Multivariate analysis of morphological variation in enset (*Ensete ventricosum* (Welw.) Cheesman) reveals regional and clinal variation in germplasm from south and south western Ethiopia

Zerihun Yemataw[1,2*], Kassahun Tesfaye[3,4], Murray Grant[5,6], David J. Studholme[6] and Alemayehu Chala[7]

[1]**Areka Agricultural Research Center, P.O. Box 79, Areka, Ethiopia**
[2]**Department of Microbial, Cellular and Molecular Biology, Addis Ababa University, Addis Ababa, Ethiopia**
[3]**Addis Ababa University, Institute of Biotechnology, P.O. Box 1176, Addis Ababa, Ethiopia**
[4]**Ethiopian Biotechnology Institute, Ministry of Science and Technology, P.O. Box 32853, Addis Ababa, Ethiopia**
[5]**Life Sciences, University of Warwick, Coventry, CV4 7AL. United Kingdom**
[6]**Biosciences, University of Exeter, Geoffrey Pope Building, Stocker Road, Exeter EX4 4QD, United Kingdom**
[7]**Hawassa University, College of Agriculture, P.O. Box 05, Hawassa, Ethiopia**

*Corresponding author: yemataw.zerihun@yahoo.com; alemayehuchala@yahoo.com

Abstract

Enset (*Ensete ventricosum* (Welw.) Cheesman) is cultivated by millions of people across Ethiopia in diverse agro-ecological and cultural settings, selecting for various agronomic traits. However, as for other underutilized crops, our understanding of the diversity and utilization of enset remains limited. This work sought to redress this limitation by estimating morphological diversity among enset accessions collected from major enset growing regions, including across altitudinal gradients. In total, landraces comprising 387 accessions originating from nine regions of Ethiopia were characterized using multivariate analysis of 15 quantitative traits. Cluster analysis grouped accessions in to five distinct classes with maximum number of accessions 338 in cluster (I) and minimum 1 in cluster (V). The clustering of accessions did not show grouping on the basis of region of origin. The first four principal components accounted for ~74% of the total variance. Linear discriminant analysis indicated that around 40.8% (160 accessions) and 45.2% (175 accession) of the studied accessions were correctly classified to their respective regions of origin altitude groups, respectively. The breadth of phenotypic differences in these 15 traits suggests significant degrees of genetic variation. These traits will be exploited to identify potential donors for future enset improvement efforts.

Key words: Agro-morphological variation, *Ensete ventricosum*, Ethiopia, landrace, multivariate analysis.
Abbreviations: AAU_ Addiss Ababa University; SARI_ Southern Agricultural Research Institute; SNNPRS_ Southern Nations, Nationalities and Peoples' Regional State.

Introduction

Enset (*Ensete ventricosum* (Welw.) Cheesman) belongs to the order Zingiberales, family Musaceae and the genus *Ensete*. It is a giant herbaceous monocotyledonous plant consisting of an adventitious root system, an underground stem structure known as corm and a pseudostem, which is formed from leaf sheaths, leaves and inflorescence (Smeds, 1955). Enset is a multipurpose crop used for human food and animal feed and traditional medicine as well as for ornamental purposes. Farmers say that enset is their food, their cloth, their house, their bed, their cattle feed and their plate (Brandt et al., 1997). Enset is eaten either as *kocho* (a bread-like food made from fermented corm and pseudostem), *bulla* (dehydrated starch rich juice collected during decortication of the pseudostem and grating of the corm, subsequently rehydrated from concentrate and

prepared as pancake or porridge), or *amicho* (boiled corm pieces, eaten like potato) (Brandt et al., 1997).
Enset is an indigenous crop grown in the highlands of Ethiopia, usually between 1200 and 3100 meters above sea level (masl) but scattered plants can be found at lower altitudes (Haile et al., 1996). Ideal growth conditions appear to be between 2000 and 2750 masl (Diro and Taboge, 1994) and for optimum growth enset requires an average rainfall of 1100 - 1500 mm per year and a mean temperature of 16–20 ^{0}C. Enset is not tolerant to freezing; frost damage on upper leaves is commonly observed at 2800 masl and serious stunting is seen above 3000 masl (Taboge, 1997). The productivity of the crop is very high compared to other root and tuber crops but varies depending on edaphic factors, altitude, cultural practices and varietal differences

(Birmeta, 2004). Ethiopian farmers usually grow enset as a mixture of different morphotypes (Tsegaye, 2002; Tesfaye, 2008; Yemataw et al., 2014). Enset landraces have been studied using morphological (Taboge, 1997; Yemataw et al., 2012; Bekele et al., 2013; Yeshitla, 2014) and molecular (Negash, 2001; Birmeta, 2004; Tobiaw and Bekele, 2011; Getachew et al., 2014; Olango et al., 2015) markers. The different studies have contributed to identifying useful germplasm for plant breeding purposes and for developing appropriate collection and conservation strategies. However, most studies concentrated on samples collected from only a subset of the growing regions. On the basis of observed diversity, Areka Agricultural Research Centre has collected and maintained about 600 enset landrace accessions (Yeshitla and Yemataw, 2012). Detailed knowledge about the collections, evaluation and cataloguing will help to elucidate the patterns of variation and the paths of evolutionary history, which are required for efficient utilization of the genetic potential held in germplasm collections (Brown, 1978). Multivariate methods like principal component analysis (PCA), cluster analysis (CA) and, discriminant analysis have proved useful for characterization and classification of plant genetic resources. These are powerful tools to estimate extent of genetic diversity for choosing potential parents in breeding programs, and to elucidate the patterns of variation in germplasm collection (Cowen and Frey, 1987). Selection of enset accessions displaying useful agronomic characteristics is dependent on knowledge of the extent of genetic diversity. Different investigators have described the usefulness of multivariate methods in many crops like barley (Abebe et al., 2010), tef (Assefa et al., 2003), sorghum (Ayana and Bekele, 1999), wheat (Damania et al., 1996), rice (Kanwal et al., 1983) and also enset (Bekele et al., 2013; Yeshitla, 2014). However, these studies, especially those on enset, were focused on the level and structure of diversity on enset within a given region, while regional variations was seldom considered. In addition, most of the enset landraces collected and preserved at the Institute of Agricultural Research have not yet been studied for their genetic diversity. This study was initiated to survey morphological diversity among enset accessions collected over 20 years from nine major enset growing regions and sampled across an altitudinal gradient. Farmers' selection practices were a driver to identify areas with high variation, which is important for selection and establishment of *in situ* conservation sites. The objectives of the present study were to: i) determine the extent and pattern of distribution of morphological variations for 15 quantitative characters in 387 enset accessions, representing a wide geographical range and, ii) to identify groups of accessions quantitatively similar in respect of those characters, using a range of multivariate statistical methods.

