever increasing size of the germplasm collections has obstructed efforts to enhance utilization of the existing diversity because of the difficulties to properly characterize and evaluate such large holdings and bearing the attached cost of managing the germplasm collections.

Cognizant to the large size of genetic resources collections that could possibly deter utilization, Frankel (1984) proposed establishment of manageable size or "core collection" from the existing collections. According to Frankel (1984) and Frankel and Brown (1984), core collection is the process of pruning the sheer size of large collections into manageable size with minimum redundancy but representative samples in terms of genetic diversity spectrum of the original collection. This idea comes as a response to the challenge of the already growing size and number of plant genetic resources collections in most of the genebanks around the globe including the Ethiopian genebank. Core collection is at present a widely recognized concept and accepted as efficient tools for improving conservation and utilization. The Global Plan of Action for Conservation and Sustainable Utilization of Plant Genetic Resources for Food and Agriculture (1996)

ISSN 1817-3322

recommend core collection establishment as one of the activities to enhance utilization of genetic resources.

The important step in the establishment of core collection is to decide on how many entries the final core set should contain. On the basis of the neutral allele model, Brown (1989) suggested that a core collection should contain 10% of the whole collection, with a maximum of 2000 entries, which allows preservation of about 80% of alleles. Most of the core collections so far established, however, was 5 - 20% of the size of the original (whole collection) collection (Charmet and Balfourier 1995, Van Hintum 1999).

Several options and models are available in the literature for structuring and estimating the existing variability in *ex situ* collection as well as selection and validation of accessions for core establishment. The rationale behind the available models is to identify a natural hierarchical structure in the entire collection and establish a representative small subset that provides access to existing diversity in a given species. Stratifying the entire collection based on geographic origins or morphological descriptors and employing random sampling approach for selecting a representative set of small size was suggested (Frankel and Brown, 1984). Alternative approaches such as the use of genetic markers and coefficient of parentage (Van Hintum and Haalman 1994) and principal component scores strategy (PCSS) (Noirot et al 1996, Balakrishnan et al 2000, Mahajan et al 2007) were also forwarded. The latter approach (PCSS), was no longer random unlike the one proposed by Frankel and Brown (1984), and was designed to maximize diversity and avoid redundancy (Balakrishnan et al 2000). The size of the core set was decided either by comparing the diversity of the core collection with that of the original collection using the Shannon-Weaver Diversity Index (e.g. Balakrishnan et al 2000) on the qualitative traits or other statistical procedures such as the coincidence rate on quantitative traits as suggested by Hu et al (2000).

The Ethiopian genebank (founded in 1976) that holds the largest collections in sub-Saharan Africa, except for barley (Tsehaye et al 2012), has not yet attempted to establish core collection of any crop species including finger millet. The cultivated species of the genus Eleusine, finger millet, is of very ancient cultivation in Ethiopia, since 5000 BC (Hilu et al 1979). It is the sixth important crop and comprises about five percent of the total land devoted to cereals (CSA 2013) and has gain little attention and relatively few researchers involved in its improvement. This was partly due to little understanding of the level of genetic diversity of the genetic stock conserved *ex situ* and lack of proper characterization and evaluation data base in the national genebank. This work was initiated to establish representative core set of finger millet *ex situ* collections as a first effort of formulating core collection to offer to users' (breeders and farmers) better opportunity for utilization and efficient conservation.

MATERIALS AND METHODS

Plant materials and research procedure

The Ethiopian national genebank is a pioneer genebank in Africa (established in 1976) and has germplasm collection not only from Ethiopia and Eritrea (former part of Ethiopia), but also from other African countries such as Zimbabwe and Zambia. The genetic materials from Zimbabwe and Zambia were collected in 1982 and donated along their passport data to the Ethiopian genebank. At the time of this particular study, the total finger millet holdings in the Ethiopian genebank were close to 1000 accessions (farmers' varieties collected from farmers field) and about 30% of these collections were donated from the southern African countries (Zimbabwe and Zambia).

The finger millet *ex situ* collections were planted at Arsi Negele Research Center (Altitude, 1960 m.a.s.l, $7^0$30' N and 39^0 00' E) during 2008/2009 planting season. This research station (Arsi Negele) was nationally designated as sorghum and finger millet evaluation site. The accessions were planted in two-row plots of 5 m row length with a row-to-row spacing of 0.75 m and plant-to-plant spacing of 0.10 m in an augmented design. The finger millet genetic materials were randomly divided into 20 incomplete blocks (50 accessions and two released varieties [Tadesse and Padat-1] as control). The two control varieties were replicated and randomized throughout the blocks to estimate the error mean square and block effect.

After eliminating accessions with incomplete passport data and poor establishment in the field, 677 accessions (391, 10, 250 and 26 accession of Ethiopian, Eritrea, Zimbabwe and Zambian origin, respectively) were considered for analysis. On average, 20 plants from each accession were sampled. A total of 13458 individuals were characterized and evaluated for 11 quantitative and five qualitative agro-morphological traits (Table 1).

ISSN 1817-3322

Table 1. List of qualitative and quantitative traits along with their unit of measurement

Qualitative traits	Code	Trait classes
Panicle shape	PS	1: droopy, 2: straight, 3: compact
Finger branching	FB	0: absent, 1: present
Seed color	SC	1: White, 2: orange to red, 3: black
Glume color	GC	1: White to yellow, 2: reddish brown, 3: black
Pericarp persistence	PP	1: none - persistence, 2:partially persistence, 3: persistent
Quantitative traits	**Code**	**Units**
Plant height	PLH	meter
Number of culm branches	CBR	count
Finger number	FIN	count
Finger width	FIW	cm
Finger length	FIL	cm
Peduncle length	PEL	cm
Peduncle exertion	PEX	cm
Yield per plant	YPP	gram
Hundred seed weight	HSW	gram
Days to 50% flowering	DTF	Days (days from sowing)
Days to 75% maturity	DTM	Days (days from sowing)

Data analysis

Model fitting and classification analysis`

A Best Linear Unbiased Estimator (BLUE's) was estimated for each of the 11 quantitative traits. The data were standardized to mean zero and standard deviation of one, prior to the execution of most of the statistical analysis procedures.

To formulate core sets as per the suggested model (Figure 1), a correlation matrix generated using the BLUE values of the 11 quantitative traits were subjected to principal components analysis (PCA).

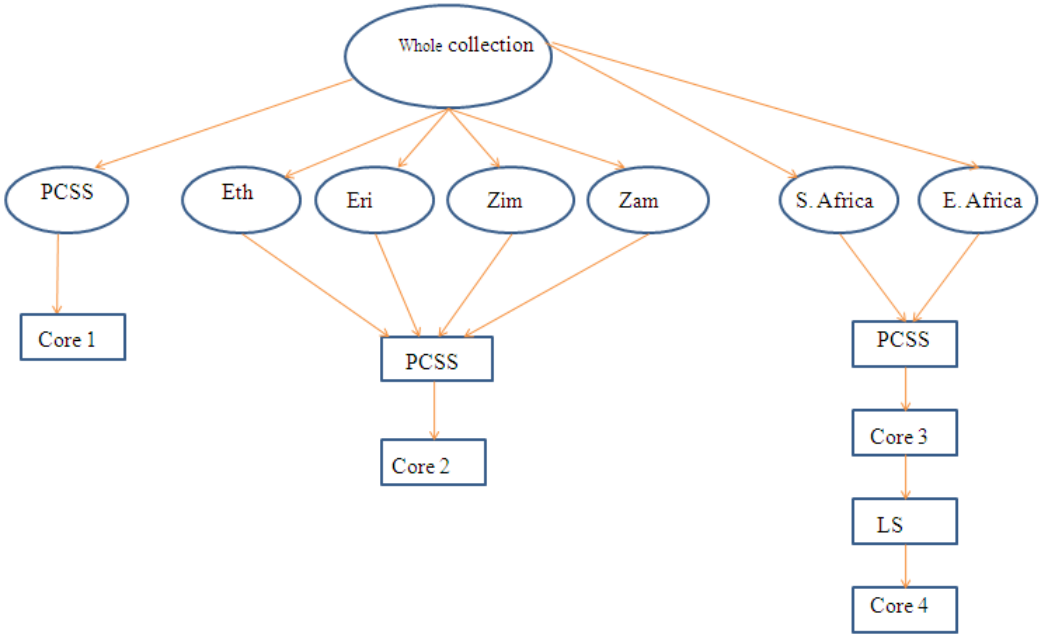

Figure 1. Model used to establish the core subsets, Eth: Ethiopia; Eri: Eritrea; Zim: Zimbabwe; Zam: Zambia; S. Africa: South Africa; E. Africa: East Africa; LS: logarithmic sampling

The contribution of the i^{th} accession to the total variability was computed using the Principal Components Scores Strategy (PCSS) as described by Noirot et al (1996) and given as follows:

ISSN 1817-3322

$$P_i = \sum_{j=1}^{t} y^2{}_{ij} \quad (1)$$

where y_{ij} is the component score of the i^{th} accession on the j^{th} principal component. The relative contribution (RC) of i^{th} accession was computed as:

$$RC_i = \frac{p_i}{(p \times t) \times 100} \quad (2)$$

where p stands for total number of accessions and t stands for the number of principal components. The cumulative inertia was computed by successively adding the contribution of each accession (sorted in descending order according to their contribution).

The representative accessions were objectively decided by fitting a logistic regression model (Van de Plank 1963) as follows:

$$\frac{y}{(A-y)} = exp(a + bn) \quad (3)$$

As described by Balakrishnan et al (2000) the linear form of the logistic regression model was given as:

$$ln\left[\frac{y}{(A-y)}\right] = a + bn \quad (4)$$

where ln is log base e; a and b are intercept and regression coefficient respectively; y is the cumulative inertia, n is the serial number (sorted data) and A stands for asymptote of the curve, which equals 100 (Balakrishnan et al 20000). The rate of progress of inertia ($\frac{dy}{dn}$) was computed as:

$$\frac{dy}{dn} = by(A-y) \quad (5)$$

To classify the accessions based on their eco-geographic region of origin (Figure 2) and assess the level of gene flow among countries, a linear discriminant analysis was executed on the standardized data and the level of misclassification was cross-validated (Table 2).