Results and Discussion

Cluster analysis

Cluster analysis demarcates genotypes into clusters, which exhibit high homogeneity within a cluster and high heterogeneity between clusters (Jaynes et al., 2003). The 387 accessions distinguished five clusters of varying size

(Table 12 and Supplementary figure 1). Furthermore, the diversity was reflected by the substantial variation among the cluster means for the 15 different characters (Table). Cluster I contained the highest number of accessions (338), accounting for 87.3% of the accessions tested in this study. Enset plants that fall into this cluster have intermediate maturity time, plant height, pseudostem height, pseudostem circumference, leaf sheath number and fermented squeezed *kocho* yield per hectare per year. These accessions were scattered across all regions but the majority were grouped in altitude group II (2001-2400 masl) and III (2401-2800 masl) (Table 3).

Cluster II was the second largest cluster comprising 45 accessions, representing 11.6% of the tested enset accessions. Accessions with the highest agronomic and yield-related traits were included in this cluster. This cluster included accessions from all regions, with the highest representation (15) being from the Kembata & Hadiya region.

Cluster III consisted of only two accessions with inferior morphological and agronomical characteristics. Enset accessions in this cluster have the shortest plant height, lowest pseudostem height, the narrowest pseudostem circumference and leaf width, the lowest leaf sheath before decortication, leaf sheath after decortication and the lowest fermented squeezed *kocho* yield per hectare per year.

Clusters IV and V each contained single accessions derived from lowland (≤2000 masl) and intermediate (2001-2400 masl) elevations in Gurage and Gamogoffa (Table 3). Accessions in cluster IV demonstrated intermediate values for yield-related traits, while cluster V constituted the best yielding accession with the highest values for morphological traits.

Overall, this study demonstrates that enset phenotypic traits could classify accessions according to their phenotypic similarity/differences using multivariate analysis. Hence, selection and crossing of enset germplasm accessions included in different clusters would provide greater heterosis in enset breeding programme in Ethiopia.

Our results indicate that accessions from different regions might have similar genetic backgrounds. This could be due to several possible reasons, most likely because farmers will select for a given character based on the adaptive role of characters for the environment, as well as sourcing and exchanging planting material between regions. Even though the geographical location of Sidama is distant from Gurage and from Kembata and Hadiya, accessions from these locations showed some similarity (Fig 1). Indeed, this example reflected a general pattern whereby the distribution of accessions in different clusters did not follow a definite pattern with regard to geographical origin. On the other hand, accessions from different regions based on the 15 pheno-morphic and agonomic traits were closely related regardless of their geographic origin. The collection of enset landraces at the Areka Agricultural Research Centre were recorded by the name identified by the farmer. While this nomenclature may be uniform across a specific the region that speaks the same language, some landraces (even possibly identical landraces) may have different names in different ethnic groups or different languages (Taboge, 1997; Yemataw et al., 2014) and the variation in utilization of the same landraces by local farmers (Negash, 2001). This

is illustrated by the clustering of genotypes with greatest morphological similarity, and the composition of these clusters did not include all genotypes derived from the same geographic origin. The result agrees with pevious reports in wheat (Ali et al., 2008) and field pea (Singh and Tripathi, 1985), which noted the absence of clear interrelationship between geographic origin and genetic diversity despite the presence of high intra-and interregional diversity among those crop accessions. Zubair et al. (2007) and Ahmad et al. (2008) also reported lack of association between morpho-agronomic traits and place of origin in mung bean and barley genotypes, respectively.

All the 10 pairs of clusters were significantly (P<0.01) different (Table 4). The maximum inter-cluster distance was observed between clusters IV and V (D^2= 221.17) (Table 5). Each of these "clusters" constituted a single accession from the Gurage and Gamogoffa regions. The second most divergent clusters were clusters III and V (D^2= 190.18); the accessions grouped in these clusters were collected from Dawro and Gamogoffa. On the other hand, the lowest (12.6) inter-cluster distance between clusters I and II indicates that the genotypes of these clusters were probably more genetically similar.

Maximum segregation and recombination in the subsequent generation is expected from crosses that involve parents from the greatest inter-cluster distance clusters. Thus these data when combined with other desirable agronomic attributes, provide an evidence-based selection method for future breeding programmes. Such genotypes can also be used in breeding programmes for developing bi-parental crosses between the most diverse and closest groups to break the undesirable linkages between yield and its associated traits (Haddad et al., 2004). However, Singh (1990) suggested that the selection of parents should also consider the specific traitsof each cluster and each accession within a cluster depending on the specific objectives of hybridization. Yan and Donaldh (1998) also recommended that in addition to genetic diversity, parents should express the optimum level of the desired component traits for enhanced yield, while being maximally resistant to biotic and abiotic stresses, and that they should fulfill quality parameters required in the breeders' target area.