Core subset formulation

A model was suggested to formulate core sets (Figure 1), and apparently four core subsets were developed. Three of them were created using non-random sampling strategy (PCSS approach) whereas, the fourth one was developed using a random sampling approach (logarithmic sampling strategy). The number of accessions included at the inflection point (cut-off point) when the PCSS was executed on the whole collection using the 11 quantitative traits was designated as Core set 1 (Core 1) (Figure 1 and 3). Core set 2 was constituted by running PCSS in each region (region of origin) and accessions at the inflection point of each region (accession contributed more to the cumulative inertia from each region) were assembled together. Core 3 was formulated by selecting accessions that contributed the most to the east African (Ethiopia and Eritrea) and southern Africa (Zimbabwe and Zambia) groups (hereafter eastern/southern Africa zonation) using the PCSS approach. In line with the original definition of the concept (Frankel and Brown 1984), a core set (Core 4, a random subset) that constitutes 10% of the whole collection was formed (Figure 4) based on a random sampling approach, the logarithmic strategy (LS) (Brown 1989), from the accessions assembled as Core 3 (that represent both the eastern and Southern African zonation). The LS approach appeared to standardize the sample size differences (between groups of different sizes). It also identifies rare variants (Brown 1989) with significance to crop adaptations and better sampled poorly represented samples (Zeuli and Qualset 1993) and may reduce genetic redundancy (Grenier et al 2000).

Validation of the core subsets

To confirm the representativeness of the formulated core sets, the a Shannon-Weaver diversity index (H'), which has been used in diversity studies of germplasm collections (e.g. Jain et al 1975), was estimated on the frequency data of the five qualitative traits (Table 3); and the index was estimated as follows:

$$H' = -\sum_{i=1}^{n} \frac{pi\ ln\ pi}{ln\ n} \quad (6)$$

where p_i is the proportion of the total number of individuals (accessions) in the i^{th} class (trait class) and n is the number of phenotypic classes for a given trait.

Further clustering of the core set

To facilitate utilization and simplify advanced evaluation of the *ex situ* collections (*vis.* physio-agronomic evaluation as well as molecular marker analysis), accessions in Core 4 (10% of the whole collection) were further grouped into 'mini' cores (Upadhyaya et al 2012). In this analysis, four optimal clusters (Figure 5) were found using the Ward's clustering (agglomerative, hierarchical classification technique with incremental sums-of-squares sorting strategy) algorithm (Ward 1963). The distinctiveness of the clusters and divergence among them was assessed using descriptive statistics (Table 4) and linear discriminant analysis (clusters were used as a classifying variable) (Figure 6) and Mahalanobis's distance (D^2) (Mahalanobis 1936). The D^2 estimates obtained from the pairs of clusters (Table 5) were considered

ISSN 1817-3322

as calculated Chi-square (χ^2) (Singh and Chaudhary 1985) values and were tested for their significance both at 1% and 5% probability.

RESULTS

Classification based on eco-geographic region of origin

Discriminant analysis executed on the 11 quantitative traits, which were standardized to mean zero and unity variance, has shown a clear eco-geographic region of origin based pattern (with 95% confidence ellipses), with slight misclassification (Figure 2, Table 2). Three dimensions (DIM's) were found to be significant and the first two dimensions (DIM 1 and DIM 2) accounted for about 98.6% of the total variation (Figure 2).

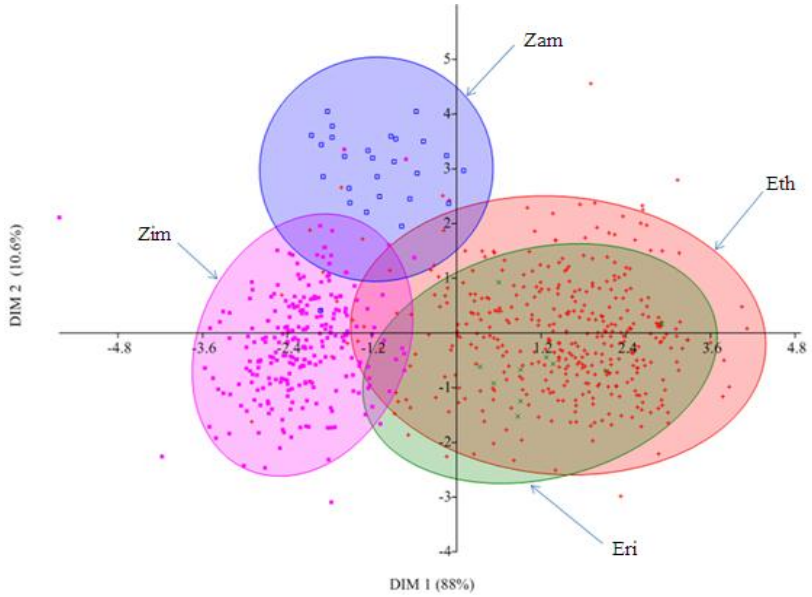

Figure 2. Classification of accession based on their eco-geographic region of origin using a linear discriminant analysis and 95% ellipses, Eth: Ethiopia; Eri: Eritrea; Zim: Zimbabwe; Zam: Zambia

The first dimension or axis (DIM 1) clearly separated the east African from the South African finger millet germplasm (accessions) and accounted for about 88% of the total variation. About 81% and 80% of the original as well as cross-validated data respectively, were correctly classified into their respective region of origin (Table 2).

Table 2. Classification of accessions (in percent) based on the 11 standardized quantitative traits

Data	Regions	Predicted group membership			
		(Eth)	(Eri)	(Zim)	(Zam)
Original	Eth	69	23	5	4
	Eri	10	90	0	0
	Zim	0	0	97	3
	Zam	0	0	4	96
Cross-validated	Eth	69	23	5	3
	Eri	20	80	0	0
	Zim	0	0	96	4
	Zam	0	0	4	96

Eth: Ethiopia; Eri: Eritrea; Zim: Zimbabwe; Zam: Zambia

The percentage of misclassified accessions was relatively higher in the Ethiopian germplasm which shared 23%, 5% and 3% with Eritrea, Zimbabwe and Zambia respectively. Despite few accessions of Eritrean origin (10 accessions) that were incorporated, yet about 10% of them were misclassified with the Ethiopian materials. The frequency of misclassification has increased to 20% in the cross-validated data (Table 2). The southern African accessions were highly confined to their eco-geographic region of origin where about 4% of the accession from both regions, Zimbabwe and Zambia, respectively were shared between each other (Table 2). None of the southern African

ISSN 1817-3322

regions were found to share even a single accession with the eastern Africa regions (Ethiopia and Eritrea).

Core development

A total of 11 principal components (PC's) (as equal as the number of the quantitative traits) were extracted. Component scores from all the PC's were used for computation purposes and establishment of the core sets. The PCSS approach was used to select the accessions that contributed the most to the overall diversity in three independent groups (whole collection, respective region of origin and eastern/southern African zonation) (Figure 1). The PCSS involved in the selection of accessions with higher relative contribution (RC) to the overall inertia up to the inflection point (cut-off point) was identified by the linearization of the logistic regression model (Figure 3). Based on the non-random sampling approach (PCSS), three core sets (Core 1 - 3) were formulated. This sampling approach, PCSS, picked up about 31% (218 accessions) from the whole collection, and altogether 28% (185 accessions) from eastern/southern Africa zonation (Figure 3) and formulated as Core 1 and 3 respectively. Furthermore, about 29% (195 accessions) accessions were sampled (bulked samples of the four respective countries) using the PCSS approach, and designated as Core 2 (Figure not shown).

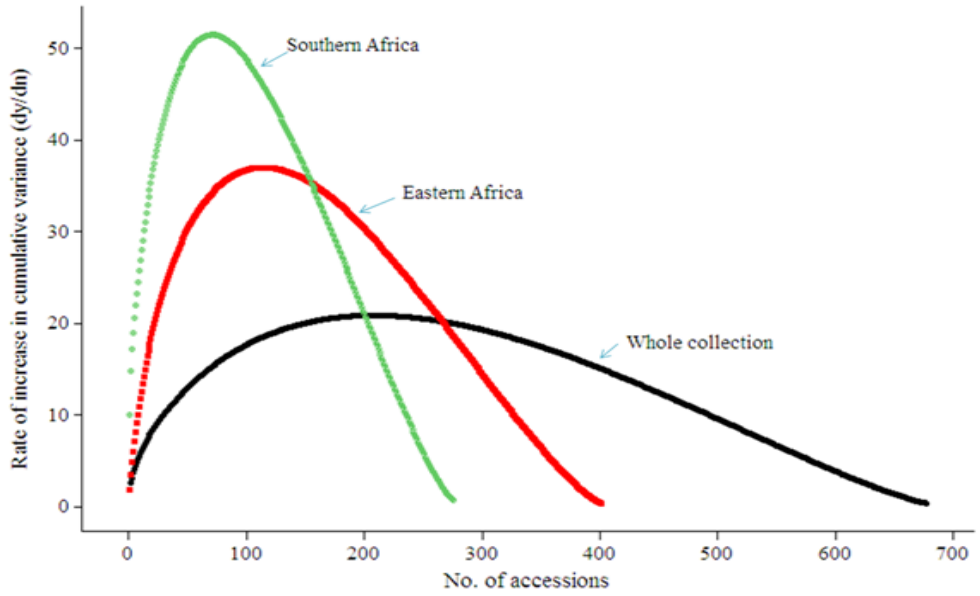

Figure 3. Rate of progress curve of cumulative contribution of accessions to the cumulative variance and the number of accessions selected as core set based on PCSS and fitted logistic regression

The relative size of Core 4 (random subset) was adjusted to 10% of the entire collection (68 accessions) where accession were sampled randomly using the logarithmic strategy (LS) from the specific assemblage of accession designated as Core 3. In Core 4, about 36 eastern African and 32 southern African accessions (35, 23, 9 and 1 randomly selected accessions of Ethiopia, Zimbabwe, Zambia and Eritrea) were assembled together and the remaining accessions were considered as a reserve collection (Figure 4).