Generally, the results of clustering and the D^2 analyses have shown that genotypes from the same collection sites were in different clusters and likewise accessions from different collection sites may cluster together, probably reflecting that the environment and farmers' selection criteria for a given character that enhances performance/yield of enset drives the similarity of most of the accessions from different origin.

Genetic distances from quantitative data allow inferences about the adaptation of populations (Camussi et al., 1985). Thus, classification using multiple agronomic characters identifies adaptation of a genotype and would improve the evaluation of genotype for potential adaptation (Souza and Sorrells, 1991). Moreover, Zhong and Qualset (1995) have suggested that the evolution of co-adaptive association of quantitative characters might contribute to the observed grouping together of accessions from geographically similar areas. The number of enset landraces grown at a given locality, their genetic similarity and the areas they occupy

over time and space are influenced by introduction (planting material source), exchange and supply. The accessions might have originally been introduced from the same source, followed by frequent exchange of planting materials among farmers across regions of the country and subsequent selection criteria for the same trait of interest (de Boef et al., 1996). Even if the original sources might vary, the crop might have also been forced to evolve in the same direction by this kind of local breeding for the same targets which may come from similar economic, social, cultural and ecological drivers in the area.

Principal Component Analysis

Principal component analysis (PCA) showed that the first four principal components accounted for about 74% of the total variance of the 387 enset accessions for the 15 quantitative traits (Table 5). Of these, the first three PCs explained about 65% of the gross population variance. The first PC contributed nearly 42%, and the second PC ~12% of the total variation. We used a criterion that characters with larger absolute values closer to unity within the first principal component influence the clustering more than those with lower absolute values closer to zero (Chahal and Gosal, 2002). Furthermore Hair et al. (1998) suggested that eigenvalues greater than one are considered significant and component loadings greater than ±0.3 were considered to be meaningful. As shown in Figure 2, fourteen parameters occupied the right side of the biplot whereas only one trait was observed in the upper left side.

The relative magnitudes of eigenvectors for the first PC indicated that traits like leaf sheath before decortication, leaf sheath after decortication, unfermented squeezed *kocho* yield per hectare per year, and pseudostem circumference contributed most to the total diversity and they were the ones that most differentiated enset populations. In the second PC, maturity time, corm weight before grating, and unfermented *kocho* yield per hectare per year had significant contributions. The analysis also revealed that plant height, leaf sheath number, leaf number and pseudostem height constituted a larger part of the total variance explained by the third PC. Leaf height, leaf number and fermented squeezed *kocho* yield per hectare per year were the major contributory parameters in the fourth principal component.

PCA was also conducted using the respective means of regional and altitude zones for the 15 quantitative traits to assess the regional and altitudinal pattern of variation. Considering an eigenvalue of more than one, the first three and two PCs explained 82.8 and 88.66% of the total variance among regions and altitude zones, respectively, (Table 6). The first three PCs contributed ~83% of the total regional variation. Central shoot weight before grating, corm weight before grating, pseudostem circumference, leaf number and leaf sheath weight after decortication had high loadings for the first PC, which accounted for >51% of variation. Similarly, ~18% of the overall variability among regions was accounted by the second PC, largely due to plant height, leaf height and unfermented *kocho* yield per hectare per year. Unfermented squeezed *kocho* yield per hectare per year and fermented squeezed *kocho* yield per hectare per year

Table 1. Clustering of 387 enset genotypes into eight clusters using mean of 15 quantitative characters

Cluster	Enset accessions[a]	No. of accessions	Percentage of accessions
I	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 14, 16, `17, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39, 40, 42, 44, 45, 46, 47, 49, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 76, 79, 80, 81, 82, 83, 85, 86, 87, 88, 90, 91, 92, 93, 96, 97, 98, 99, 100, 101, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, `155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 196, 197, 198, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 216, 217, 218, 219, 220, 221, 222, 223, 225, 227, 229, 230, 231, 232, 235, 236, 237, 238, 239, 240, 241, 242, 244, 245, 246, 247, 248, 249, 250, 251, 253, 254, 255, 256, 258, 259, 261, 262, 263, 265, 266, 267, 268, 269, 270, 271, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 285, 286, 287, 288, 289, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 302, 303, 304, 305, 306, 307, 309, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330. 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 362, 363, 364, 365, 366, 369, 370, 372, 373, 374, 475, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387	338	87.3
II	4, 12, 13, 15, 18, 37, 41, 43, 48, 50, 62, 74, 75, 77, 78, 84, 89, 95, 124, 125, 126, 171, 195, 199, 215, 224, 226, 228, 233, 234, 243, 252, 257, 260, 264, 283, 284, 290, 301, 308, 310, 361, 367, 368, 371	45	11.6
III	94, 102	2	0.5
IV	272	1	0.3
V	139	1	0.3

[a] numbers refer to accession codes; See the supplementary table for the details of the accession number and generic name.
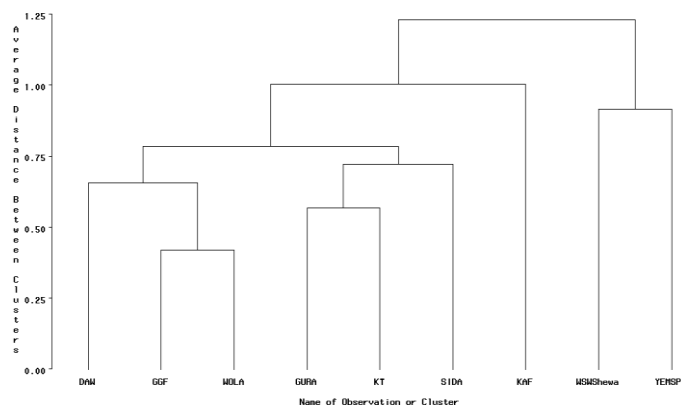

Fig 1. Dendrogram showing the clustering patterns of the nine regions of origin of enset accessions based on 15 phenotypic characteristics. DAW= Dawro, GGF= Gamogoffa, WOLA= Wolaita, GURA= Gurage, KT= Kembata and Hadiya, SIDA= Sidama, KAF= Kaffa, WSWShewa= West and South-west Shewa, YEMSPW= Yem special woreda.