ISSN 1817-3322

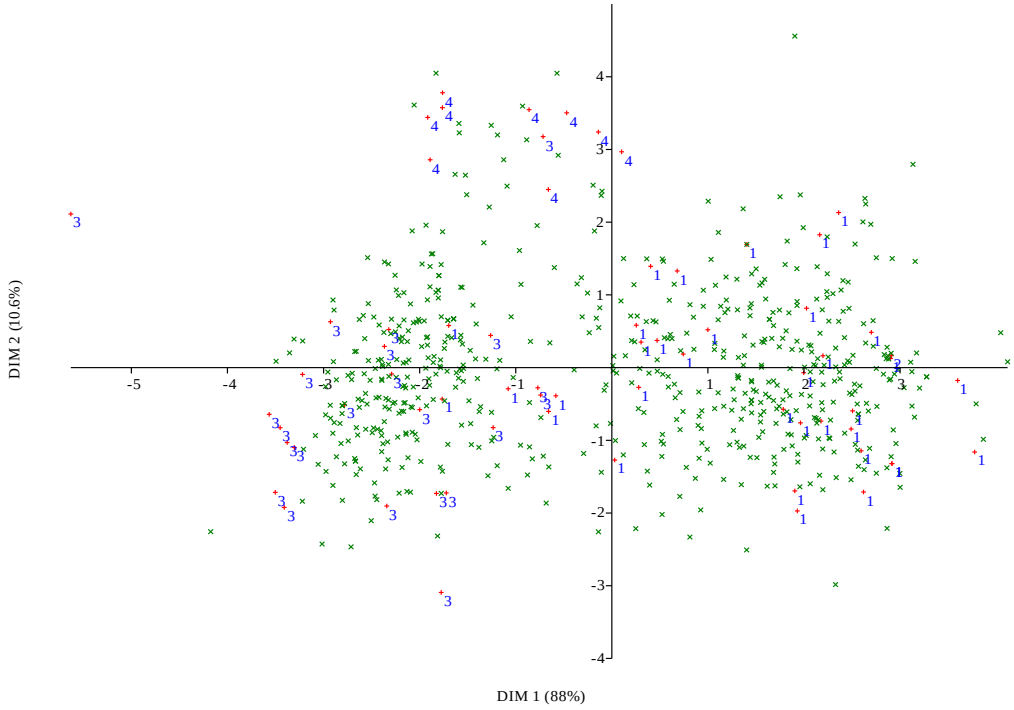

Figure 4. Accessions selected using the LS strategy (10% of the entire collection and designated as Core 4) along with reserve accession (**x** mark), Numbers refers to the selected accessions (1: Ethiopia, 2: Eritrea, 3: Zimbabwe, 4: Zambia) and accessions with **x** mark are reserve accessions

Representativeness of the core sets

The standardized and pooled values of the Shannon-Weaver index (*H'*) estimated from the five qualitative descriptors were used to compare the representativeness of the established core sets (Core 1 - 4) with the whole collection (Table 3). The pooled estimated values (average of all qualitative traits) of all the core samples including the Core 4 were found to be more or less equivalent and fairly comparable to the whole collection (Table 3).

Table 3. Estimated H' values of the whole collection and the core subsets

Traits	**H'**				
	Whole	**Core 1**	**Core 2**	**Core 3**	**Core 4**
PS	0.88	0.88	0.87	0.87	0.85
FB	1.00	0.98	1	1	0.96
SC	0.69	0.77	0.80	0.81	0.83
GC	0.91	0.87	0.91	0.90	0.86
PP	0.87	0.78	0.77	0.77	0.79
Mean	**0.87**	**0.86**	**0.87**	**0.87**	**0.86**

H': estimated on the five qualitative traits (PS: panicle shape; FB: finger branching; SC: seed color; GC: glume color; PP: panicle persistence

Despite the little discrepancy in the estimated *H'* values for some traits such as SC and PP, the pooled *H'* value of the core constructed using the LS sampling strategy (Core 4) was less only in 0.01 (insignificant quantity) from the whole collection (Table 3) that indicates its representativeness.

Grouping within the random subset

Ward's clustering and discriminant analysis were further employed in Core 4 (random subset) and the cluster analysis clearly identified four distinct groups (Figure 5). The clustering algorithm appeared to group accession of different origin together (different eco-geographical settings) that may provide wide variability for users (breeders and farmers). Cluster 1 comprised 25 accessions that had relatively higher number of fingers and finger width, better grain yield and seed weight, and relatively late maturing accessions (Table 4).

ISSN 1817-3322

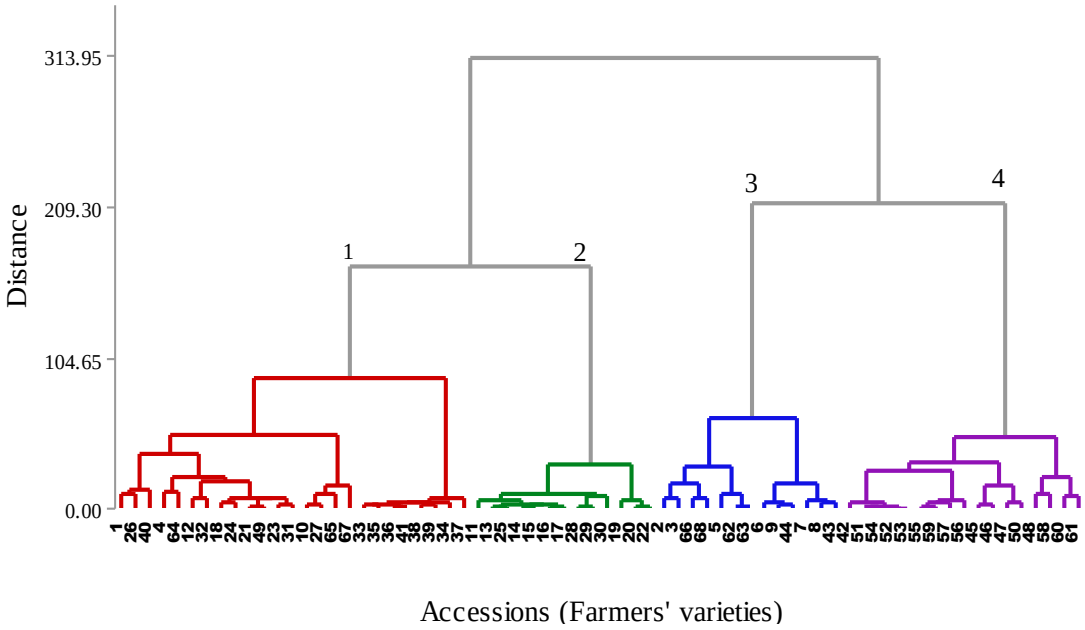

Figure 5. Clustering of accessions in the random subset (Core 4) using the Ward clustering algorithm (accession numbers are re-coded)