Table 2. Mean values for the 15 quantitative traits in the five clusters identified in Table 1.

Cluster	MT[a]	PLHT	PSHT	PSCIR	LFNO	LFHT	LFWTH	LFSTHNO
I	4.48	5.13	1.69	1.15	10.25	3.34	0.65	18.95
II	4.13	6.47	2.01	1.44	11.46	3.79	0.75	21.78
III	5.01	2.62	0.96	0.61	5.73	1.49	0.35	14.25
IV	3.58	6.24	2.06	1.11	11.5	3.06	0.64	48.5
V	5.13	7.13	2.36	1.91	10.00	3.53	0.86	21.5

Cluster	LFSTHBD	LFSTHAD	CSBG	CORMBG	UFK	UNSQKOCH	SQKOCHO
I	56.479	27.15	15.17	25.34	31.98	13.41	9.37
II	107.94	54.16	23.84	44.78	63.07	25.43	17.55
III	10.46	3.71	2.21	11.71	7.06	2.17	1.34
IV	67.46	34.46	18.46	31.46	50.34	24.21	15.5
V	120.96	66.96	60.46	72.46	83.52	30.65	21.27

[a] MT= Maturity time, PLHT= Plant height (m), PSHT= Pseudostem height (m), PSCIR= Pseudostem circumference (m), LFNO= Leaf number, LFL= Leaf length (m), LFWTH= Leaf width (m), LFSTHNO= Leaf sheath number, LFSTHBD= Leaf sheath weight before decortication (kg), LFSTHAD= Leaf sheath weight after decortication (kg), CSBG= Central shoot weight before grating (kg), CORMBG= Corm weight before grating (kg), UFK= Unfermented *kocho* yield per hectare per year, UNSQKOCH= Fermented unsqueezed *kocho* yield per hectare per year, SQKOCHO= Fermented squeezed *kocho* yield per hectare per year.
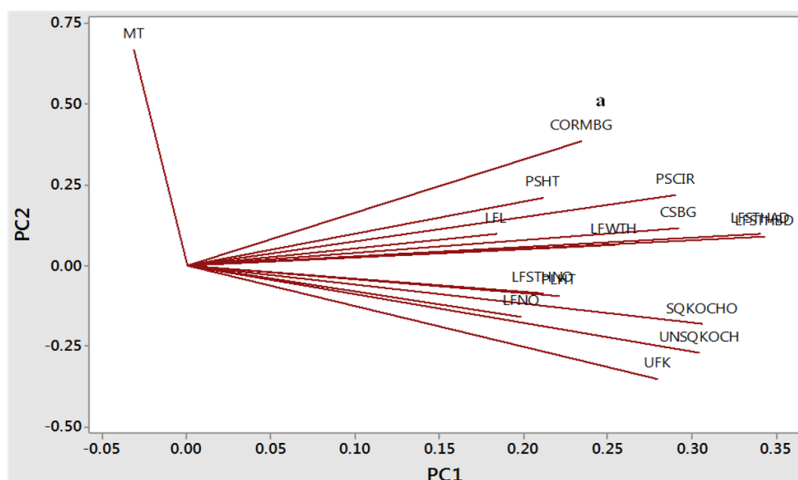
Fig 2. Loading plot of PC1 and PC2 for 387 enset accessions. [a]MT= Maturity time, PLHT= Plant height (m), PSHT= Pseudostem height (m), PSCIR= Pseudostem circumference (m), LFNO= Leaf number, LFL= Leaf length (m), LFWTH= Leaf width (m), LFSTHNO= Leaf sheath number, LFSTHBD= Leaf sheath weight before decortication (kg), LFSTHAD= Leaf sheath weight after decortication (kg), CSBG= Central shoot weight before grating (kg), CORMBG= Corm weight before grating (kg), UFK= Unfermented *kocho* yield per hectare per year, UNSQKOCH= Fermented unsqueezed *kocho* yield per hectare per year, SQKOCHO= Fermented squeezed *kocho* yield per hectare per year.

Table 3. Distribution of 387 enset accessions over five clusters by region of origin and altitudinal zone.

Region or altitudinal zone	Clusters					Total
	I	II	III	IV	V	
Region						
Kembata & Hadiya	63	15				78
Dawro	47	6	2			55
Gamogoffa	43	1			1	45
Wolaita	34	2				36
Sidama	33	8				41
Gurage	30	6		1		37
Yem special woreda	37	3				40
West & SW Shewa	31	1				32
Kaffa	20	3				23
Altitude zones						
≤2000 masl	39	4		1		44
2001-2400 masl	114	22	1		1	138
2401-2800 masl	160	17	1			178
>2800 masl	25	2				27
Total	338	45	2	1	1	387

Table 4. Pairwise generalized squared distances (D^2) among 387 enset accessions in five clusters.

Cluster	I	II	III	IV	V
I	0	12.6**	28.23**	139.47**	115.75**
II		0	61.17**	145.18**	83.14**
III			0	176.31***	190.18**
IV				0	221.17**
V					0

*** Indicates highly significant differences (p<0.001) and ** significant at p<0.01 level.

Table 5. Eigenvectors, eigenvalues, and percent variance explained by the first four principal components (PCs) for 15 different traits in 387 enset accessions.