The number of accessions in Cluster 2 was 13 and were chiefly characterized as having long stature, late in flowering and maturity, higher number of culm branching and relatively large number of fingers. Cluster 3 comprised 13 accessions, relatively early flowering/maturing ones and had shorter stature and peduncle length. The average finger length of this group was, however, very short (4.06, see Table 4). Accessions in cluster 4 (17 accessions) were characterized as early flowering and maturing and also having long stature, large number of fingers and finger length as well as long peduncle exertion. Moreover, they had high seed yield per plant and seed weight.

Table 4. Cluster mean of the 11 quantitative traits

Cluster	PLH	CBR	FIN	FIW	FIL	PEL	PEX	YPP	HSW	DTF	DTM
1	0.87	1.04	8.58	1.07	5.32	1.24	12.35	11.11	0.37	100.30	166.22
2	1.03	1.13	8.07	0.69	8.55	1.77	9.97	5.95	0.20	105.43	169.49
3	0.56	1.34	6.85	0.94	4.06	0.45	7.94	7.55	0.31	90.82	147.20
4	1.46	0.70	9.55	0.54	8.83	0.99	17.44	9.40	0.36	89.93	153.44

The linear discriminant analysis (where the two DIM's accounted for 88% of the variation) also supported the distinctness of the four groups acquired from the cluster analysis (Figure 6).

ISSN 1817-3322

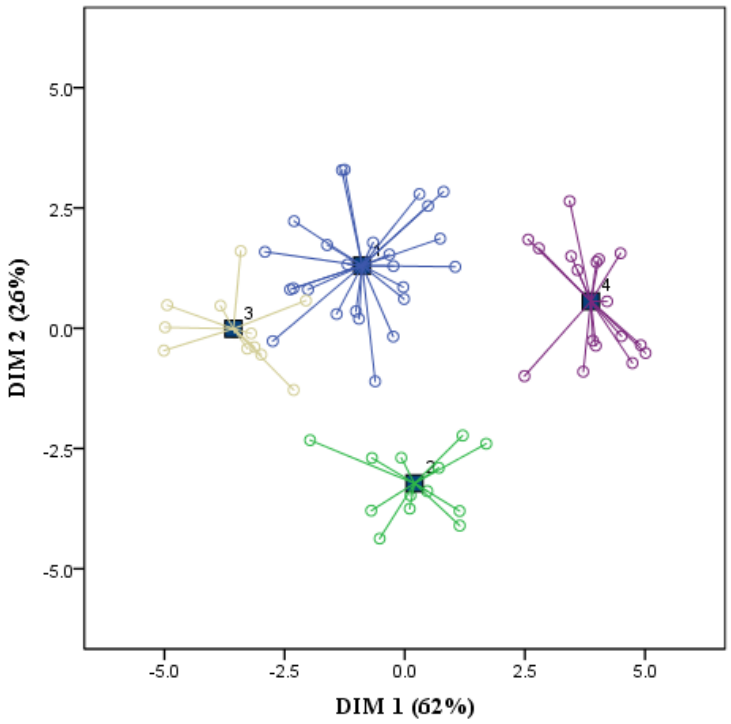

Figure 6. Ordination of the accessions in the random subset (Core 4) using a linear discriminant analysis

The accuracy of divergence among all clusters (groups) constituted in Core 4 using cluster analysis and linear discriminant analysis was confirmed using a Mahalanobis's distance algorism (Table 5). The difference among the group centroids as estimated by the Mahalanobis distance was statistically significant. Moreover, the divergences between accession assembled in cluster 4 and the other clusters were fairly large and may have a breeding significance.

Table 5. Mahalanobis's distance between clusters

Cluster	1	2	3	4
1	-			
2	27.2**	-		
3	18.4*	35.9**	-	
4	36.4**	40.0**	72.0**	-

* P < 0.05, ** P < 0.01

DISCUSSIONS

The alarming threat of genetic erosion to the wealth of the Ethiopia's genetic resources has led to an urgent need of launching plant expedition, collection and conservation mission to salvage the country's still abundant genetic resources. Thus, large number of accessions were collected and assembled in the *ex situ* genebank (under the Ethiopian Biodiversity Institute), operational since 1976. The existing holdings of the Ethiopian genebank constitutes genetic materials collected from farmer's field and acquired through transfer from several breeding centers within the country, donations by various national and international organizations. At present, the Ethiopian genebank overall holdings have reached about 75,000 accessions (NBASP 2014). Such huge number of accession of different species accumulates, however, questions of efficiency in management and utilization may inevitably arise. In large collections such as the Ethiopian *ex situ* holdings, virtually impossible to screen all available germplasm of a given species that carries the variability required for crop improvement.

The finger millet holding in the Ethiopian *ex situ* genebank contained a great level of spatial and genetic diversity. However, the available diversity has not been adequately tapped in breeding programs and so far only four improved and registered varieties are available (MOA 2013), and yet their adoption is still minimal. Despite the untapped breeding potential, the so far limited utilization of the *ex situ* collections of the crop in the country may attributed to its size and also unavailability of proper characterization and evaluation data. The major benefit of establishing a representative subsample in this work would, therefore, be to enhance the evaluation work in breeding programs that require fewer resources.

The eco-geographic region of origin based classification analysis (discriminant analysis) executed on the 11 quantitative traits scored on the

ISSN 1817-3322

entire collections provided good region level uniformity of distribution, clear structure between groups (eco-geographic region of origin) with very little overlap. This clear eco-geographic based population structure observed in the *ex situ* finger millet collections may indicate definite region based evolution and adaptation to specific cultural settings and system of cultivation.

The PCSS approach of sampling has set an optimal sampling fraction (sampling intensity) at the cut-off point (dy/dn curve). There were also identified accessions ranging from 28-31% (in Core 1 - 3) of the whole collection. Balakrishnan et al (2000) pursued similar approach and at the point of inflection they found about 27% of the entire collection as optimal sampling fraction, relatively higher than the originally suggested size (Brown, 1989). The sampling intensity of most of the core sets established so far, particularly using the random sampling approach appeared to range between 5 to 10% of the entire collection (Brown 1989, Charmet and Balfourier 1995, Van Hintum 1999, Brown and Spillane 1999). This was due to the anticipated core size that was predetermined concomitant with the original concept as defined by Frankel and Brown (1984). However, there are still some disagreements with such low sampling intensity (sampling up to 10%) in which sample may be insufficient to represent the level of diversity contained in the original collection (Diwan et al 1995). Apart from the relatively big sized core sets constituted by the PCSS approach, the selected subsamples appeared to represent well the level of diversity estimated in the whole collection as measured by H'. Close correspondence in the estimated H' value between an entire collection and a core set established using the PCSS approach was also reported (Balakrishnan et al 2000).

Despite the well representation of the selected subsets to the parental population, the number of accession sampled based on the PCSS may still be large that may restrict a thorough evaluation work, utilization as well as management. To come across this problem and to stick to the original definition of the concept (Brown 1989), a logarithmic sampling strategy (LS) was deployed on the already established (*via* non-random approach) core set (Core 3) and accessions were selected up to the predetermined size of 10% randomly. Apparently, a core set (Core 4) of manageable size (68 accessions [10% of the entire collection]) was established. This was in line with the newly postulated concept - 'mini core' set (Upadhyaya et al 2012) and yet concurrent to the original concept, 10% of the entire collection (Brown 1989).

As confirmed by the pooled H' value estimated in the qualitative traits, better representation of the existing diversity of the entire collection was obtained in the formulated 10% sized subset (Core 4). Nonetheless, the fairly similar phenotypic diversity observed in all the formulated core sets with that of the entire collection (as estimated using H') perhaps indicate the preponderance of duplicate (redundant) materials in the *ex situ* collections.

The further grouping of the random subset (Core 4) using the Ward's clustering algorism has constituted four distinct clusters that were divergent in flowering/maturity time as well as yield and its components. In areas with a restricted growing season resulted from terminal drought, the early maturing/flowering groups (cluster 3 and 4) may be targeted as a practical option. For locations receiving relatively better rainfall, accessions assembled in cluster 1 can fit very well (relatively late maturing and high yielding). Accessions assembled in cluster 4 were the most divergent ones (as revealed by the Mahalonobis distance) and crossing this group with either of the clusters (groups) may lead the development of transgresive segregants for most of the important traits. This may boost up the identification of superior genotypes having high yield and other component traits.

Thus, the most pressing question now may be how to manage and efficiently utilize the already established core subsets. From the genebank curators' point of view, the core subsets can be preserved as active collection at $+4^0$C (following the contemporary conservation procedure of the Ethiopian genebank), so that the subsets along with the characterization/evaluation data can be made available to breeders and other users. Regarding breeders, perhaps two options are available - splitting and/or bulking. Following a thorough evaluation work, the formulated core set could be splitted into homogenous components (e.g. based on seed color and spike morphology). For splitting into different components, a single plant selection can also be done *via* the single seed decent (SSD) approach. The purified individuals can be used as a parent material in crossing and multi-environments testing against biotic and abiotic stresses. Perhaps the other possible option is bulking the groups (respective clusters) as a single entry (Van Hintum, 1999).

The accessions clustered together may equally be mixed. The mixture of genotypes can be treated as a single variety and can be sown together (by farmers or other users) in an evolutionary breeding platform that can serve as a safeguard against climate change extremes (heat, cold and drought). Subsamples can be distributed (in splitted or bulked form) into an *in situ* conservation sites and possibly preserved in community seed banks (CGB's) for farmers to have easy access. This can enhance

ISSN 1817-3322

indirectly farmer's access to the national genebank. Bulked seed samples can be made available and diffuse to farmers through organized seed diversity fairs and through published agro-biodiversity catalogs (which contains characterization as well as evaluation data, particularly phenology and quality traits). Accessions in the core set can also be planted in a diversity block platform in non-replicated plots either in farmers training centers (FTC's) or *in situ* sites where farmers would encourage stopping and observing diversity and exercising selection. This would be one way where farmers obtain new seeds and planting materials so as to exchange knowledge and increase diversity in their farms and villages.