Variable	Eigenvectors			
	Prin1	Prin2	Prin3	Prin4
MT[a]	-0.03	0.67	0.20	-0.03
PLHT	0.22	-0.09	-0.53	0.09
PSHT	0.21	0.21	-0.33	0.17
PSCIR	0.29	0.22	0.10	0.03
LFNO	0.20	-0.16	0.39	0.44
LFHT	0.18	0.10	-0.25	0.54
LFWTH	0.25	0.07	-0.19	0.23
LFSTHNO	0.21	-0.09	0.43	0.31
LFSTHBD	0.34	0.09	0.07	-0.02
LFSTHAD	0.34	0.10	0.07	-0.14
CSBG	0.29	0.12	0.17	-0.10
CORMBG	0.23	0.38	-0.12	-0.36
UFK	0.28	-0.35	-0.19	-0.15
UNSQKOCH	0.30	-0.27	0.09	-0.27
SQKOCHO	0.31	-0.18	0.16	-0.28
Eigen Value	6.65	1.91	1.23	1.01
Percent of variance explained	0.4437	0.1272	0.0824	0.0671
Cumulative percent of variance explained	0.4437	0.5709	0.6532	0.7203

[a] MT= Maturity time, PLHT= Plant height (m), PSHT= Pseudostem height (m), PSCIR= Pseudostem circumference (m), LFNO= Leaf number, LFL= Leaf length (m), LFWTH= Leaf width (m), LFSTHNO= Leaf sheath number, LFSTHBD= Leaf sheath weight before decortication (kg), LFSTHAD= Leaf sheath weight after decortication (kg), CSBG= Central shoot weight before grating (kg), CORMBG= Corm weight before grating (kg), UFK= Unfermented *kocho* yield per hectare per year, UNSQKOCH= Fermented unsqueezed *kocho* yield per hectare per year, SQKOCHO= Fermented squeezed *kocho* yield per hectare per year.

Table 6. Eigen vectors and Eigen values total for the first three and two principal components of the nine regions and altitude classes of origin of 387 enset accessions.

	Regions				Altitude		
	PC1	PC2	PC3	PC4	PC1	PC2	PC3
MT[a]	0.29	-0.26	-0.28	0.06	0.32	0.11	0.13
PH	-0.16	0.52	-0.13	0.07	0.27	0.28	0.01
PSH	0.25	0.25	0.09	0.40	0.19	-0.25	0.49
PSC	0.31	-0.01	-0.29	-0.17	0.26	-0.27	0.21
LN	0.31	-0.15	0.15	0.04	0.08	0.47	0.09
LH	0.22	0.40	0.15	0.25	0.15	-0.44	0.04
LW	0.12	-0.05	0.04	0.64	0.30	0.09	-0.29
LSN	0.26	0.20	-0.05	-0.43	0.10	0.28	0.59
LSBD	0.27	0.37	-0.03	-0.15	0.29	0.16	-0.23
LSAD	0.30	0.21	-0.04	-0.20	0.30	0.13	-0.26
CSBG	0.35	-0.05	-0.06	-0.07	0.29	-0.20	-0.19
COBG	0.33	0.02	-0.07	0.15	0.23	0.34	0.13
UFK	-0.22	0.38	0.27	-0.11	-0.29	0.16	-0.21
USQKB	0.13	-0.07	0.64	-0.08	-0.31	0.18	0.004
SQKB	0.20	-0.19	0.51	-0.15	-0.31	0.09	0.22
Eigen Value	7.71	2.66	2.04	1.55	9.23	4.07	1.70
Per cent of variance explained	51.45	17.73	13.62	10.33	61.50	27.13	11.34
Cumulative per cent of variance explained	51.45	69.18	82.80	93.13	61.53	88.66	1.00

[a] MT= Maturity time, PLHT= Plant height (m), PSHT= Pseudostem height (m), PSCIR= Pseudostem circumference (m), LFNO= Leaf number, LFL= Leaf length (m), LFWTH= Leaf width (m), LFSTHNO= Leaf sheath number, LFSTHBD= Leaf sheath weight before decortication (kg), LFSTHAD= Leaf sheath weight after decortication (kg), CSBG= Central shoot weight before grating (kg), CORMBG= Corm weight before grating (kg), UFK= Unfermented *kocho* yield per hectare per year, UNSQKOCH= Fermented unsqueezed *kocho* yield per hectare per year, SQKOCHO= Fermented squeezed *kocho* yield per hectare per year.

Table 7. Discriminant analysis of 387 enset accessions for region of origin and altitude based on 15 quantitative characters.

Regions	Original No.	Accession	Kembata-Tembaro /Hadiya	Dawro	Gamo Goffa	Wolaita	Sidama	Gurage	Yem	West &SW Shewa	Kaffa	Accessions classified under their regions of origin(%)
Kembata & Hadiya	78		46	8	10	4	4	5	1	0	0	58.97
Dawro	55		6	13	8	4	3	9	11	0	1	23.64
Gamogoffa	45		5	7	15	5	7	2	4	0	0	33.33
Wolaita	36		14	6	9	4	1	1	1	0	0	11.11
Sidama	41		15	4	5	2	13	2	0	0	0	31.71
Gurage	37		15	8	4	1	2	7	0	0	0	18.92
Yem special woreda	40		0	2	1	0	1	1	28	7	0	70
West & SW Shewa	32		1	1	2	0	0	0	2	23	3	71.88
Kaffa	23		0	3	0	1	0	0	2	6	11	47.83

Altitude zones	Original No.	Accession	≤2000 m.a.s.l	2001-2400 masl	2401-2800 masl	>2800 masl	Accessions classified under their regions of origin(%)
≤2000 masl	44		6	23	14	1	45.45
2001-2400 masl	138		2	71	65	0	47.3
2401-2800 masl	178		2	40	134	2	46.07
>2800 masl	27		1	11	11	4	48.15

Table 8. Regions of origin, altitude and numbers of landraces used for this study.

Collection region or altitude zone	Total no. populations
Region	
Kembata & Hadiya	78
Dawro	55
Gamogoffa	45
Wolaita	36
Sidama	41
Gurage	37
Yem special woreda	40
West & SW Shewa	32
Kaffa	23
Altitude zones	
≤2000 masl	44
2001-2400 masl	138
2401-2800 masl	178
>2800 masl	27

contributed 51 and 64% for the third PC which explained about 14% the total regional variance.

Altitude classes resulted in two PCs with eigenvalue greater than one, and the two PCs together explained 88.6% of the total variation. Maturity time, unfermented squeezed *kocho* yield per hectare per year and fermented squeezed *kocho* yield per hectare per year were the most loading contributors in the first principal component. Similarly, leaf number, leaf length and corm weight before grating showed greatest loading in the second principal component.

This PCA analysis confirmed that all morphological characters measured made contributions to the variance existing across the landraces. This in turn indicated the contribution of a number of traits towards the overall observed diversity. These results align with the findings of Assefa et al. (2003) where four PCs contributed 93.9% of total variation in tef populations. Ayalew et al. (2011) extracted four PCs that contributed 81% of the total variation in tef landraces. Similarly, in highland maize accessions of Ethiopia, 71.8% of total variation was accounted by the first four PCs (Beyene et al., 2005).

The combined PCA data indicated that agro-ecologies played a major role in discriminating accessions based on morphological traits, compared to the regions from which the landraces were collected. It was thus possible to observe how altitude and important ecological factors are related and influenced the diversity of enset in the country as previously suggested by Yeshitla and Yemataw (2012).

Discriminant analysis

Discriminant analysis of accessions using the region of origin as a grouping variable showed that 40.8% (160) of the accessions were correctly classified to their respective regions of origin (Table 7). The percentage of accessions correctly classified varies with regions. The proportion of enset accessions correctly classified with their region was the highest (72%) for West and South West Shewa. This was followed by Yem special district, Kembata & Hadiya, and Kaffa with 70, 59 and 48 % of the accessions respectively correctly classified within their collection region. On the other hand, Wolaita has the smallest percentage of accessions (11%) classified within their region of origin. Nearly 39% of accessions from Wolaita and over 40% of accessions from Gurage were grouped under the Kembata and Hadiya region. Accessions originated from Dawro were grouped under all regions except West and South West Shewa.

The discriminant analysis of the four altitudinal classes of enset accessions showed that 45.2% (175) of the accessions were correctly placed in their respective altitudinal class. The percentage of accessions correctly classified was higher for the second and fourth altitude classes compared to the others. Notably, around 59 and 55 % of studied accessions were misclassified according to the respective regions of origin and altitude groups. Evaluation of the predicted membership for the misclassified accessions of each region revealed that most of these accessions were flow among regions through planting material exchange and flow.

This variability is in line with those of Pecetti and Damania (1996), who reported that the higher the diversity of the group, the higher the probability of misclassification and separation. Similarly (Ayana and Bekele 1999) stated that the role of environmental factors has greater importance than regions of origin in discriminating sorghum accessions. When clustering of accessions did not follow their geographic origin, more emphasis should be given to agro-ecological parameters than to geographic origins as a source of diversity (Alemayehu and Becker, 2002).

Materials and methods

Description of the study site

Enset accessions were planted during April 2011 to April 2017 at the Areka Agricultural Research Centre, South Ethiopia, which hosts the coordination of the National Enset Improvement Program and is situated in the heart of one of the major enset producing areas of the country. The centre is located at 7^0 09' N latitude and 37^0 47' E longitude at an elevation range of 1750 to 1800 msl. The soil is silt loam with a pH of 4.8 to 5.6 and low to medium organic matter content (2.6 - 5.7%). The total amount of rainfall for the period from 2011 to 2017 was 1539 mm, and minimum and maximum mean temperatures were $14.5^{o}C$ and $25.8^{0}C$, respectively. Thus, the weather conditions were within the normal range for the growth and development of enset.

Plant materials

A total of 387 enset accessions, representing 381 landraces and six released varieties (controls), were used in this study. The 381 enset landraces were collected from nine different regions across varying altitudes (Table 8). All the materials are preserved *in situ* at Areka Agricultural Research Centre. The detailed passport data of the landraces including the regions and altitudes of collection and the vernacular names is summarized in Supplementary material 1.

The experiment was laid out in an augmented randomized complete block design consisting of three blocks in which the 381 landraces were planted in un-replicated plots and the six controls were replicated three times (once in each block) to estimate error variance. Eight suckers of each of the 387 landraces/accessions were planted in two rows with intra and inter-row spacing of 3m and 1.5m. All pre- and post-stand establishment management practices were performed as per the recommendations.

Plant characters and data recording

Data were collected on a total of 15 important quantitative (metric or count) pheno-morphological and agronomic traits (Supplementary material 2). All the traits were measured based on the procedure used by Taboge (1997) using four plants from each accession.

Statistical analysis

The data were standardized to a mean of zero and a variance of unity to avoid differences in scales used for recording data on the different characters before undertaking a series of multivariate analyses. Multivariate statistical analysis including cluster analysis (CA), principal component analysis (PCA) and discriminant analysis, as well

as non-hierarchical, and hierarchical clustering of accessions based on the average linkage method were performed using SAS V9.1.3 (SAS, 2004). Statistics, pseudo F statistics, pseudo t^2 statistic generated by SAS were examined to decide the number of optimum clusters. Genetic distances between clusters, as standardized Mahalanobis's D^2 statistics were calculated following the recommendation by Singh and Chaudary (1985) as follows:

$$D^2ij= (Xi - Xj) \, S^{-1} \, (Xi - Xj)$$

Where D^2ij is the distance between cases i and j; Xi and Xj are the vectors of the values of the variables for cases i and j; and S^{-1} is the pooled within groups variance-covariance matrix. The D^2 values obtained for pairs of clusters were considered as the calculated values of Chi square (X^2) and were tested for significant at (1 and 5%) probability levels against the tabulated value of X^2 for 'P' degree of freedom, where P is the number of parameters considered. Principal component analysis was performed using the correlation matrix to define the patterns of variation both between accessions, and between their regions of origin and altitudinal classes. In this study, only PCs with an eigenvalue greater than unity were considered important in explaining the variability. As suggested by Johnson and Wichen (1988) trait coefficient or eigenvectors greater than half divided by the standard deviation (square root) of the eigenvalue of the respective PC were employed as general guidelines for weighing the relative significance of traits contributing to the PCs. Linear discriminant analyses were also employed to examine the validity of the origin-based grouping of the accessions with respect to their regions and altitude groups of collection, and also to check the grouping of the accessions obtained through cluster analysis. The PROC DISCRIM procedure of SAS V9.1.3 (SAS, 2004) was used for analysis of discriminant analysis.

Conclusion

Findings of the current work revealed that there is likely to be high genetic diversity present within the Ethiopian enset accessions, though this is not uniformly distributed across the regions and altitudinal gradients. It was observed that accessions from the southern part of the country sharing similar ethno linguistic bases were closely related regardless of their geographic origin, though accessions from the same regions of origin might have different genetic backgrounds. There was no definite association between geographic origin and genetic diversity. All of the genetic distances between clusters are significant, suggesting the ability to incorporate desirable agronomic traits in subsequent generations from crosses that involve parents from the clusters characterized by significant distances. Future sampling of enset germplasm as a source of diversity should take place in areas with relatively large variation is evident, with due consideration to the cause of genetic erosion and depletion of resource. It must be reemphasized however, that morphological variation alone does not reflect the total variation, but this work opens the way to select accessions for a more comprehensive investigation of genetic diversity using molecular marker and quantitative characters. Such studies are underway and will probably provide a fuller picture about the genetic variation in Ethiopian landrace enset.

References

Abebe T, Bauer A, Lean J (2010) Morphological diversity of Ethiopian barley in relation to geographic regions and altitudes, Hereditas. 147: 154-164.

Ahmad Z, Ajmal S, Munir M, Zubair M, Masood M (2008) Genetic diversity for morpho-genetic traits in barley germplasm. PAK J BOT. 40: 1217-1224.

Alemayehu N, Becker H (2002) Genotypic diversity and patterns of variation in a germplasm material of Ethiopian mustard (*Brassica carinata* A. *Braun*). – Gen Resour Crop Ev. 49: 573 – 582.

Ali Y, Atta M, Akhter J (2008) Monneveux, P., Lateef, Z. Genetic variability, association and diversity studies in wheat (*Triticum aestivum* L.) germplasm. PAK J BOT. 40: 2087-2097.

Assefa K, Merker A, Tefera H (2003) Multivariate analysis of diversity of tef (*Eragrostis teff* (Zucc. Trotter) germplasm from western and southern Ethiopia. – Hereditas. 138: 228 – 236.

Ayalew H, Genet T, Wondale L (2011) Correlation and path coefficient analysis among yield component traits in tef (*Eragrostis teff* (Zucc. Trotter) landraces. LARCJI. 2: 180-185.

Ayana A, Bekele E (1999) Multivariate analysis of morphological variation in sorghum [*Sorghum bicolor* (L.) Moench] germplasm from Ethiopia and Eritrea. Gen Resour Crop Ev. 46: 273–284.

Bekele A, Diro M, Yeshitla M (2013) The diversity and associated yield components of enset (*Ensete ventricosum*) based on its agro-morphological Traits from southern Ethiopia. *SINET: Ethiop J Sci.* 36(1):49–54.

Beyene Y, Botha A, Myburg A (2005) A comparative study of molecular and morphological methods of describing genetic relationships in traditional Ethiopian highland maize. Afr J Biotechnol. 4: 586 – 596.

Birmeta G (2004) Genetic variability and biotechnological studies for the conservation and improvement of *Ensete ventricosum*. Doctoral thesis, Swedish University of Agricultural Sciences, Alnarp, 91 pp.

Brandt S, Spring A, Hiebsch C, McCabe S, Taboge E, Diro M, Welde-Michael G, Yentiso G, Shigeta M, Tesfaye S (1997) The 'Tree Against Hunger'. Enset-based Agricultural Systems in Ethiopia. American Association for the Advancement of Science, 56 pp.

Brown A (1978) Isozymes, plant population genetic structure, and genetic conservation. Theor Appl Genet. 52: 145–157.

Camussi A, Ottaviano E, Calinski T, Kaczmarek Z (1985) Genetic distances based on quantitative traits. Genetics. 111, 945–962.

Chahal G, Gosal S (2002) Principles and Procedures of Plant Breeding: Biotechnological and Conventional Approaches. Narosa Publishing House, New Delhi.

Cowen N, Frey K (1987) Relationships between three measures of genetic distance and breeding methods in oat (*Avena sativa* L.). Genome. 29: 97–106.

Damania A, Pecetti L, Qualset C, Humeid B (1996) Diversity and geographic distribution of adaptive traits in *Triticum turgidum* L. (durum group) wheat landraces from Turkey. Gen Resour Crop Ev. 43, 409–422.

de Boef W, Berg T, Haverkort B (1996) Crop genetic resources. In: Bunders J., Haverkort B. and Hiemstra W. (eds), Biotechnology; Building on Farmers' Knowledge. Macmillan, London and Basingstoke. pp. 103–128.

Diro M, Tabogie E (1994) Preliminary investigation on enset culture. In: Herath, E., Dessalegn, L. (eds) Proceedings of the Second National Horticultural Workshop of Ethiopia, 1-3 Dec. 1992, IAR, Ethiopia, Pp. 120-131.