In general, as large collection may hamper the assessment of the genetic worth and available genetic diversity, core set of small size as established in this study may serve as an entry point for the assessment of the genetic variability, population structure and targeted allele mining. Core sets of such small size, is expected to enhance and facilitate germplasm evaluation against a number of biotic stresses in finger millet such as blast (a series disease that challenges finger millet production) and abiotic stresses (mainly drought). The constituted core set may also simplify the effort of screening accessions for quality traits particularly essential amino acids and malting quality. As finger millet is rich in important amino acids such as eleusinin, cystine, tryptophan and methionine that are very crucial to human health (NRC 1996). Thus, the accession in the core sets (particularly Core 4 with only representative 68 accessions) can allow proper analysis for such quality parameters with limited cost.

ACKNOWLEDGEMENTS

This work was logistically supported by the Ethiopian Biodiversity Institute (EBI). The author is thankful to Mr. Tegegnwork Wendwossen (from EBI) for his assistance in data collection.

REFERENCES

Balakrishnan R, Nair NV and Sreenivasan TV (2000): A method for establishing a core collection of *Saccharum officinarum* L. germplasm based on quantitative-morphological data. Genet. Resour. Crop Evol. 47: 1-9.

Brown AHD and Spillane C (1999): Implementing core collections-principles, procedures, progress, problems and promise. In: R.C. Johnson, R.C. Hodgkin, T. (Eds.), Core collections for today and tomorrow. Crop Sci. Soc. Am. Madison, pp. 1-10.

CSA (2013): Agricultural sample survey, report on area and production for major crops. Statistical Bulletin, Central Statistical Authority (CSA), Addis Ababa, Ethiopia.

Diwan N, McIntosh MS and Bauchan GR (1995): Methods of developing a core collection of annual Medicago species. Theor. Appl. Genet. 90: 755-761.

FAO (1996): Global Plan of Action for the Conservation and Sustainable Utilization of Plant Genetic Resources for Food and Agriculture, FAO, Rome, Italy.

Frankel OH (1984): Genetic perspectives of germplasm conservation. In: Arber, W.K., Llimensee, K., Peacock, W.J. and Starlinger, P. (Eds.), Genetic manipulation: impact on man and society, Cambridge University Press, Cambridge, England, pp. 161-170.

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

Grenier C, Hamon P and Bramel-Cox PJ (2000): Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs non-random sampling procedures, Using morpho-agronomical and passport data. Theor. Appl. Genet. 101: 190-196.

Harlan HV and Martini LM (1936): Problems and Results in Barley Breeding, U.S. department of agriculture, yearbook of agriculture.

Hilu KW, Dewet JMJW and Harlan JR (1979): Archaeobotanical studies of *Elueusine coracana* ssp. *coracana* (finger millet). Am. J. Bot. 66: 330–333.

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. Theor. Appl. Genet. 101: 264-268.

Jain SK, Qualset CO, Bhatt GM and Wu KK (1975): Geographical patterns of phenotypic diversity in a world collection of durum wheat. Crop Sci. 15: 700-704.

Mahajan RK, Bisht IS and Dhillon BS (2007): Establishment of a core collection of world sesame (*Sesamum indicum* L.) germplasm accessions. SABRAO J. Breed. Genet. 39: 53-64.

Mahalanobis PC (1936): The generalized distance in statistics, Pro. India Nat. Inst. Sci. 2: 49-55.

MOA (2013): Crop variety register, Addis Ababa, Ethiopia.

NBASP (2014): Ethiopia's fifth national report to the convention on biological diversity. Ethiopia Biodiversity Institute. Addis Ababa, Ethiopia.

ISSN 1817-3322

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

NRC (1996): Lost crops of Africa: Grains, National Academy Press, Washington, D.C.

Singh RK and Chaudhary BD (1985): Biometrical methods in quantitative genetic analysis. Kalyani Publ., New Delhi.

Suneson CA (1956): An evolutionary plant breeding method, Agron. J. 48: 188-191.

Tsehaye Y, BjØrnstad Å and Abay F (2012): Phenotypic and genotypic variation in flowering time in Ethiopian barleys. Euphytica 188: 309-323.

Upadhyaya HD, Dronavalli N, Gowda CLL and Singh S (2012): Mini core collections for enhanced utilization of genetic resources in crop improvement. Indian J. Genet. Resour. 25: 111-124.

Van der Plank JE (1963): Plant Disease: Epidemics and Control, Academic Press, New York, NY. 1963.

Van Hintum TJL and Haalman D (1994): Pedigree analysis for composing a core collection of modern cultivars, with examples from barley (*Hordeum vulgare*). Theor. Appl. Genet. 88: 70-74.

Van Hintum TJL (1999): The general methodology for creating a core collection. In: Johnson, R.C., and Hodgkin, T. (Eds.), Core collections for today and tomorrow. International Plant Genetic Resources Institute, Rome, Italy,

Vavilov NI (1951): The origin, variation, immunity and breeding of cultivated plants. Chron. Bot. 13: 1-366.

Ward JH (1963): Hierarchical groupings to optimise an objective function. J. Am. Stat. Ass. 58: 236-244.

Zeuli PLS and Qualset CO (1993): Evaluation of five strategies for obtaining a core subset from a large genetic resource collection of Durum wheat. Theor. Appl. Genet. 87: 295-30.

ISSN 1817-3322