Getachew S, Mekbib F, Admassu B, Kelemu S, Kidane S, Negisho K, Djikeng A, Inosters N (2014) A Look into Genetic Diversity of Enset (Ensete ventricosum (Welw.) Cheesman) Using Transferable Microsatellite Sequences of Banana in Ethiopia, J Crop Improv. 28:2, 159-183.

Haddad N, Boggo I, Muchibauer J (2004) Genetic variation of six agronomic characters in three lentil (Lens culinaris Medik.) crosses. Euphytica. 31:113-120

Haile B, Diro M, Tabogie E (1996) Agronomy Research on Enset. In: Enset-based Sustainable agriculture in Ethiopia (Tsedeke Abate, Clifton Hiebsch and Steve Brandt eds.). Proceedings of the First international workshop on Enset. Dec 13-21 1993. IAR, Addis Ababa, Ethiopia, Pp.235-241.

Hair J, Tatham R, Andrson E, Black W (1998) Multivariate data analysis. 5th Edn., Prentice-Hall International Inc., London, UK. ISBN-13: 978-0138948580.

Jaynes D, Kaspar T, Colvin S, James D (2003) Cluster analysis of spatio temporal corn yield pattern in an Iowa field. AJ. 95, 574–586.

Johnson A, Wichern W (1988) Applied Multivariate Statistical Analysis, Prentice-Hall, Englewood Cliffs, NJ.

Kanwal S, Singh M, Singh J, Singh B (1983) Divergent gene pools in rice improvement. Theor Appl Genet. 65: 263–267.

Negash A (2001) Diversity and conservation of enset (*Enset ventricosum* Welw. Cheesman) and its relation to household food and livelihood security in South-western Ethiopia. PhD dissertation, Wageningen University, The Netherlands, 247pp.

Olango T, Tesfaye B, Mario A, Mario E, Marcello C (2015) Development of SSR markers and genetic diversity analysis in enset *(Ensete ventricosum* (Welw.) Cheesman), an orphan food security crop from Southern Ethiopia. BMC Genetics. 16:98, DOI 10.1186/s12863-015-0250-8.

Pecetti L, Damania B (1996) Geographic variation in tetraploid wheat (*Triticum turgidum* convar. *durum*) landraces from two provinces of Ethiopia. – Genet Resour Crop Ev. 43: 395–407.

SAS (2004) Statstical Software, version 9.1.3. SAS Institute Inc., Cary, NC, USA

Singh P (1990) Bridging parents for incompatible crosses between Mesoamerican and Andean common beans. Annu. Rpt. Bean Improv. Coop. 33: 111.

Singh K, Chaudhary D (1985) Biometrical Methods in Quantitative Genetic Analysis. Kalyani Publishers, New Delhi–Ludhiana**.**

Singh S, Tripathi B (1985) Genetic divergence in pea. *Indian Journal of Genetics.* 45: 389-393.

Smeds H (1955) The Enset planting culture of Eastern Sidamo, Ethiopia. Acta Geographica. 13 :40.

Souza E, Sorrells E (1991) Relationships among 70 American oat germplasm. I. Cluster analysis using quantitative characters. Crop Sci. 31: 599–605.

Taboge E (1997) Morphological characterization of enset (*Ensete ventricosum* (Welw) Cheesman) landraces s and the association of yield with different traits. M.Sc. thesis Presented to School of Graduate Studies, Alemaya University, Alemaya, 89 pp

Tesfaye B (2008) On Sidama folk identification, naming, and classification of cultivated enset (*Ensete ventricosum*) varieties. Gen Resour Crop Ev. 55:1359–1370.

Tobiaw D, Bekele E (2011) Analysis of genetic diversity among cultivated enset (*ensete ventricosum)* populations from Essera and Kefficho, southwestern part of Ethiopia using inter simple sequence repeats (ISSRs) marker. Afri J Biotechnol. 10(70): 15697-15709.

Tsegaye A (2002) On indigenous production, genetic diversity and crop ecology of enset (*Ensete ventricosum* (Welw.) Cheesman). PhD dissertation, Wageningen University, The Netherlands, 197pp.

Yan W, Donaldh W (1998) Simulation and prediction of plant phenology for five crops based on photoperiod by temperature interaction. Annals of Botany, 81: 705-716.

Yemataw Z, Mohamed H, Diro M, Addis T, Blomme G (2014) Ethnic-Based Diversity and Distribution of Enset (Ensete ventricosum) Landraces s in Southern Ethiopia. J Ecol Nat Environ. 6(7): 244-251.

Yemataw Z, Mohamed H, Diro M, Addis T, Blomme G (2012) Genetic Variability, Inter-Relationships and Path Analysis in Enset (*Ensete ventricosum*) Landraces s. AJPSB. 6(1): 21-25.

Yeshitla M (2014) Cluster analysis for evaluation of genetic diversity in Enset *(Enset ventricosum* (Welw.) Cheesman) landraces s at Areka Condition. JPS. 2(1): 55-69.

Yeshitla M, Yemataw Z (2012) Past Research Achievement and Existing Gaps on Enset (Ensete Ventricosum (Welw.) Cheesman) Breeding. In Mohammed, Y., Tariku, H., (eds.). Enset Research and Development Experiences in Ethiopia. Proceedings of Enset National Workshop, 19-20 August 2010, Wolkite, Ethiopia, P 22-36.

Zhong Y, Qualset O (1995) Quantitative genetic diversity and conservation strategies for an allogamous annual species, dasypyrum villosum (L.) Candargy (Poaceae). Theor Appl Genet. 91: 1064–1073.

Zubair M, Ajmal S, Anwar M, Haqqani M (2007) Multivariate analysis for quantitative traits in mungbean [Vigna radiate(L.) Wilczek]. PAK J BOT. 39: 103-113.